

ADMINISTRATIVE DATA FOR
POLICY-RELEVANT RESEARCH:
ASSESSMENT OF CURRENT UTILITY
AND RECOMMENDATIONS
FOR DEVELOPMENT

A Report of the
Advisory Panel on Research Uses of Administrative Data
of the
Northwestern University/University of Chicago
Joint Center for Poverty Research



V. JOSEPH HOTZ, ROBERT GOERGE,
JULIE BALZEKAS, AND FRANCIS MARGOLIN,
Editors

Members of Advisory Panel on Research Uses of Administrative Data:

V. Joseph Hotz*

Chair, UCLA

Julie D. Balzekas

*Executive Director;
Communications Director,
Joint Center for Poverty Research*

Norman Bradburn†

*University of Chicago and
Committee on National Statistics*

Henry E. Brady

University of California and UC-DATA

Gerald Gates

U.S. Bureau of the Census

Robert Goerge

*Chapin Hall Center for Children
University of Chicago*

Carol Luttrell

Massachusetts Department of Revenue

Frances Margolin**

American Hospital Association

Bruce Meyer

Northwestern University

Deanna Schexnayder

*University of Texas
Center for the Study of Human Resources*

□ Werner Schink‡

*Department of Social Services
State of California*

Michael Wiseman

University of Wisconsin-Madison

*Research Affiliate of the Joint Center for Poverty Research

†Faculty Affiliate of the Joint Center for Poverty Research

‡Member of External Advisory Board of the Joint Center for Poverty Research

**Original Executive Director of the Advisory Panel

CONTENTS

PREFACE □	vi
EXECUTIVE SUMMARY	viii

CHAPTER 1

Introduction:

An Increasing Role for Policy-Relevant Administrative Data	1
1.1 The Research Potential of Administrative Data	3
1.1.1 The advantages of administrative data □	3
1.1.2 Examples that demonstrate the potential of administrative data	4
1.1.3 The use of administrative data outside of social	5
1.2 The Purpose of this Report □	6

CHAPTER 2

Developing Administrative Data for Research:

Definitions and Procedures	□ 9
2.1 Administrative Records and Management Information Systems □	10
2.2 Acquiring Agency Administrative Records □	10
2.3 Documentation of Source Data	11
2.4 Designing the Analytical Database	13
2.4.1 Tracking an individual over time □	13
2.4.1.a Duration of Service	14
2.4.2. Linking an individual across programs	14
2.4.2.a Probabilistic record matching	14
2.4.3 An individual in relation to characteristics and circumstances	15
2.4.3.a Relational databases	15
2.5 The Larger Context: What's at Stake?	16

CHAPTER 3

Developing Administrative Data for Research: Addressing Confidentiality and Privacy □

3.1 Confidentiality and Privacy □	18
3.2 Federal Protections	19
3.2.1 Nongovernment protections □	21
3.3 State Protections □	21
3.3.1 Unemployment insurance wage record data	22
3.3.1.a Creating a distributed wage database from state UI data	22
3.3.1.b State sharing of UI wage record data	22
3.3.2 Other state administrative systems □	23
3.4 Fair Information Practice Principles	24
3.4.1 Functional separation.	26
3.4.2 Informed Consent.	26
3.4.2.a The notice principle	27

.....	
3.4.2.b The fairness principle□.....	27
3.4.2.c Informed consent in statistical research	27
.....	
3.5 Safeguarding Privacy and Confidentiality.....	28
3.5.1 Restricting content and disclosure limitation techniques.....	28
.....	
3.5.2 Restricting access	28
3.5.3 Remaining challenges.....	29

CHAPTER 4

Assessing the Relative Strengths of Administrative and Survey Data for Research	31
4.1 A Comparison of Key Dimensions of Administration Data and Survey Data.....	32
4.1.1 Populations represented	32
4.1.2 Obtaining outcome measures and background variables and their quality.....	34
4.1.3 Time frames for which information is gathered	36
4.1.4 Obtaining information on program parameters and context □.....	37
4.2 The Strengths and Weaknesses of Administrative and Survey Data in Alternative Types of Research□.....	39
4.2.1 Descriptive research and trend analysis□.....	39
4.2.2 Data for use in causal inferences and evaluation of impacts of social programs and policies■■.....	40
4.2.3 Implications for relative merits of administrative data■■■.....	42
4.3 Data for Performance Monitoring and Accountability	44
4.4 The Potential for Linking Administrative and Survey Data□.....	45

CHAPTER 5

Examples of Ongoing Research Capacity	46
5.1 Common Lessons	47
5.2 Case Histories □.....	48
5.2.1 California Work Pays Demonstration Project	48
5.2.1.a Research using the CWPDP data sets	50
5.2.1.b Key lessons from the CDSS / UC DATA experience.....	51
5.2.2 The Integrated Database on Children's Services in Illinois.....	52
5.2.2.a Research using the IDB.	53
5.2.2.b. Key lessons from the IDB experience.	54
5.2.3 Massachusetts Department of Revenue Child Support Enforcement Division	54
5.2.3.a Research using CSE data.	55
5.2.3.b The Massachusetts Longitudinal Database for Research on Social Service Programs.....	55
5.2.4 Continuing development of archived data by Texas.....	57
5.2.4.a Research using Texas administrative data■■.....	58
5.2.4.b Archiving efforts across agency lines	59
.....	
5.2.4.c Key lessons from the CSHR experience	60
5.2.5 Building a shared information system in Oregon	60

5.2.5.a Research using Oregon SIS □.....	62
--	----

CHAPTER 6

Developing the Research Potential of Administration Data: Summary of Findings and Recommendations	63
6.1 Findings: Where We Are and Where We Need to Go	63
6.2 Key Findings	67
6.3 Recommendations for Developing the Research Value of Administrative Data	69
6.3.1 Fostering institution building□.....	69
6.3.2 Further assessment of confidentiality and privacy concerns □.....	71
6.3.3 Assessing and improving the quality and cross-state comparability of administrative data for public assistant programs □.....	72
6.3.3.a Assessing the quality and validity of administrative data	73
6.3.3.b Improving cross-state data comparability of administrative data.....	74
6.4 Recommendations in Three Key Areas	76
6.5 Concluding Observations □.....	77
 GLOSSARY OF TERMS	 78
 REFERENCES □.....	 80
 APPENDIX I <i>Examples of Successful Data-Sharing Laws and Agreements</i>	 84
APPENDIX II <i>Amendment to Illinois Public Aid Code (305 ilcs 5/124.33 new)□.....</i>	89
APPENDIX III <i>Evaluation of CalWORKS Program Implementation□.....</i>	92
APPENDIX IV <i>1997 Legislature, Florida, New Shared Information Statute</i>	96

PREFACE

The Advisory Panel on Research Uses of Administrative Data was formed under the auspices of the Northwestern University/University of Chicago Joint Center for Poverty Research in the fall of 1996. One aspect of the Poverty Research Center's mission is to support research on the effectiveness of policies aimed at reducing poverty. In support of that goal, the center sought funding from the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, to form an Advisory Panel to assess the development of research-ready data from state administrative sources in the areas of public assistance, public health and welfare and for use in policy and academic research. The Advisory Panel, composed of researchers, state and federal officials, and experts in the area of data protection and archiving, first met in the fall of 1996 and continued to meet and correspond through 1997.

From the beginning, the Advisory Panel sought to understand as thoroughly as possible, past and current uses of administrative data sources for policy-relevant and evaluative research and to use what was learned to inform a variety of audiences, including program managers, policymakers, and policy and academic researchers, about the present research utility of administrative databases. More important, the panel sought to identify the issues central to the continued productive use of this rich source of information on our nation's poor populations.

The panel spent a year gathering and synthesizing information from a variety of sources. This report characterizes existing state administrative databases capable of sustaining various types of research; assesses key concerns that must be addressed for administrative data to become a widely used basis for research; describes instances where administrative data sources have been developed; identifies the strengths and weaknesses of administrative data compared with survey data; and makes recommendations for enhancing the quality, availability, and use of administrative data.

This report is limited in its scope, owing, in some part, to limitations on the time and resources the panel could bring to bear on this topic, but more important to the nascent stage of ongoing programs to build, maintain, and use state administrative databases in research. It is the panel's hope that its findings will stimulate expansion of that capacity and a wider use of such data in conducting informative, policy-relevant research on program effectiveness and the well-being of the nation's poor and disadvantaged.

There are many people who deserve a great deal of thanks. As the final stage of producing the report, we invited a select group of external reviewers representing various interests and concerns of the utility of administrative data for research to read and respond to our findings. We wish to thank the following people for their critical assessment and considered suggestions of a draft of this report: Barbara Blum, National Center for Children in Poverty, Columbia University; Professor Catherine Born, University of Maryland-Baltimore County; John Bouman, Poverty Law Project, National Clearinghouse for Legal Services, Inc.; Dr. Virginia DeWolf, Chair, American Statistical Association's Committee on Privacy and Confidentiality; Linda Gage, Chief Demographer, State of California; Dr. John Haltiwanger, Chief Economist, U.S. Bureau of the Census; Fredericka Kramer, Welfare Information Network; Professor Julia Lane, American University; Kathlene Larson, Iowa Department of Human Services; Professor Robert MoYtt, Johns Hopkins University; Howard Rolston, Agency for Children and Families, U.S. Department of Health and Human Services; Dr. Matthew Stagner, Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services; and Professor David Stevens, University of Baltimore. Their comments proved invaluable, making our report more focused and more accessible.

I want to thank Professor Rebecca Blank, the founding director of the Northwestern/University of Chicago Joint Center for Poverty Research, for her encouragement and

commitment to this panel and its work, and her successor, Professor Susan Mayer, who not only continued that support, but also read and provided extremely useful comments on an earlier draft. More personally, I wish to thank them both for their support, counsel and willingness to lend an ear at those times when I thought the task of guiding this report to its conclusion was hopeless.

I also wish to express my sincere thanks to two people who truly made this report possible. Without the persistence, organizational skills, and wise counsel of Frances Margolin, the panel's original executive director, this panel would have never gotten off the ground. Before she left to join the Hospital Research and Educational Trust of the American Hospital Association, Francie helped to give structure to the report, framing the issues we needed to address to fulfill our mission. We are also indebted to Julie Balzekas, the Poverty Center's Communications Director, for her willingness to take over as executive director after Francie's departure. Julie's intellectual commitment to the enterprise and her ability to make sense of the disjointed prose of the ten members of this panel was absolutely crucial to turning a jumble of ideas into what I think is a coherent report. To both, let me offer my personal thanks. It has truly been my pleasure to have had the opportunity to work with each of them.

Finally, let me take the unusual step of thanking my fellow members of this panel. While always a challenge to get nine independent and opinionated individuals to agree on anything, I can truly say that chairing this group was an enjoyable and rewarding experience for me. Each member of the panel made important contributions to this report and did so in a pleasant and collegial way. I thank them for their cooperation and for all that they taught me about administrative data over the past year.

V. Joseph Hotz
Chair, Advisory Panel on Research Uses of Administrative Data

January 1998

EXECUTIVE SUMMARY

The report of the Advisory Panel on the Research Uses of Administrative Data¹ is concerned with administrative data collected at the state and local levels in the operation of government programs for the poor, such as AFDC/TANF, Food Stamps, Medicaid, and foster care. In addition to their “record-keeping” function, administrative data increasingly are used to monitor and evaluate program performance and ensure agency accountability.

The panel undertook this study at a time when American public assistance policies—particularly those aimed at families with dependent children—are changing. Governing authority and financial responsibility for public assistance have always been shared by federal, state, and local governments. However, as a consequence of the Personal Responsibility and Work Opportunities Reconciliation Act of 1996 (PRWORA), control of the policies and programs that affect poor families with children has largely devolved from Washington to the states.

At the same time, the aim of cash assistance to poor families with children has also changed from one of a federal entitlement to income support that sometimes provided education and job training to a time-limited benefits program principally focused on moving able-bodied adults into the labor force and their families off the welfare rolls. To achieve these new goals, state-designed programs must now develop and implement large-scale activities and services to help individuals and families become self-sufficient.

These changes are occurring in the presence of considerable uncertainty. Under the old law, the federal government had the means to collect relatively comparable state-generated program data. With the recent and profound devolution of family-related policies and programs to states and localities, the federal government no longer has a reliable means for monitoring what states are doing and how recipients are making out. Moreover, state and local governments have mixed experiences in producing reliable intrastate information on the effectiveness of alternative policies, much less reliable and valid data that permit interstate comparisons. Indeed, the latter, absent federal guidelines, is nearly impossible to achieve.

Consequently, in this report, the Advisory Panel seeks to:

- a. Describe the key practical and political considerations of transforming the information in these programmatic records into research-ready databases;
- b. Identify the strengths and weaknesses of administrative data, relative to that gathered in national surveys, for use in descriptive and evaluative research and in accountability-based monitoring of program performance;
- c. Describe examples of several states’ efforts to develop an ongoing capacity to use administrative data for both program and policy evaluations; and
- d. Make initial recommendations that will improve the quality and usefulness of administrative data for policymakers and program administrators.

¹ The Advisory Panel on the Research Uses of Administrative Data was formed under the auspices of the Northwestern University/University of Chicago Joint Center for Poverty Research in the fall of 1996. The center sought funding from the Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, to form an Advisory Panel to assess the development of research-ready data from state administrative sources in the areas of public assistance, public health, and welfare for use in policy and academic research. The Advisory Panel, composed of researchers, state and federal officials, and experts in the area of data protection and archiving, first met in the fall of 1996 and continued to correspond and meet through 1997.

The panel examined several states' efforts to develop intrastate databases from administrative records and reviewed how these databases are used to monitor and evaluate key public assistance programs and the disadvantaged populations they serve. In addition, we consulted many additional experts on the production and use of administrative data. We sought to obtain their assessments of the potential value that these sources of data may have in the future, and what issues need to be addressed if these nascent local efforts are to be replicated in other states.

Our report synthesizes the findings from our investigation and recommends ways that various groups—including policymakers, administrators, researchers, and foundations—can develop better administrative data for monitoring and evaluating the implemented welfare reforms. The remainder of this section summarizes our findings and recommendations.

Findings

1. The Value of Administrative Data

The Personal Responsibility and Work Opportunities Reconciliation Act of 1996 (PRWORA), P.L. 104-193, has set in motion an array of new policies and programs, including Temporary Assistance for Needy Families (TANF), which replaces the Aid to Families with Dependent Children (AFDC).

To answer the critical policy and program questions of “what works,” “for whom,” and “at what cost,” the Advisory Panel finds that:

- **Policymakers and program administrators will require more and better data sources than they now have if they are to adequately monitor program operations and evaluate program outcomes.** Program administrators and policymakers will need reliable state and local data if, among other things, they are to: summarize program operations; determine who is being served by which programs, who is being underserved, who is not being served but should be, and how services can best be targeted to those in greatest need; determine which strategies and services are most cost-effective; track individual work histories; track individual and family earnings and income; and describe the conditions of poor children and their families relative to the conditions of other households.
- **Current national survey research data cannot adequately monitor the diverse, local programs currently being established by state and local governments.** In the context of devolution, none of the national cross-sectional and longitudinal data sets is large enough to support separate analyses of poverty-related issues for any but the few largest states.
- **To obtain reliable information over time and across programs and agencies, it will be necessary to augment current administrative databases and to link them together.** Administrative data provide detailed and accurate program information, large sample sizes that allow for more types of analyses, and state-specific data that reflect variations in state and local programs.
 - For example, *administrative data offer the advantage of allowing for substate analyses*, thereby allowing the many TANF waiver programs operating in a limited number of counties to be better studied.
 - *Administrative data can also provide information on the same individual or case over long periods of time.* Such capabilities are increasingly important if we are to understand how, for example, recipient behavior and well-being change

in response to both time-limited benefits and varying economic and labor force conditions.

- *And due largely to advances in computer technology, linking administrative databases is easier, less expensive, and more reliable than ever before.* Such efforts can provide richer, more comprehensive information on how the poor, the working poor, and others are faring and how, and to what extent, they contribute to or consume public tax and transfer benefits.

2. Key Operational Issues for Developing Research-Relevant Administrative Databases

A clear case can be made for greater reliance on state and local administrative data systems for monitoring and evaluating public assistance programs in the future. To date, administrative databases have mostly been used in one-time evaluations based on random-assignment designs. While that experience offers many valuable lessons for improving administrative data, the structural changes in the welfare system under PRWORA may mean that states will be less likely to use experimental designs to evaluate their public assistance programs, if they conduct any impact evaluations at all. In this context, states and local governments are likely to make even greater use of administrative data for whatever evaluations they conduct.

Based on our review of past evaluations and research using administrative data and our investigation of present efforts to develop an ongoing capacity to provide research-ready administrative data, the Advisory Panel finds that:

- **There are three operational issues that can “make or break” the development of administrative databases: (1) negotiating appropriate interagency agreements; (2) negotiating agreements in which agencies retain adequate control over any new demands researchers impose on agency employees and the nature of the information researchers may disclose about agency operations; and (3) developing protocols that protect the privacy of clients and the confidentiality of data.**

EXAMPLES OF WORKING MODELS

The Advisory Panel found that when the issues of interagency cooperation, researcher-agency agreements, and client privacy are worked out in a mutually satisfactory way for all parties, solid, ongoing working relationships developed. In turn, over time, and across discrete projects, such relationships more readily remove obstacles to using administrative data for research. Indeed, in some instances, researchers and agencies were able to agree to take additional steps to improve and use administrative data for research on an ongoing basis. The Advisory Panel describes five successful collaborative efforts in this report:

California, where the collaboration of state and university researchers on data collection and evaluation for California’s federal AFDC waiver, the 1992 California Work Pays Demonstration Project, has led to the creation of five ongoing databases and the construction of analytic data sets for continued program evaluation and research.

Illinois, where a collaborative effort between university researchers and a single state agency in the early 1980s has evolved into a multiservice, integrated research database constructed from administrative data gathered by numerous public agencies serving children and families in Illinois, and now used to track the impact of reforms on caseloads and across agencies.

Massachusetts, where efforts to link tax administrative databases with other agency databases for the purpose of enforcing child support payments demonstrated the potential of administrative data for research and evaluation and earned support for the creation of a longitudinal database for research and evaluation of social service programs, which is now in use.

Texas, where the availability of administrative data for research has been largely facilitated by the establishment of performance measures by the state legislature in the early 1990s, evaluation of the Texas JOBS program, implemented in 1990, and the multiagency data collection and data sharing that both required and permits the establishment of key performance measures for workforce development and other programs.

Oregon, where an integrated database project, still under development, was mandated by the state legislature in 1993 in order to provide a database for future evaluations of welfare reform.

In examining these efforts, the Advisory Panel noted the following aspects of successful operations:

- **A collaboration between one or more state agencies and outside academic and independent research groups or institutions is key to developing successful, ongoing administrative databases for research.**
- **Initial development of ongoing administrative databases tends to be idiosyncratic and entrepreneurial, where someone involved in the research enterprise possesses a larger vision of what administrative data might do and is able to implement the vision.**
- **Ongoing databases are more typically the result of “bottom-up” rather than “top-down” development efforts. That is, they tend to be the result of localized and more idiosyncratic efforts as opposed to mandates from above.**
- **A key feature of the “entrepreneurial” effort that initiated and sustained the existing databases is the presence of someone or some group that holds a longer-run perspective so that the database is not viewed as a “one-shot effort,” useful only for a single project or contract.**

From these observations, the Advisory Panel concludes that a mutual investment by social assistance agencies, policymakers, and researchers in an entrepreneurial effort to create and sustain ongoing capacity will improve the quality of administrative data for research and make possible the monitoring and evaluation demanded by emerging programs and policies.

QUALITY AND INFORMATION CONTENT OF ADMINISTRATIVE DATA: POTENTIAL LIMITATIONS

Another set of challenges has to do with the nature and quality of administrative data. Primarily because administrative data are gathered in the context and for the purpose of administering a program, issues concerning the quality and information content of administrative data pose considerable obstacles to drawing inferences about program trends and impacts. The Advisory Panel identified the following potential concerns.

- **Administrative data possess some limitations that diminish their value in certain types of research, including: (1) inability to estimate such things as the**

rates of program participation; (2) inability to measure all outcomes, such as indicators of well-being that would not be tracked in the program-based data, or to measure anything when a person is “off the program”; and, (3) difficulty in comparing programs across states in the absence of standardized information collection.

However, the panel also sees significant opportunity to offset these limitations by linking data. While administrative data from one program seldom contain enough information for a useful evaluation, by linking administrative data from different programs it is possible to obtain an array of explanatory and outcome variables. Further, linking information from state administrative databases with survey data on individuals or households holds considerable promise.

RECOMMENDATIONS TO FOSTER THE DEVELOPMENT OF ADMINISTRATIVE DATA

Based on these findings, the Advisory Panel offers the following recommendations to foster the development of administrative data as an integral data source for public assistance research in the future. The recommendations cover three key areas: (1) *institution building*; (2) *confidentiality and privacy protection*; and (3) *assessing the quality and comparability of interstate administrative data*. They are summarized below.

Fostering institution building

Across the country, opportunities are emerging for developing ongoing administrative databases for research on social assistance programs and policies. It is the Advisory Panel’s view that states and the nation need to build on these promising efforts and develop permanent, ongoing administrative data capacities. To help realize that goal, the panel offers three sets of recommendations to foster the construction of permanent administrative data “institutions.”

- **A centralized and ongoing repository of information on administrative data efforts should be established (and funded).**
- **States without administrative databases organized for research should be encouraged to establish partnerships with independent research organizations, such as those at universities, to help develop, maintain, and use administrative databases on an ongoing basis for program monitoring and evaluation.**
- **National organizations, such as American Public Welfare Association (APWA) or the Welfare Information Network (WIN), as well as organizations and groups within the academic community, such as Association for Public Policy Analysis and Management (APPAM) and the National Association of Welfare Researchers and Statisticians (NAWARS) need to find ways to recognize and encourage the use of administrative data in research.**

Further assessment of confidentiality and privacy concerns

While the Advisory Panel found that existing principles and recommendations regarding confidentiality and privacy apply to the research uses of administrative data, some new issues need to be addressed, having primarily to do with disclosure limitation techniques and the applicability of federal legislation to particular states. Therefore,

- **The Advisory Panel calls on independent organizations (such as the National Research Council’s Committee on National Statistics) as well as professional organizations (such as the American Statistical Association) to conduct a more thorough assessment of the adequacy of existing principles and practices that will protect the privacy and confidentiality of the information contained in administrative databases.**

Assessing and improving the quality and crossstate comparability of administrative data for public assistance programs

Great strides have been made in the “science” of developing administrative databases, especially those that contain longitudinal information on program participants and those that consist of data linked across various databases. Nonetheless, it is the Advisory Panel’s assessment that many unanswered questions persist regarding the quality and usability of administrative data for many types of research, and this is especially true for evaluating the impacts of emerging state- and county-based welfare programs under PRWORA. The panel strongly believes that more research on the comparability of administrative and survey data needs to be done if administrative data are to become a trusted and appropriately used source of data in high-quality research.

- **The Advisory Panel urges that funding be provided by agencies such as the National Science Foundation, private foundations, and government agencies themselves to further research and analysis on such questions as: (1) quality of administrative data; (2) comparability with other data sources, such as survey data; (3) methodological strategies for dealing with such analytic issues as the denominator problem, which affect the range of use of data; and, (4) the interactions of research and management objectives and how this affects the structure and quality of such data.**
- **The panel also urges research organizations, such as the Joint Center for Poverty Research, and academic publishers and journals to encourage and help legitimize such research by creating “outlets” for it, including convening conferences and supporting volumes or special issues of journals on these topics.**
- **Further, the panel would urge those working on the “management” side of the equation, including professional organizations for the public sector, to collaborate and help support efforts to develop higher quality administrative data.**

The Advisory Panel’s final recommendation concerns data comparability across states. If cross-state comparisons are to prove informative as alternative programs and policies are monitored and evaluated, data that contain comparable measures and populations at the state level are essential. Clearly, administrative databases can play a crucial role in cross-state comparisons. But to play that role, attention must be paid to achieving greater comparability of information and populations in these databases. Accordingly, the panel offers the following recommendation to highlight this important issue.

- **Develop guidelines and standards to ensure that comparable and high-quality data are gathered across states and across agencies within states.**

We suggest that the National Research Council’s Committee on National Statistics be commissioned to establish an expert panel to assess and make recommendations on ways to foster such data comparability. This expert panel should include and work with representatives from state and local governments. It should also seek input from

professional organizations such as the APWA and NAWRS. Finally, this panel should recommend structures and institutional arrangements that will encourage an ongoing partnership between the states and federal government for gathering objective, high-quality, and comparable data on populations receiving (or at risk to receive) public assistance provided under PRWORA and related programs. The National Center for Health Statistics and National Center for Education Statistics might be considered possible institutional models for public assistance data-gathering.

CHAPTER 1
**Introduction: An Increasing Role
 for Policy-Relevant Administrative Data**

American social assistance policy is changing. Federal programs are being restructured. Governing authority and financial responsibility for social assistance is moving from Washington to the states. States are experimenting with a wide range of changes in the types of social services they provide, and differences across the states are growing. In this new era of social assistance policy, it has become essential to monitor these policy changes and assess their impacts. In particular, information or data are needed to¹:

1. **Describe the conditions of children and the poor.** The conditions include income levels, housing conditions, work effort, rates of teenage childbearing, educational outcomes including test scores, health status, immunization, and other use of health facilities.
2. **Summarize program operations.** This information includes counts of those entering and leaving programs, reasons for program use, and costs. In localities where outcomes- or performance-based accountability has been instituted, summary information provides an essential tool for managing programs.
3. **Determine who is being served by programs and who is not.** This area includes specific information to meet reporting requirements under Temporary Assistance for Needy Families (TANF) and broader information such as the degree of self-sufficiency among program participants and the well-being of children.
4. **Evaluate what “works” and what does not.** Good answers are needed to questions such as: What training programs get people jobs and allow them to keep them? What strategies prevent teenagers from having children and teen mothers from having second children?
5. **Determine how services can be targeted to those who most need or benefit from them.** Reliable answers are needed to such questions as: Which women need assistance finding a job immediately and which need to be trained first? Which women need help with childcare before either strategy can begin?

Currently, most of the data used to address these research objectives are obtained from household surveys. Because of their national scope and the variety of information gathered, the Current Population Survey (CPS) and Survey of Income and Program Participation (SIPP)² have served academic-based evaluation research on social policy and trends for over 30 years. Data from such national longitudinal surveys as the Panel Study of Income Dynamics (PSID), and the National Longitudinal Survey of Youth (NLSY) have also been

¹ Certain terms used throughout this report have specific meaning in the context of our analysis. We have included definitions of these key terms in the glossary.

² In the recently enacted Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA), the U.S. Congress has called on the U.S. Bureau of the Census to conduct a new survey, the Survey of Program Dynamics (SPD), “on a random national sample of recipients of assistance under State programs funded under this part [of PRWORA] and (as appropriate) other low-income families” to “enable interested persons to evaluate the impact of the amendments made by title I of the [Act]” (Public Law 104, 104th Congress, August 22, 1996, 110 STAT. 2156-7). The 1992 and 1993 SIPP panels will provide the core of the special Survey of Program Dynamics.

put to work in such analyses. In addition, more specialized or state- or region-based surveys, such as the National Health Interview Survey (NHIS), and the Behavioral Risks Factors Surveillance System (BRFSS), have been employed to monitor public health.

But as one of the consequences of the greater independence that states now have in setting social assistance policy, national household surveys will no longer be adequate for several reasons:

- It will be impossible to study the many state-level initiatives using national data sets as they will not have sufficiently large state-level samples.
- Survey designs and questionnaires are unlikely to respond quickly enough to program changes to insure that the appropriate people are surveyed and the appropriate questions are asked.
- Moreover, with different programs in different states, the information gathered from surveys is unlikely to be well tailored to each state.
- To answer long-term questions and meet specific legislative requirements (such as time limits), longitudinal data will be needed.
- The occasional, one-time evaluations of the past (usually conducted as part of federal program waivers) will not keep states informed of how their new programs are working.

The increased need for administrative data has come at a time when more can be expected from these data. Already, some states are gathering some of the information mentioned above. Nevertheless, with regard to research, social assistance administrative data often have been neglected, and their use has been piece-meal. One of the main limitations has been the sheer size of the data sets. Further, much of the information was extraneous to particular uses, while important pieces of information were missing. These problems have lessened in recent years due to increases in computer speed. Improvements in speed and methods have also allowed improved matching techniques. Key attributes of individuals or families that once were missing from individual data sets now may be obtained from another data set by matching. Administrative data are just beginning to reach their research potential.

1.1 The Research Potential Of Administrative Data

This report is about a particular source of information that can be used to monitor and evaluate emerging programs and the nation's changing policy context as it moves into the 21st century—administrative data. More precisely, this report deals with data collected in the operation of state and local public assistance programs for the poor and disadvantaged. In the face of the data challenges noted above, it is the Advisory Panel's assessment that administrative data will need to play an integral role in future efforts to monitor and evaluate the effects of America's evolving social and health policies and programs. Furthermore, it is this panel's belief that the capacity of administrative databases must be expanded to meet the administrative and research needs outlined above.

1.1.1 The advantages of administrative data

For many types of analyses and research, administrative data have always offered advantages that are only accentuated by recent policy and program changes. As discussed further in Chapter 4, these advantages include:

- **Detail and accuracy of program information.** Administrative data are detailed and very accurate measures of program status and outcomes. Detailed information on the

characteristics of participants, the services they have received, and the actions they have taken cannot be obtained in any other way. The types of information available include monthly information on welfare receipt, family composition, training, earnings, health services received by children, and many other items. Data on these topics will be essential to meet legislative requirements such as time limits and requirements on the fraction of TANF participants engaged in work or training.

- **Large sample sizes permit more types of analyses.** Administrative data sets typically include thousands—if not millions—of records and often include the entire universe of participants. The large sample sizes usually permit state-level analyses even in the smallest states. The data may even allow substate analyses, which have assumed increased importance as many AFDC/TANF waivers have created programs that operate in only a few counties. The large sample sizes allow small program effects to be more easily detected, and studies of the different effects of different programs (or different effects of the same program on different groups of individuals) easier to perform.
- **State-specific data can reflect state programs.** If welfare programs have different forms in different states, then only a state-level data collection effort is likely to insure that the right information is obtained from the right people.
- **Low cost relative to the alternatives.** The cost of obtaining administrative data for research may be low (and comparably lower than implementing a range of state-level surveys) if most or all of the information is already collected for management purposes.
- **Data on the same individual over a long period.** Administrative data are often longitudinal or can be made longitudinal through matching over time. Longitudinal data are needed to enforce and study time-limits. They are also needed to capture a family's entire history of program participation and work.
- **Data about the same conditions or program over a long period.** Data spanning many years may be required to provide a sufficiently long background prior to reforms in order to put subsequent changes in perspective. With such a background one can determine if recent changes are unusual and not just the continuation of an existing trend.
- **The ability to obtain many kinds of information through matching.** A given administrative data set can be made more useful with links to other data. Administrators and researchers need to know about multiple program use in order to see the entire picture of services received and actions of program participants.

Despite these advantages, relying primarily on information from administrative records to monitor and assess the impacts of changes in social policies presents clear challenges. While state-specific data can reflect state programs, data specific to a single state will not be useful for comparing programs across states without standardized information collection. And while administrative data provide detailed information on the characteristics of program participants, the services they have received, and the actions they have taken, they are not useful for estimating such things as the rates of program participation because they do not account for the entire population eligible or at risk for the program. Further, the reliability of the data may be compromised by “key-in” errors, or by the “creaming” of program participants, especially where positive program outcomes are a significant management concern.³ Protecting the privacy of program participants and the confidentiality of the data

³ The institution of outcomes-based accountability by many state and local governments presents a potential for introducing bias into the data, an issue that is discussed at length in Chapter 4.

when they are used for research continues to raise concerns. Chapters 2, 3 and 4 devote considerable attention to these and other challenges.

Linking information from administrative and survey data sources appears to offer the best of both worlds, and current efforts in this direction (including those by the Census Bureau to match data on firms and individuals) confirm this potential. Matching firm and individual data would permit analyses of employers' characteristics and information on employment growth along geographic and industry lines. This information would be very useful in identifying employment opportunities for welfare participants.

1.1.2 **Examples that demonstrate the potential of administrative data**

The potential research value of social assistance administrative data can already be demonstrated by the following examples, highlighted from efforts by states and outside researchers to develop and use ongoing administrative data sets (described at length, with one exception, in Chapter 5).

- **Improving Foster Care and Adoption.** The Child Welfare Research Center at the University of California, Berkeley, used administrative data on foster care from California counties to show that many children who enter foster care have multiple placements and often remain in foster care for many years, even though federal law has mandated "permanency" by 18 months. This and other research led to the "Adoption and Safe Families Act of 1997," which tries to improve the likelihood of finding permanent placements for children. In a November 1997 appearance at the University of California to announce the passage of the new law, First Lady Hillary Clinton credited the university's Child Welfare Research Center with helping to provide the research that led to the new legislation (McBroom, 1997).
- **Establishing performance measures for workforce development programs in Texas.** In March 1997, the Texas governor approved two key performance measures for workforce development programs that rely exclusively on Unemployment Insurance (UI) wage records. Adoption of these measures resulted from a series of projects at the Center for the Study of Human Resources at the University of Texas (CSHR) and Texas State Occupational Information Coordinating Committee (SOICC) over the past 10 years (Gula and King 1989). The measures were also adopted based on CSHR's recent demonstration (using program data from each of these programs and UI wage data) that calculation of such measures could actually be achieved (Texas SOICC 1997).
- **Learning how many Illinois children use state human services.** The Chapin Hall Center for Children at the University of Chicago was able to answer the basic question of which human services Illinois children used and how frequently by linking administrative data from five programs: Aid to Families with Dependent Children (AFDC), Food Stamps, Medicaid, special education and foster care (Goerge, Sommer, Lee, and Harris 1995). The findings showed that in June 1990, 25 percent of Illinois's child population was using at least one of the health and human service programs. By 1995, the percentage had increased to 27 percent. The finding that one in four children in Illinois used at least one major state human service at a given point in time calls into question any notion that only a small minority of users consume the majority of human services in Illinois.
- **Projecting time limits in Maryland.** Researchers at the University of Maryland, School of Social Work, used a multiyear administrative database from the Maryland Department of Human Resources to look at the five-year limit imposed by TANF and provided a preliminary answer to the question confronting all states: who is at greatest risk of reaching the five-year lifetime limit? (Caudill and Born 1997). Continuing research on Maryland's welfare population that draws heavily on administrative data,

wage/employment files, and interviews will track a random sample of 2,000+ Maryland families that exit welfare during the first 12 months (from October 1996 through September 1997) of welfare reform. An interim report, “Life After Welfare” was released in September 1997 (University of Maryland SSW).

1.1.3 **The use of administrative data outside social assistance**

While this report is about the use of administrative data from the AFDC/TANF, Food Stamp, Medicaid, foster care and related programs, the importance of administrative data in other areas has not been discounted. Administrative data have been widely used in many areas outside social assistance, attesting to their utility for research and program improvement. The use of administrative data in other areas also indicates that analytical difficulties such as data quality issues and matching, as well as political and administrative problems, such as privacy and confidentiality, can be overcome. Administrative data have been successfully used for research in many areas.

- **Research using vital statistics is probably the longest standing research using administrative data.** Information from birth certificate records has been used by demographers to analyze trends in birth rates and by public health specialists to develop indicators of infant health status, such as low birth weight. For some time now, information from death certificates has been used to track alternative causes of mortality. Local and state public health departments use administratively gathered data on disease incidence to form surveillance systems for infectious and communicable diseases. These uses of vital statistics data were greatly facilitated by states’ moves toward a common set of items in these data, which accelerated toward the end of the 19th century.
- **Medical and health policy research** has made extensive use of administrative data since the 1970s. Beginning in 1974, the U.S. Department of Health, Education, and Welfare mandated the submission of a uniform set of data items from all acute hospital discharges paid through Medicare and Medicaid. Data from the Medicare system have been used to determine the costs of treating certain types of patients and the outcomes of treatment.
- **Unemployment Insurance (UI) claims data and wage records** have been used by managers and researchers to gauge employment service performance, evaluate education and training programs and reforms of unemployment insurance systems, and to study displaced workers. Many of these efforts have used data from several states collected in a comparable longitudinal format under the U.S. Department of Labor’s Continuous Wage and Benefit History project of the late 1970s and early 1980s. The Bureau of Labor Statistics recently investigated the possibility of creating a national wage record database by assembling state UI wage records.

The more advanced use of administrative data in areas other than social assistance suggests opportunities in this area for improving the data for better program administration, policy design, and research.

1.2 **The Purpose Of This Report**

The report of the Advisory Panel on the Research Uses of Administrative Data informs a variety of audiences—including policymakers, agency officials, program managers and policy and academic researchers—about the present utility of administrative databases in policy-relevant research. Perhaps more important, the report identifies the issues central to improving the data’s usefulness. Specifically, the Advisory Panel has demonstrated the need for:

- **Ongoing and sustainable administrative databases** that will make possible more sophisticated policy and program monitoring and evaluation research, as well as long-term, analytic research that will inform the development of the antipoverty and welfare policies and programs of the future.
- **Mutual investment** by program managers, policymakers, and researchers in the process of program monitoring and analysis.

The Advisory Panel spent the last year gathering and synthesizing information from a variety of sources. This report describes what was found:

Chapter 2—Developing Administrative Data for Research: Definitions and Procedures. This chapter walks the reader through the process of developing administrative data for research. This process includes:

- Gaining access to the data;
- Determining the quality of the data;
- Identifying the methodological concerns that arise when using administrative data in research; and
- Addressing the larger political context in which the decision to make administrative data available for research is made.

Chapter 3—Developing Administrative Data for Research: Addressing Confidentiality and Privacy. The legal and ethical obligations to protect the privacy of agency clients and to maintain confidentiality of client data are considered in this chapter. We define what is meant by confidentiality and privacy; identify some of the principles that have been developed to guide data gathering organizations in safeguarding the privacy of information; and present new considerations for future protection of privacy and confidentiality in the face of advanced computer technologies and the mixed experience of states in legislating protections of privacy.

Chapter 4—Assessing the Relative Strengths of Administrative and Survey Data for Research. This chapter discusses the relative strengths of administrative data compared with survey data to perform three basic research activities: generating indicators for description and monitoring; measuring specific program outcomes (often in the context of performance-based management); and performing research to evaluate program effects.

Chapter 5—Examples of Ongoing Research Capacity. This chapter chronicles case histories of efforts in different states that have developed the capacity to provide administrative data for research on an ongoing basis.

Chapter 6—Developing the Research Potential of Administrative Data. This chapter summarizes the Advisory Panel’s findings and offers a set of recommendations for the further development of the research potential of administrative data.

It is appropriate to conclude this section with some comments on what this report does and does not deliver. First, for those who have produced or used administrative data in the past, some of the discussions and findings will not be new. But for those who are unfamiliar with using administrative data for research, or are more familiar with survey data, the Advisory Panel believes the discussions and findings will be quite useful. Moreover, an effort has been made to convey the potential demonstrated by several emerging, if underpublicized, efforts to develop and use ongoing administrative databases.

Second, while the Advisory Panel makes recommendations, some may find them too modest. The panel does not, for example, urge states and the federal government to devote large sums of money to fund large-scale administrative databases, although we hope new

resources will be devoted to such efforts. Nor does the panel offer specific recommendations on standards for states to follow in an effort to promote comparable databases. Such strong stands would be premature given the current developing state of the methodologies, technologies, and experiences of using administrative databases to conduct research. The panel does recommend encouraging the ongoing “bottom-up” efforts to develop administrative databases, the existing partnerships between state agencies and academic and research centers, and the initial commitments by states to develop their information systems and administrative databases. Together these will create the laboratories in which the groundwork needed to sustain a more comprehensive set of recommendations can be created over the coming years.

CHAPTER 2

Developing Administrative Data for Research: Definitions and Procedures

The administrative records agencies possess are not suited for most research purposes in their raw state. This chapter outlines the steps involved in moving from administrative records to analytic data sets usable for program monitoring and evaluation as well as for analytic research.

- Section 2.1 describes how administrative records are stored in management information systems and how the data are “dumped” to create databases.
- Section 2.2 outlines the steps involved in acquiring access to administrative records for research.
- Section 2.3 discusses the documentation of source data and ascertaining its reliability.

Similar to the issue of protecting individual privacy and maintaining data confidentiality discussed in the following chapter, acquiring access and documentation constitute “make-or-break” issues in developing administrative data for research.

- Section 2.4 reviews the ways to link administrative databases to create analytic data sets usable for research.
- Section 2.5 addresses the larger context in which decisions about using administrative data research are made.

In considering this process, it is important to keep in mind that it is not fixed. As Stephen Freedman of Manpower Demonstration Research Corporation (MDRC) advised participants at a recent conference on the research uses of administrative data,¹ “In considering how to use administrative data for research...researchers [must] make choices and set priorities. The choices and priorities are not fixed [but] are related to the researcher’s primary goals, as well as to his or her position in the research world. They may differ whether the researcher belongs to an outside private organization, such as MDRC, or to a research department within a state or federal agency, or if the researcher is a college professor within a state university system, working in tandem with an agency within the same state” (Freedman 1997, p. 4).

2.1 Administrative Records and Management Information Systems

Starting at the most basic level, administrative records (on paper or computerized) are housed in management information systems (MIS), a way of organizing information for *program management* purposes. This distinction is essential: most of the problems that must be overcome when using administrative data for research stem from the management purpose for which the data are collected. The MIS is used to make decisions on individual cases; it supports online transactional processing, that is, the information from a transaction

¹ “Evaluating State Programs: The Effective Use of Administrative Data,” hosted by the Joint Center for Poverty Research, June 1997.

is entered, processed, and stored as data.² The data may be information from clients receiving AFDC or Food Stamps, or today, TANF benefits. Or the data may be vital statistics, such as birth, marriage and death records.

By their very nature, management information systems focus on a program and *do not extend beyond the program*. Typically, an agency will have different management information systems to support different functions. For instance, a social service agency may have one MIS for case records and another for the personnel records of caseworkers. The design of an MIS and procedures for “dumping” data can be developed so as to anticipate the creation of a database. When the data are dumped, a database representing the information stored in the MIS at that point in time is created and can be analyzed. The number of cases in the last month, for instance, can be counted.

2.2 Acquiring Agency Administrative Records

Lengthy negotiations may be necessary to procure agency data. The difficulty in reaching an agreement may depend in part on the structure of the agency, the number of data sets requested, and the privacy or confidentiality rules and regulations governing the agency. The decision to share administrative data must include the highest levels of an agency’s officials, usually from the outset. A general framework tends to be agreed on at this senior level and includes, at a minimum:

1. What information will be collected and how it will be processed;
2. What linkages will be permitted;
3. How data confidentiality will be maintained; and,
4. How the research results will be disseminated.

Memoranda of agreement (MOA) specify each party’s duties and responsibilities. Researchers and program staff negotiate database issues, including choice of database software, hardware configuration, record layouts, search capabilities, and statistical software interface. Before requesting an agency’s administrative data, researchers will want basic information about the agency’s systems and files so they can choose which records and fields to copy and file, as well as the record formats for delivery.³

Catherine Born of the University of Maryland’s School of Social Work has performed state-level welfare research using administrative data for nearly 20 years. Obtaining access to data is mostly a matter of building relationships, according to her. “Academic researchers are advised to build relationships with state agency personnel before they try to build administrative databases. One can move forward without these relationships in place, but they do have important advantages. An obvious one is that an agency’s willingness to actually give you its data is greatly enhanced when there is something more substantial in place than just a request for data files. Why should you be entrusted with this precious and potentially powerful resource? What are you going to do with it, and what are you going to say about the agency and its programs?” (Born 1997, p. x).

Maintaining the confidentiality of data and protecting the privacy of program participants must be addressed to the satisfaction of the agency providing the data. The following chapter is devoted to discussion of these issues.

2.3 Documentation of Source Data

² These are three discrete steps. Just because data are “entered” does not mean they will be processed, and particularly does not mean they will be stored. Administrative purging practices have been the researchers bane for decades and remain so.

³ Researchers will also want to begin developing a plan for matching records across systems. Inevitably, this plan will be revisited as researcher work with the data, but even a rough idea of how records will be matched will aid in making an initial request that will reduce the number and type of later changes.

Researchers using administrative data can expect to spend a great deal of time familiarizing themselves with the characteristics of source data and documenting them. Data comparability is a serious concern. Definitions of even such basic fields as race or ethnicity vary from one agency to the next, as do service or household units, definition of case and identifiers, to say nothing of what triggers data entries, status changes, etc.

MDRC's program evaluations in different localities underline the importance and complexity of the data comparability problem. A research objective might be to estimate whether a welfare-to-work program in Grand Rapids, Michigan, reduces AFDC and Food Stamp payments as much as a program in Columbus, Ohio. To make a reliable comparison, MDRC must determine whether \$200 in AFDC recorded in the welfare payment field in January 1995 in Michigan's system means the same thing as \$200 in AFDC recorded in January 1995 in Ohio's system. "That is not nearly as easy as it sounds. It requires considerable investment and—parenthetically—[comparability] will be a huge problem in research on TANF" (Freedman 1997).

On a practical level, the documentation process involves learning how each data item was defined by the source agency and compiling reference catalogues (written and on-line) documenting every item. This requires researchers and program staff to spend a lot of time reading through systems manuals and code books, and testing data sets. As Freedman of MDRC says, "Our efforts at making results comparable have been time-consuming and expensive but important and. . . beneficial to the research" (1997).

Documentation should include:

- Variable definitions;
- Value codes, including "messy" value codes;
- Original data entry rules;
- Re-code rules (when agencies redefine variables or discontinue their use) and their effective dates;
- Changes in definitions and their effective dates; and,
- Pertinent information about the service context in which data were collected (such as agency protocols for tracking clients and case openings and closings).

Administrative data usually come from systems so large and complex that no one person alone understands all they contain. All of the information listed above is not generally included in the data guides and reports of public assistance agencies, so it must be obtained by interviewing various agency staff. "Program administrators can teach you much about how data are recorded—for example, when someone is listed as the payee for a benefit or when she is not. But only systems people may understand how and when databases are updated, or under what circumstances records are archived or purged from the systems" (Freedman 1997, p. x).

As the database grows and becomes operational, and as people lacking historical association with the database begin using it, thorough documentation becomes increasingly important. The process of documenting data must be routine as new data and data sources are added, definitions of fields change, and new variables are created.

Documenting source data also provides an opportunity for assessing its completeness and reliability. It is often the case that even when most items in a database are reliable, others are inaccurate or missing. Inaccuracies may range from typographical errors, such as date inversion and misspellings, to highly systematic errors, as when an entire county uses an incorrect code for a type of service. Data are more likely to be accurate when an agency relies heavily on them. For instance, information used to establish an agency's compliance

with state and federal laws, an original function of many administrative databases, tends to be complete and reliable.

It is reasonable to expect that the reliability of administrative data will improve as program managers and researchers mutually invest in their use.⁴ Feedback from researchers provides agencies with information they need to improve specific aspects of data collection and entry. Interest in research findings increases the incentives within agencies to respond to this feedback.

The potential introduction of bias poses another significant issue in assessing the reliability of administrative data. How bias is introduced and what can be done to mitigate its effects on research results is discussed at length in Chapter 4.

2.4 Designing the Analytic Database

This section discusses the issues involved in linking data in order to track over time and to create periods of duration of service for:

- An individual;
- An individual with respect to his or her changing relationship with a program (such as head of household, principal recipient, or household or benefit unit member); or
- An individual in relation to a various set of characteristics and circumstances.

2.4.1 Tracking an individual over time

Tracking over time creates longitudinal data that are especially useful for understanding the dynamics of program participation. To track an individual over time, decisions must be made about how to follow the individual, and rules for doing so must be established and applied consistently across the data.

Many social program databases use *cases* as their basic unit of analysis. Because the case is the basic unit of analysis, information on the persons within the case is sometimes not collected in a very useful form, or not at all.⁵ Sometimes they are not identifiable within the case, or, even if they are, their relationship to other members of the case is not always clear. In California, for example, it has been very hard to identify teenage mothers “nested” within cases with their own mothers or relatives because it is impossible to distinguish a case where a baby is the biological offspring of the teenager’s mother from a case where a baby is the biological offspring of the teenager. Further, cases dissolve and reform in new ways as adults marry and divorce.

What should be done when a case splits up? One could decide to follow the youngest person in the case, on the grounds that the youngest child is most likely to continue receiving assistance. Other rules are also possible, but some rule must be established and followed consistently in order to link the data.

⁴ The research community is also working on formal, systematic ways to improve the quality and uniformity of administrative data. One such effort is the Mental Health Statistics Improvement Program of the National Institute of Mental Health (NIMH), which is encouraging state agencies to adopt the reporting standards for mental health statistics set forth in the NIMH’s Data Standards for Mental Health Decision Support Systems (1989). Another can be found in the 1997 California welfare legislation, which includes a requirement that the California Department of Social Services work to “standardize state and county data collections infrastructure” (see Appendix III, §11521). These efforts will promote both standardization and reliability of data, a development that will benefit both agency administration and research.

⁵ These anomalies often make it very hard to follow individual people, and especially children. Indeed, with regard to children, there is a basic paradox embedded in many of these systems and in the new welfare reform itself. The bulk of social service attention is placed on adults, and consequently the databases are designed in such a way that they track adults much more readily than children. To examine outcomes for children, then, data systems must be designed that allow children to be followed and the outcomes of programs for them to be measured.

Understanding the way files are updated is crucial because those updates directly bear on the data quality. Cases may disappear at some calendar date because bureaucratic routine calls for cleaning out discontinued cases at that time. The correction of clerical errors in the updating process may lead to changes in identifiers, or cases may be assigned different case numbers from one spell of welfare receipt to the next. Researchers also have to be alert to what might happen when a program converts from an old management information system to a new one: is all the old history “wiped out?” Is there some cut-off date for “old” data being retained in the new system? Are all cases converted with a “service” or some other relevant date that is set to the system conversion date? Do old, nonconverted data get saved?

2.4.1.a □ DURATION OF SERVICE

Longitudinal data make it possible to analyze the spells or periods during which individuals participate in programs. The event dates and personal identifiers contained in administrative data provide the necessary elements for converting event data into episode data. Roughly, this conversion is accomplished by identifying all event records associated with a particular person, sorting them into chronological order, and creating a new data record for each between-event interval.

Records constructed in this way are entirely new analytic entities in which the time interval replaces the event as the unit of analysis. New records may be created to represent significant periods in the person’s service history and to express other facts that may be inferred from given data. The new variables, or *summary records*, are produced by manipulating or combining existing variables. The most important summary records for many purposes are records expressing time or duration, such as the duration of an individual’s eligibility for TANF or duration of a child’s foster care placements.

Summary records allow for the kind of simplified statistical analyses familiar to social scientists but less frequently applied in policy research. For example, summary records may be used to analyze patterns of service provision and utilization within a single agency, or to assess the relationships of use among several agencies. Chapin Hall researchers have used summary records to study patterns of duration and re-entry in foster care (Goerge 1990; Wulczyn and Goerge 1991). Summary records may also be used to analyze the timing of service interventions and long-term patterns of service use.

2.4.2 □ Linking an individual across programs

Linking across programs or data sets can also greatly increase the analytic power of a database, but it requires linking of identification data that might be recorded in quite different ways. As we have seen, linking the records of individual clients across agencies is complicated by the fact that no single variable can be relied on to establish the identity of a person from the records of various agencies. Though each client receiving a service is typically given an identification number (ID) unique to a particular program, each agency and department often uses its own system of identification numbers. Indeed, a single agency may issue a single client more than one ID because IDs may be assigned each time a case is opened or a child or family receives services. Other variables that might be used to establish an “all-or-nothing” match are equally problematic: even names and birth dates that “match perfectly” may refer to two different individuals, as a result of incorrectly entered data or other human error.

2.4.2.a □ PROBABILISTIC RECORD-MATCHING

The most reliable means of matching records proves to be a process called probabilistic record-matching, first developed by researchers in the fields of demography and epidemiology (Newcombe 1988; Winkler 1988; Jaro 1985, 1989; Baldwin, Acheson, and Graham 1987). Probabilistic record-matching is based on the assumption that no single match between variables common to the source databases will identify a person with complete reliability. Instead, probabilistic record-matching calculates the probability that two records belong to the same person using multiple pieces of identifying information. Such

identifying data may include name, birth date, gender, race/ethnicity, county of residence, and possibly Social Security numbers.⁶ When multiple pieces of identifying information from two databases are comparable, the probability of a correct match is increased.

Once a match has been determined, a unique number is assigned to the matched records so that each record can be uniquely identified. The end result of matching is a new file—a “link-file” that contains the unique number assigned during matching, the individual’s identifying data (e.g., name, birth date, race/ethnicity, gender, and county of residence), and all the identification numbers assigned by agencies from which the person received service. For example, if Janie Smith has been a foster child and received mental health services, the new file will contain her new unique number, her foster care and mental health ID numbers, and her name, birth date, race/ethnicity, gender, and county of residence. In the aggregate, link-files serve to establish the relationships among data in source databases and provide a means of retrieving groups of records that meet specific criteria.

2.4.3 **An individual in relation to characteristics and circumstances**

The sophisticated monitoring of emerging programs is likely to require states to identify all cases in which, for example, the head of the household:

- 18 years or older;
- not enrolled in a job-training program;
- has not worked an average of 20 hours or more in the last month;
- has already accumulated 24 months of welfare assistance; and,
- has no condition exempting him or her from the time limit (e.g., disability).

For states to monitor their programs, they must look at an individual in relation to a number of characteristics and circumstances simultaneously.

2.4.3.a **RELATIONAL DATABASES**

Relational databases have created one of the most powerful tools for organizing and accessing data. They are essentially linked tables of information on different units of analysis. A relational database can produce a table of individual characteristics such as age, sex, race, and marital status, listing every individual in the database; a table of family relationships, showing how these people are related to one another; a table of household characteristics, such as family size and date of first service receipt for every case in the data; a table listing all instances of a particular event along with the characteristics of the event, such as receiving a TANF check, enrollment in a job-training program, and perhaps receiving wages for an average of 20 hours or more per week in the last month. Within the relational database, inquiries with many qualifiers are very simple to construct, making it possible to “cut” the data in many different ways without an excessive amount of programming each time a different “slice” is needed.

For example, the database of investigated child abuse and neglect reports in Illinois maintains three separate record types (for investigations, caretakers, and children). As a result, there are at least three records for each investigation. Each record type contains summary information that is also contained in the others, and a new set of records is created for each investigation, even if they involve the same child or caretaker. The resulting data set not only contains substantial duplication, but its structure does not clearly reflect the

⁶ In the coming years, the Social Security Number (SSN) may prove somewhat more useful as a unique identifier of all clients. New rules of the U.S. Internal Revenue Service require that each claimed dependent be assigned an SSN. At best, though, the SSN is likely to become only an increasingly useful component in the process of record-matching, not a completely reliable or universal client identifier. It will be a very long time before everyone using social services will have an SSN. And because even unique identifiers may be duplicated or entered incorrectly into databases, multiple pieces of information will always be required to identify an individual conclusively.

relationships of the actors and events involved. A principal object in designing the relational database is to preserve the universe of relationships, actors, and events implied in the original data while minimizing storage costs and maximizing speed of retrieval.

Relational databases also have another feature that makes them useful for organizing data. They enforce a kind of discipline called “normalization” on how tables are constructed, which reduces ambiguity and simplifies the process of updating files. This normalization feature could prove to be especially useful in the construction of social program databases where considerable confusion results from having cases composed of persons who can move in and out of them and also form new cases.

2.5 The Larger Context: What is at Stake?

The previous discussion in this chapter has focused on the operational and procedural, as well as some of the methodological issues that must be addressed as administrative data are developed for research. What has been largely ignored is the larger context in which an agency or group of agencies decides to make their administrative data available for research. That context tends to be highly politicized for reasons that are as understandable as they are obvious. Those responsible for programs and policies are held accountable for their success or failure, generally in a very public light. Research in the forms it has been described in this report is often what determines either outcome.

“If you cannot measure it, you cannot manage it” has become a mantra for public social service managers who administer programs under a popular demand to prove their effectiveness (Barth and Needell 1997). Performance-based accountability systems have been adopted by many states (G.A.O. 1994), either to monitor the performance of particular agencies or the entire state government, and they are currently being implemented in all agencies of the federal government as a result of the Government Performance and Results Act (GPRA) of 1993 (U.S. G.A.O. 1996). Such a system is also a cornerstone of PRWORA. The act stipulates a series of outcomes that states must achieve (for example, percentages of TANF caseloads engaged in work-related activities) or a target for improvement (for example, annual rates of out-of-wedlock teenage pregnancies in the state). States must report to the federal government on these outcomes and are subject to either penalties (loss of a fraction of the state’s TANF block grant) or performance-based financial bonuses to the top performing states.

These systems are fundamentally concerned with evaluation and attribution of outcomes, issues at the core of evaluation research and causal inference. Yet, what can be measured, and how to reliably arrive at those measures, and how to compare them across states are thorny questions, often without certain answers. The problems of measurement are what researchers have long grappled with. Indeed, the standards and procedures of good research practice are what lend credibility to the use of research in program management. As John Bouman of the Poverty Law Project puts it, “The value of research to me, as an advocate, is its credibility in the policy debate, most of which has to do with its objective scientific respectability, but some of which also has to do with the absence of any apparent nonscientific motivation of the researcher” (1997. P. x).

Policymakers, program managers, data technicians, and outside researchers need to mutually invest in the process of procuring useful data if the potential of administrative data for the more sophisticated monitoring and continued evaluation of programs is to be realized. That mutual investment has demonstrated the most promise, in the opinion of this panel, in the development of ongoing administrative databases for research, a few of which are highlighted in Chapter 5.

CHAPTER 3

Developing Administrative Data for Research: Addressing Confidentiality and Privacy

The longstanding tension between the individual protection of privacy and the researcher's need to know has grown severe in recent years due to rapid and powerful advances in information technology combined with a general distrust of any entity, public or private, to safeguard confidentiality. The Advisory Panel found the concerns about privacy and confidentiality among the most important issues confronting—and potentially jeopardizing—the continued development of administrative data for research. This chapter considers which safeguards should be provided to ensure that the information on individuals and households contained in administrative databases remains confidential and that privacy interests of individuals are maintained.

- Section 3.1 distinguishes between individual privacy and confidentiality of data, the widely used terms in the discussion of personal information. In this report, “confidentiality” is used to mean restricting the pool of persons who have access to individually identifiable information. “Privacy” is used to refer to one’s right to set the conditions of disclosure of personal information.
- Section 3.2 reviews the federal protections that safeguard privacy and confidentiality, including the concept of “functional separation,” which ensures that individually identifiable information collected or compiled for research or statistical purposes may enter into administrative and policy decision-making only in aggregate or anonymous form.
- Section 3.3 provides an overview of state protections, drawing largely on the use of Unemployment Insurance wage record data.
- Section 3.4 discusses fair information principles and established research ethics that protect individual privacy and safeguard data confidentiality. These include “informed consent,” perhaps the most significant issue in protecting privacy.
- Section 3.5 considers how confidentiality and privacy can be safeguarded today, when advances in technologies for linking and storing electronic information systems make it possible to construct extensive databases on individuals and households without soliciting or directly involving these parties at all.

3.1 Confidentiality and Privacy

The definitions of confidentiality and privacy, the widely used terms in the discussion of protecting personal information, are not universally agreed on. In this report, “confidentiality” means restricting the pool of persons who have access to individually identifiable information. Its most important consequence is that confidential data are not made publicly available in a form that allow the data to be linked to identifiable individuals or units, such as families, institutions, companies, etc. Providing confidentiality protection for information collected by government agencies is essential in assuring individuals that the information requested of them will not be given to persons or organizations that eventually might make the data public.

“Privacy” is a concept that complements and, to some degree, underlies confidentiality. It refers to one’s right or privilege to set the conditions of disclosure of personal information, including not disclosing the information at all. Privacy concerns center on the use to which the data are to be put, regardless of whether the data are to be kept confidential. Individuals expect that the agencies and organizations asking them for their personal information will tell them how they intend to use this information and will not allow the information to be used in other ways unless they are informed about or give consent to such uses.

Assurance of confidentiality by itself does not address privacy concerns. Informing individuals that the information they give will be used for statistical or research purposes and not for any purpose that will affect their individual well-being addresses the privacy issue. The complexity of the privacy issue with regard to possible other uses of data originally collected for a stated purpose is discussed in Section 3.4.

Personal information collected from government program applicants is sensitive and often protected by legislation. Federal laws protect information collected by federal agencies or state agencies when federal funds are involved. The ability to control the use of one's information is recognized by the courts as a fundamental right of individuals that should be denied only in cases where the needs of society outweigh the individual's privacy interests, for instance when a public health concern arises (*Whalen v. Roe*, 429 U.S. 589 (1977)).

3.2 Federal Protections

Although the same level of confidentiality protection is given to all individuals applying for a particular program, exceptions are often provided to permit disclosures for law enforcement, program administration, and research uses of the information. At the federal level, the Privacy Act of 1974 (Title 5, U.S.C., Section 552a) offers a general level of protection to all information collected by federal agencies, but it does not apply to information collected by state governments or by businesses.

The Privacy Act prohibits disclosures that are not explicitly permitted by it, or authorized by the individual. The disclosures explicitly permitted are many, and include some relevant to research. For instance, disclosures may be made to the Bureau of the Census for censuses and surveys within its authority. Disclosure to the Census Bureau presumes that the information will be used only for a statistical purpose (defined as "not used...in making any determination about an identifiable individual"). This is supported by the Census statute (Title 13, USC) that not only limits the uses that may be made of the records but also makes them immune from compulsory disclosure. (*OMB 1975*) Significantly, the act also permits federal agencies the discretion to make other disclosures, not explicitly provided for in the text of the act, if the disclosure is compatible with the purpose for which the information was collected. These disclosures, called "routine uses," must be announced to the public in the Federal Register, and individuals providing information must be advised of them, but the disclosures may be made even with respect to individuals whose information has been collected before the use is announced. Some agencies have published routine uses to permit disclosure of administrative information for research purposes, often with conditions to be met prior to disclosure and conditions restricting further use.

Some federal agencies also have specific legislation that may provide additional protections. For instance, the Internal Revenue Service confidentiality statute (Title 26, U.S.C. Section 6103) sharply restricts disclosure of taxpayer information and imposes strict penalties, beyond the Privacy Act penalties, for improper disclosure.

In addressing the privacy rights of individuals as subjects of research and statistical studies in 1977, the Privacy Protection Study Commission determined that information collected and used for administrative purposes could be used for statistical purposes, but recommended that

no record or information contained therein collected or maintained for a research or statistical purpose under Federal authority or with Federal funds may be used in individually identifiable form to make any decision or take any action directly affecting the individual to whom the record pertains ... (p. ##).

This principle was labeled "functional separation," which means that individually identifiable information collected or compiled for research or statistical purposes should never be used to affect the individual case, and may enter into administrative and policy

decision-making only in aggregate or anonymous form. (For further discussion, see Section 3.4.1 on Functional Separation and Section 3.5 on Informed Consent.)

The commission went on to support disclosure of administrative records for research and statistical purposes under careful conditions in its recommendation that:

unless prohibited by Federal statute, a Federal agency may be permitted to use or disclose in individually identifiable form for a research or statistical purpose any record or information it collects or maintains without the authorization of the individual to whom such record or information pertains only when the agency:

- determines that such use or disclosure does not violate any limitations under which the record or information was collected;
- ascertains that use or disclosure in individually identifiable form is necessary to accomplish the research or statistical purpose for which use or disclosure is to be made;
- determines that the research or statistical purpose for which any disclosure is to be made is such as to warrant risk to the individual from additional exposure of the record or information;
- requires that reasonable procedures to protect the record or information from unauthorized disclosure be established and maintained by the user or recipient, including a program for removal or destruction of identifiers;
- prohibits any further use or redisclosure of the record or information in individually identifiable form without its express authorization; and,
- makes any disclosure pursuant to a written agreement with the proposed recipient which attests to all the above, and which makes the recipient subject to any sanctions applicable to agency employees (p. ##)

The National Academy of Science's Committee on National Statistics, in its 1993 report *Private Lives and Public Policies*, reaffirmed the need for a consistent set of statutes and regulations that provides protections for statistical records. The NAS report stipulated seven mandatory requirements for federal legislation governing the collection of statistical information:

- a definition for statistical data that incorporates the principle of functional separation as defined by the Privacy Protection Study Commission;
- a guarantee of confidentiality for data;
- a requirement of informed consent or informed choice when participation in a survey is voluntary;
- a requirement of strict control on data dissemination;
- a requirement to follow careful rules on disclosure limitation;
- a provision that permits data sharing for statistical purposes under controlled conditions; and,
- legal sanctions for those who violate confidentiality requirements.

With these statutory protections, a statistical agency or research organization operating under a government grant can effectively resist the demands of outside interests for the individually identifiable information.

A June 1997 order by the U.S. Office of Management and Budget recognizes the need to ensure that statistical data are protected. This Statistical Confidentiality Order offers a basic level of protection for information collected by U.S. statistical agencies and prohibits nonstatistical agencies or components within the same agency from demanding access to data collected solely for statistical purposes for their own program uses.

3.2.1 Nongovernment protections

Academic institutions, without legal protections, must also protect against efforts by others, including the courts, to obtain individual records. According to the drafters of the International Statistical Institute's (ISI) Declaration of Professional Ethics, "There is a powerful case for identifiable statistical data to be granted 'privileged' status in law so that access to them by third parties is legally blocked in the absence of permission of the responsible statistician (or his or her subjects). Even without such legal protections, however, it is the statistician's responsibility to ensure that the identities of subjects are protected."

3.3 State Protections

At the state level, federal or state laws will apply. For example, the disclosure of individual records maintained by the states to support the issuance of Food Stamps is subject to federal laws pertaining to the Department of Agriculture (Title 7, U.S.C., Section 2020). The law limits the use or disclosure of information obtained from applicant households to persons directly connected with the administration or enforcement of the provisions of the Food Stamp law, the regulations pursuant to the Food Stamp law, federal assistance programs, or federally assisted state programs. State welfare offices are subject to these limitations when deciding who can access Food Stamp records.

3.3.1 Unemployment Insurance wage record data

An example of state-level administrative data controlled by state law is the state-administered Unemployment Insurance (UI) program. Although the Employment and Training Administration within the Department of Labor establishes standards for UI programs in the states, each state has discretion to develop and administer a UI program best suited to its situation. The accounting data collected and maintained by the states are protected by state law. Federal program uses of these data are authorized by the Federal Unemployment Tax Act (Title 22, U.S.C., Section 3304), which provides for federal approval of state UI laws. Data collected for the Covered Employment and Wages Program (commonly referred to as the ES-202 program) are further subject to the confidentiality protections outlined in the Bureau of Labor Statistics' Commissioner's Confidentiality Order (2-80).¹

3.3.1.a CREATING A DISTRIBUTED WAGE DATABASE FROM STATE UI DATA

Efforts to create a Distributed Wage Record Database offer some interesting lessons for providing access to other state administrative data while protecting confidentiality and privacy. Significantly, a Department of Labor report describing the effort notes that none of the potential for such a database will be realized if wage record data are not accessible by non-UI users (ALMIS 1997). Recognizing the disparity of access laws and policies in the states, the authors find that:

Most state unemployment insurance laws include a confidentiality provision stating, in somewhat different words state-to-state, that authorities "may not publish or allow public inspection of information obtained under this section in any manner that reveals the identity of the employer except to public officials in the performance of their public duties." A similar prohibition applies to disclosure of an employee's identity (p #).

This statutory provision continues to be a subject of administrative and legal review in the states, and is likely to be a candidate for federal legislative action in the 105th Congress. Highlighted here is the urgent need for a single reliable source of information where interested parties can keep apprized of how this important screening factor, which

¹ Also relevant is the BLS Administrative Procedure Number 196.

determines who gains access to wage records for what purposes, is being handled by other states (ALMIS 1997).

3.3.1.b STATE SHARING OF UI WAGE RECORD DATA

In a recent survey of State Employment Security Agencies, the Bureau of Labor Statistics canvassed states on their current uses of UI wage record data (U.S. B.L.S. forthcoming). Preliminary findings show that all states permit the use of wage data for claims purposes and for child support enforcement. Three-quarters of the states permit uses by the Food Stamp program. Four-fifths permit special requests, including court subpoenas, fugitive location, Immigration and Naturalization Service (INS) and Veterans' Administration (VA) use. Two-thirds allow "other uses," mostly by federal government agencies. Only one-third permit disclosures of UI data to nongovernment entities such as release to mortgage companies at the request of the employee. Two-thirds of the states allow UI records to be used for job-training programs. Three-fifths permit "other educational" uses, including use by universities for research.

David Stevens, in a report for the Department of Labor's America's Labor Market Information System (ALMIS), notes that "individual State Employment Security Agencies (SESA's) continue to share information among themselves and with other parties about actions taken, or not, in response to third-party (i.e. non-UI) requests for wage record data. There is no one source of reliable up-to-date information about state practices, which change from month to month and with each annual legislative cycle" (p. x). Stevens goes on to suggest a process for collaborative formation of a uniform data sharing agreement specifying the range of acceptable third-party uses. Most important to this discussion is his review of six states that have successfully developed data-sharing laws and agreements (Stevens 1996).²

3.3.2 Other state administrative systems

Other state-level administrative systems such as vital statistics records, prisoner, parolee, and probationer records, worker's compensation records, and AFDC records have general confidentiality protections in law, but exceptions vary greatly across programs and across states. Legal exceptions permitting research and statistical uses are not uniform and are sometimes quite narrowly defined. Legislators may allow exceptions for research uses when the need has been expressly argued in drafting the bill. Often the research uses are limited to research related to the specific population being studied (e.g., students, medical patients, welfare recipients). Generally, in granting researchers access to personal information legislators are concerned with the effects of such uses on the individual.

The Illinois State Legislature approved an amendment to the Illinois Public Aid Code (305 ILCS 5/124.33 new) to provide research access to information on applications, terminations, and denials of benefits under the TANF program.³ The purpose is to make the process more open and to ensure accountability. The bill provides university researchers access to individual data for longitudinal studies of subgroups representing important sectors of the assistance population and requires cooperation of state welfare officials as *permitted* in the Public Aid Code. An important feature of the legislation is that the research involve no cost to the government.

California has passed legislation (AB 1542)⁴ that calls for the State Department of Social Services to conduct evaluations of the CalWORKS program implementation by an independent research entity. Interestingly, the legislation includes an article on Interagency Data Development and Use that calls for the department, with the cooperation of the University of California, to establish a project to link longitudinal administrative data on individuals and families who receive CalWORKS program benefits and to make the data

² See Appendix I

³ See Appendix II

⁴ Also known as the Thompson-Maddy-Ducheney-Ashburn Welfare-to-Work Act of 1997. See Appendix III

available to a university center with the capability of linking it with other appropriate data for ongoing assessment of program impacts. The confidentiality issue is addressed as follows:

The department shall ensure that information identifiable to individuals and families is removed so as to maintain strict confidentiality (p. x).

A particularly interesting application of an administrative data-linking system that can be used for performance-based funding purposes is the Florida Education and Training Placement Information Program (FETPIP). FETPIP is an interagency data collection system that obtains follow-up data on former students and others in Florida schools, including postsecondary institutions. FETPIP collects follow-up data that describes the employment, military enlistment, incarceration, public assistance participation, and continuing education experience of the participants being followed. It accomplishes its data collection by electronically linking participant files to the administrative records of other state and federal agencies.

FETPIP data are integral parts of the accountability data displays used by public schools, vocational institutions, community colleges, and universities in Florida. They will also be used to monitor the changes in welfare program participation that occur as welfare reform initiatives are implemented. In the future, FETPIP data will be used to derive measures related to placement of participants in adult general and vocational education, which will be an important element in determining the allocation of state funding based on performance.

Although individual data are treated confidentially, the linking of data from a wide ranging set of administrative files permits the use of the data for accountability purposes and performance-based funding decisions, thus affecting institutions, such as schools or training programs, participating in the system. Some of the privacy issues that such uses raise are discussed in section 3.4.2 on Informed Consent.

3.4 Fair Information Practice Principles

In addition to the legal issues, there are ethical concerns that must be considered in the research use of administrative data. Primarily these involve the fair treatment of the individual and his or her information. Several government groups and private-sector organizations have developed Fair Information Practice Principles that are voluntarily adopted by those who collect and process personal information.

In 1973, an advisory committee to the Secretary of Health, Education and Welfare formulated, in its report, "Records, Computers and the Rights of Citizens," a set of principles to govern use of personal information, which it recommended be enacted in the form of a "Code of Fair Information Practice."

The Privacy Protection Study Commission adopted the principles as a starting point for its 1977 study, "Personal Privacy in an Information Society." Later policy statements incorporated similar principles. The Organization for Economic Cooperation and Development (OECD) issued its "Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data" in 1980. The Clinton Administration's Information Infrastructure Task Force issued its report, "Principles for Providing and Using Personal Information" in 1995. Together, these serve to highlight some of the ethical considerations that information processors should consider.

Each of these sets of guidelines and principles share common threads. They:

- Promote openness;
- Provide for individual participation;
- Limit the collection of personal information;
- Encourage accurate, complete, and current information;
- Limit the use of information;

- Limit the disclosure of information;
- Ensure the information is secure; and
- Provide a mechanism for accountability.

They are general to all personal information and do not attempt to treat special issues related to public interests (law enforcement, the press, research, or statistics).

In recognizing the unique relationship between statistician and citizen, the Project Group on Data Protection of The Council of Europe (COE) (1997) has developed “Recommendations on the Protection of Personal Data Collected and Processed for Statistical Purposes” (see, [www.coe.fr/dataprotection/rec/r\(97\)18e.htm](http://www.coe.fr/dataprotection/rec/r(97)18e.htm)). These recommendations acknowledge key considerations for the fair treatment of information processed for statistical purposes. Principally, they recognize the need for statisticians to guarantee respect for rights and fundamental freedoms, in particular the right to privacy when personal data are collected and processed for statistical purposes. The COE recommendations include “general conditions for lawful collection and processing for statistical purposes.” In considering the implications of direct data collection on privacy, these general conditions stipulate that

In order to avoid collection of the same data again, personal data collected for non-statistical purposes may also be processed for statistical purposes where that is necessary for the performance of a task carried out in the public interest or in the exercise of official authority or for the purpose of the legitimate interests pursued by the controller except where such interests are overridden by the rights and fundamental freedoms of the data subject (Council of Europe, 1997, p. x).

An interesting corollary condition permits secondary statistical uses where the information was originally collected for statistical purposes:

Data collected for one statistical purpose may also be processed for other statistical purposes in the circumstances listed above (p #).

Key ethical issues for statisticians have also been outlined by statistical and research associations in ethical guidelines to their professions. The ISI *Declaration of Professional Ethics* (1986) describes “obligations to subjects” that include:

1. Avoiding undue intrusion;
2. Obtaining informed consent;
3. Modifications to informed consent;
4. Protecting the interests of subjects;
5. Maintaining confidentiality of records; and,
6. Inhibiting disclosure of identities.

3.4.1 Functional separation

Twenty years earlier, the U.S. Privacy Protection Study Commission (19xx) released *Personal Privacy in an Information Society*. The commission described the relationship between citizens and the government when the citizen is the subject of research or statistical studies, and identified a key element of the legal/ethical framework as “functional separation.”

The Commission believes both needs [protection from inadvertent exposure from an administrative action and public confidence in the integrity of research and statistical activities] will be met if the data collected and maintained for research and statistical use cannot be used or disclosed in individually identifiable form for any other purpose. *To erect such a barrier, however, there must be a clear functional*

separation between research and statistical uses and all other uses. The separation cannot be absolute in practice but the principle must be established that individually identifiable information collected or compiled for research or statistical purposes may enter into administrative and policy decision making only in aggregate or anonymous form. The reverse flow of individually identifiable information from records maintained by administrators and decision makers to researchers and statisticians can be permitted, but only on the basis of demonstrated need and under stringent safeguards (U.S. Privacy Protection Commission, 19XX, p. x).

3.4.2 **Informed consent**

Perhaps most significant to the discussion of protecting confidentiality and privacy is informed consent. The rapid advances in technologies for linking and storing electronic information systems make possible the construction of extensive databases on households and individuals. In fact, almost all of the information in these databases can be assembled without directly involving these parties at all. While this substantially streamlines and reduces the costs of information gathering, it also represents a significant loss of control over how this information is used by others.

The ISI principle states that “statistical inquiries involving the active participation of human subjects should be based as far as practicable on their freely given informed consent. . . .” However, one acceptable “modification to informed consent” is secondary use of records. The declaration states that “in cases where a statistician has been granted access to, say, administrative or medical records or other research material for a new or supplementary inquiry, the custodian’s permission to use the records should not relieve the statistician from having to consider the likely reactions, sensitivities and interests of the subjects concerned, including their entitlement to anonymity” (ISI 1986, p. x). In situations where statistical uses are envisioned for administrative records prior to their collection, the ethical guidelines would anticipate that subjects are told about these intended uses.

3.4.2.a **THE NOTICE PRINCIPLE**

Although not fully resolved as applied to the type of administrative data considered in this report, there are several principles in federal regulations and legislation that provide some initial guidance on informed consent. The first concerns what “notice” should be provided to these individuals. On this score, the Clinton Administration’s Information Infrastructure Task Force’s recently promulgated *Principles for Providing and Using Personal Information* (1995) seem germane. These principles provide that those who collect personal information directly from individuals should provide adequate, relevant information to those individuals whose information is being assembled. The disclosure information should include:

- Why they are collecting the information;
- How the information is expected to be used;
- What steps will be taken to protect the confidentiality, integrity, and quality;
- The consequences of providing or withholding the information; and,
- Any rights of redress.

This notice principle is key to protecting privacy.

3.4.2.b **THE FAIRNESS PRINCIPLE**

A companion principle, the “fairness principle,” states that information users should not use personal information in ways that are incompatible with the individual’s understanding of how it will be used, unless there is a compelling public interest for such use. The task force’s discussion of the principles notes that the fairness principle cannot be applied uniformly in every setting. An incompatible use is not necessarily a harmful use; in fact, it may be extremely beneficial to the individual and society. There are some incompatible uses

that will produce enormous benefits and have at most a trivial effect on the individual's information privacy interest. Research and statistical studies, in which information will not be used to affect the individual, are examples. Obtaining the consent of individuals to permit new statistical uses of existing data adds cost and administrative complexity to the process and risks impairing research projects. Nevertheless, the principles do not relieve the collector of the information from acknowledging the intended statistical uses for the information.

3.4.2.c INFORMED CONSENT IN STATISTICAL RESEARCH

Finally, it is important to note that "informed consent" in the context of statistics is different from informed consent in medical and scientific research. Most statistical research involves learning only about characteristics of populations, whereas medical research often has direct implications for the individual subject. As a result, *consent* is required for a clinical research study, whereas notice may suffice for a statistical study.

Administrative agencies should provide this notice at the time the information is originally collected. In the case of a welfare agency, it would generally be sufficient to provide the program applicant notice that his or her information will be provided to authorized researchers for research on the effectiveness of welfare programs in meeting planned objectives. If information about the types of research and names of researchers is known, this information should be communicated as well. According to the National Academy of Sciences *Report Private Lives and Public Policies* (1993), unplanned or unanticipated statistical uses can be covered by a statement suggesting unanticipated future uses of the data for statistical or research purposes.

A point of some dispute concerns unplanned or unanticipated statistical uses that may result in indirect effects on individuals whose data have been used. For example, research uses of administrative data may result in amendments to TANF that affect some individuals whose data were used in the analyses even though the data were released to the researcher in aggregate or anonymous form. The basic issue can be considered one of reciprocity: Should an individual who accepts social benefits be expected to voluntarily relinquish certain rights to privacy (but not confidentiality)? Some researchers answer yes, sometimes even without direct informed consent because of the selection-bias that direct consent introduces. Others believe that one should seek direct consent for uses that were not initially described to the participants. Many researchers feel that the protection of confidentiality is sufficient in most such cases. It should be noted that the federal regulation for protection of human research subjects, in setting out criteria that institutional review boards should use in ethical evaluation of the effect of research on subjects, explicitly excludes consideration of "possible long-range effects of applying knowledge gained in research (for example, the possible effects of the research on public policy)". (Federal Policy for the Protection of Human Subjects, generic sec. ____111(a)(2), 56 Fed. Reg. 28003, 28015 (1991), found in DHHS regulation at 45 CFR sec. 46.111(a)(2)).

3.5 Safeguarding Privacy and Confidentiality

Maintaining the confidentiality of administrative records involves protecting the identity of individuals when the information is published or disseminated to researchers. Confidentiality can be maintained in two ways: by restricting the *content* of the information or by restricting *access* to the information.

3.5.1 Restricting content and disclosure limitation techniques

To restrict content, an analytic database can be made available to researchers after disclosure limitation techniques are applied to mask the identity of program participants. Such techniques include suppressing identifying information such as name, address, and Social Security number, limiting the value of highly discriminating variables such as income, and grouping responses into categories. Tabular results can be protected by suppressing cells

that contain very few responses. The federal Report on *Statistical Disclosure Limitation Methodology* provides guidance on protecting confidentiality using these techniques (Statistical Policy Office 1994).

3.5.2 Restricting access

In instances where the content must include identifiers, restricting the access to confidential information provides an alternative means of releasing data to researchers. Under this approach, a researcher signs a binding agreement that extends to both access to the individually identifiable data and the penalties for unauthorized disclosure that are imposed on those employed by the agency collecting the data. The data remain confidential, and limits are placed on the use and release of the information. The researcher, under the agreement, is permitted to use the data for specific statistical studies but is prevented from disclosing it to third parties. Controls are placed on who can access the information, where access is permitted, and what techniques are used to limit disclosure in the resulting analyses.

There are various models that have been successfully used at the federal level. One model involves having the researcher use the data at the source agency under the supervision of agency staff. Another model involves establishing regional, secure centers (“safe sites”) where researchers can use the data under the supervision of a full time agency employee (or the employee of an organization that has been authorized to maintain a secure site). Another option is to license researchers to use confidential data at their own facility under specified security arrangements. A final option, currently used to permit research access to survey data for the Luxembourg Income Study, permits researchers to access confidential data remotely by computer and modem with software filtering the resulting products to ensure confidentiality. Examples of these and other access arrangements are provided by Tom Jabine in his report *Procedures for Restricted Data* (1993).

3.5.3 Remaining challenges

Clearly, the procedural and technical means of protecting confidentiality can be established. Some may even argue that we already know how to protect privacy and all that is needed to ensure its protection is the will to do so, through legislation, professional ethics, and other means. But a potential limitation to the approaches outlined above stems from the fact that these methods—and especially disclosure limitation techniques—have been formed primarily in the context of survey data on large-scale, national populations, within the context of federal statutes. Administrative data pose a very particular challenge to the established approaches of safeguarding confidentiality and privacy in that they represent very select populations at the outset. It is much more difficult to prevent disclosure of individual data in a public use file when it is known that a specific individual’s data are in the file than it is to prevent disclosure when one does not know the identity of any individual in the file, as is typically the case in public use data files based on sample surveys. The relative ease with which information can now be linked, given new technologies, means that disclosure limitation techniques must be re-examined in the context of administrative data where the names of individuals in the file are likely to be known more widely because of their known participation in a program or their receipt of some benefit.

Further, most states do not have the history of the federal government in dealing with concerns about confidentiality and privacy protection in the context of data-sharing because they simply have not the extent of experience in data collection efforts. Neither have they faced the same number or nature of requests for their data. Efforts to create a Distributed Wage Record Database by the Department of Labor as well as the BLS exploration of how states share their UI data demonstrate the variety and fluctuation of state practices with regard to data sharing, “which change from month to month and with each annual legislative cycle.” (Stevens 1996. Also, ITSC 1997). Much of what has been established by the federal government may apply to the states, but because these examples suggest

considerable differences regarding protections of confidentiality and privacy, each state will want to examine federal models in light of its own circumstances.

Finally, the argument that what is needed most is the will to protect confidentiality and privacy may be sufficient to the problem at hand. Researchers must continue to place the vigilant protection of the confidentiality of individual data and of the individual's privacy at a premium. In the context of public apprehension, any failure to safeguard confidentiality adequately puts at risk the use of administrative data for research.

CHAPTER 4

Assessing the Relative Strengths of Administrative and Survey Data for Research

Survey data have been a mainstay in program evaluation, policy analysis, and government monitoring efforts for decades. National surveys such as the U.S. Censuses of Population, the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), and more specialized health-related surveys have been and will continue to be used to monitor performance and evaluate the impact of public assistance and public health programs and policies.¹ But, as previously noted, the advent of computerized administrative records housed in modern information systems has made it possible to extract information, maintain data quality through computerized checks, and link data across systems and over time more easily and less expensively than even a few years ago. Administrative databases, along with data from birth, death, and marriage registries, will become increasingly important components of the nation's data collection "system" for conducting ever-widening types of research.

This chapter provides an assessment of the strength of administrative data relative to survey data for conducting various types of policy-relevant research and for monitoring government programs and policies.

- **Section 4.1** compares key dimensions of administrative data and survey data, including: the populations represented; obtaining outcomes measures and background variables and their quality; the time frames for which information is gathered; and obtaining information on program parameters and context.
- **Section 4.2** assesses the strengths and limitations of administrative data and survey data in alternative types of research, including: descriptive research and trend analysis, causal inference, and evaluation of the impact of social programs and policies. For instance:
 - **Developing** databases from administrative records makes it possible to amass large samples of program participants or individuals who have experienced events monitored by government (birth, marriage, death, disease occurrence, etc.). In contrast, the sample sizes that can be sustained for data gathered via survey interviews, especially for low-income and disadvantaged populations, are typically much smaller, given the effort and cost involved in locating and contacting respondents.
 - **Administrative data** provide more detailed and usually more accurate measures of programmatic outcomes and status than can be obtained via surveys, while surveys represent a more flexible method for gathering information, enabling one to obtain samples that are representative of general populations (difficult to represent with information from most administrative data sources) and to provide a wider array of variables that are unlikely to be found in administrative databases, such as health status or parenting skills.
- **Section 4.3** considers the use of administrative data for monitoring a program's performance to determine whether it is meeting its objectives or mandates. Though this is largely viewed as a management issue rather than a research concern, the criteria for judging the suitability of data sources for conducting research apply to outcomes- or performance-based accountability.

¹ As noted in Chapter 1, examples include the National Health Interview Survey (NHIS), Behavioral Risks Factors Surveillance System (BRFSS), and AFDC Quality Control (QC) Surveys.

- Section 4.4 examines the potential of linking administrative data and survey data for research of social programs and policies.

4.1 A Comparison of Key Dimensions of Administrative Data and Survey Data

To compare the relative strengths and limitations of administrative with survey data, their similarity and differences should be examined with respect to:

- The populations they sample (or cover);
- The types of outcome and “background” variables they measure and the quality of these measurements;
- The “contextual” information, i.e., information on characteristics of social programs and conditions (e.g., labor markets, service access, etc.) contained in these data sources for the population represented in each source; and,
- The time frames for which information is available in each data source (i.e., the extent of longitudinal data on units).

Much of the discussion is focused on data sources relevant for such public assistance programs as the former AFDC program, Food Stamps, Medicaid, and the new state programs under TANF.

4.1.1 Populations represented

A key difference between data gathered from administrative records and survey data concerns the populations for which representative information can be obtained. In general, information in administrative data characterizes, or is representative of, the universe of individuals or households who experienced some event (such as birth, marriage, or arrest) or some particular transaction (such as entry into a program or system, or receipt of payment or service from a public assistance program). In the language of the statistical sampling literature, such data are an example of a choice-based sample.²

Examples of choice-based samples include information for the entire caseload for a state’s public assistance programs under TANF, or children (and families) with open cases in a state’s foster care system. While databases for governmental programs are representative of populations involved in these programs, they are not representative of populations potentially eligible for them, or the populations targeted for public assistance, such as families with children living in poverty. Administrative databases are ideally suited for obtaining information about individuals or households who participate in certain programs. However, they generally are not sufficient to estimate such things as the rates of program participation because the data do not allow for estimating the size of the relevant populations. Administrative data only provide information on program participants (the numerator) but not the universe, or population, of those who are “at risk” to participate (the denominator). This limitation is often referred to as the “denominator problem.”

In contrast, the population for which surveys are representative is determined by the sampling frame used to draw a particular survey. Some surveys, such as the CPS or SIPP, are designed to represent the civilian non-institutional population residing in the U.S. These surveys are referred to as “area probability surveys,” in that they sample households or addresses, from areas and localities within the U.S., and typically draw samples in proportion to the population in those areas. Other surveys, such as the Quality Control surveys under the former AFDC program, were designed to be representative of a state’s AFDC caseload at a particular point in time. Samples that are drawn from populations with

² Other examples of choice-based sample include marketing data on individuals who bought a specific product or shopped at a particular store, or data on individuals who used a certain mode of transportation, such as riding the subway to commute to work.

modern sampling methods provide researchers with greater latitude with respect to the populations that can be covered by a particular data source. Moreover, these techniques enable one to control, *ex ante*, the representativeness of the resulting sample. In contrast, it is often difficult, if not impossible, to cover many populations of interest by using the units in an administrative database as a sampling frame beyond the caseload of a particular program at a particular point in time.

While the survey's use of statistical sampling methods provides greater flexibility with respect to the populations that are analyzed, this greater degree of control comes at a cost. Conducting a representative survey, whether for a general or more targeted population, requires the following steps:

- a. development of a sample frame (typically based on the primary sampling units (PSU) from U.S. Census data);
- b. obtaining lists of the households (or individuals) in the frames;
- c. drawing samples at random from the listings;
- d. locating individuals in the samples, either in-person or by telephone; and,
- e. gathering information by conducting an interview to obtain respondents' answers to questions.

Each step carries with it a set of costs that must be borne before information is available for any purpose. The size of both the "fixed" costs of finding subjects and the more "variable" costs of gathering pieces of information associated with surveys generally means that survey data are gathered for only a random subset, or sample.

Administrative data, on the other hand, in principle can yield information on an entire population, so long as it is "captured" and recorded by some administrative process or procedure. Moreover, the costs associated with identifying such populations can be very low, to the extent that locating them is a direct by-product of the administration of a program or existing data-gathering effort (for example, recording births or marriages).

There is a trade-off with respect to obtaining samples via the (implicit) data-gathering methods of administrative data and those used for conducting most surveys. Samples obtained from administrative records limit one to more restrictive and selective populations compared with gathering data by means of a survey. At the same time, the differential costs of obtaining data tend to limit the sample sizes obtained with surveys compared with those, in principle, associated with administrative records.

4.1.2 Obtaining outcome measures and background variables and their quality

Just as the populations represented in administrative data sources tend to be selective, so too is the information that can be gleaned from them. The kinds and quality of measures and variables that can be extracted from administrative records are limited by the primary purpose for which these records exist. Most administrative data are derived from program information systems used to manage and monitor ongoing programs.³ The records and information they contain are designed primarily to serve the business function of running a program, not a research function. Therefore, data from administrative records will measure such things as which household members were eligible to receive some set of benefits or services, which benefits and services they received, and when they received them.

Because of its "business" purpose, the information from administrative records on programmatic transactions is likely to be very accurate, especially compared with information gathered by survey interviews. When asked in a survey about details concerning program benefits or service receipt, respondents often have difficulty recalling exact information; they may even be reluctant to reveal their program participation or its

³ The exception to this are data from vital statistics and disease surveillance systems. The information in these systems is gathered primarily for record-keeping and monitoring purposes, and, as such, what is gathered and how it is gathered is often governed by its research uses.

extent because such participation is viewed as undesirable by many quarters of society. In a recent paper, Brady and Luks (1995) compared the survey responses of individuals who had received (or were still receiving) public assistance in California in the 1990s with information on the spells of welfare receipt contained in county public aid records; they found that respondents tended to report shorter spells than recorded in administrative data.

Typically, the data available in administrative records on the personal, demographic, and background characteristics of individuals or households are either directly relevant for determining program eligibility and benefits or they are of secondary importance. For example, a program may keep track of a household's income, the family structure (numbers, relationships, and ages of family members), and place of residence because this information is needed to determine the benefits, levels of subsidies, or services to which the household is entitled. Because this information is collected for the business purpose of determining benefit levels, the resulting measures based on the data may be biased in certain systematic ways. A recent study by Hill, Hotz, Mullin, and Scholz (1997) found that adults in AFDC assistance units in California appeared to systematically underreport their earned income to public assistance offices compared with the level of earnings reported by employers to the state's UI system or compared with what they claimed as earned income on their federal tax returns. Such underreporting is hardly surprising; these individuals had a strong incentive to underreport income given the AFDC eligibility rules and the treatment of income in benefit determination.

Data on individuals or households that are not directly relevant to the business needs of the program either tend not to be kept at all, or if recorded, tend to be inaccurate or out of date. For instance, the records kept on individuals by states for purposes of administering their unemployment insurance programs do not contain information on such things as a person's educational attainment, ethnicity, or race. Most state public assistance files do not keep accurate information on those household members who are not a part of the "assistance unit" relevant for that program, let alone their characteristics.

Administrative data from one program seldom contain enough information to make a useful program evaluation possible. By linking administrative data from different programs, it is possible to obtain an array of explanatory and outcome variables. State unemployment insurance programs collect data on quarterly wages of covered employees, and these data can be linked to welfare eligibility data and job training data to determine whether job training helps move people off welfare and into remunerative jobs.

More recently, several states and organizations have begun to develop and maintain archives of linked databases for populations drawn from program caseloads. The California Children's Services Archive developed by the Child Welfare Research Center at the University of California, Berkeley, and the multistate foster care data archive developed by researchers at the Chapin Hall Center for Children at the University of Chicago contain a rich set of data for children in state foster care or child welfare systems. These databases were developed by linking administrative records from other public assistance programs, such as AFDC programs, as well as from vital statistics data, which contain information on a child's birth weight and the marital status or age of parents at the child's birth.⁴ Linking information across program information systems and registries can greatly enrich the "variable set" for the associated populations. However, the reliable identification of an individual across systems to create a "link-file" for that person entails a great deal of effort.⁵

Gathering outcome and background information through survey methods presents a different set of advantages and difficulties. One advantage is the ability to obtain more than the business-relevant information on the individuals and households represented in an administrative database. Having incurred the costs of locating these individuals, the survey enables one to gather a much broader range of information, including the outcome and

⁴ For further discussion of both archiving efforts, see Chapter 5.

⁵ See Chapter 2, Section 2.4 "Designing the Analytic Database."

background information not directly used in the administration of a program. One can also obtain more detailed information on outcomes, such as indicators of quality of life, the adequacy of parenting, health care, nutrition, or housing, which provide a more complete characterization of a person's situation. Surveys also make it possible to obtain follow-up information on closed cases, or for periods between administrative events.

There are, however, limits to what surveys can achieve in gathering accurate outcome and background data. Individuals must be located before they can be asked any questions, and they must agree to participate. Even if they agree to participate, they still may refuse to answer certain questions. It is well known that respondents to surveys often refuse to disclose any information about their income, for instance. And even when participants do respond to questions, their responses may be subject to recall error. The head of a household may not remember the amount of the benefits or particular services received in the preceding month. While a number of different survey design strategies make it possible to minimize refusals, item nonresponse, and recall errors, each is a potentially important shortcoming.

4.1.3 Time frames for which information is gathered

The time frames covered by administrative and survey data and the timing for collecting information present different opportunities and concerns for the study of program participation dynamics, including rates of exit from programs like AFDC, TANF and Medicaid, which is facilitated by longitudinal data on individuals or households. Administrative data have the potential to provide a wealth of highly accurate, longitudinal data on program participation and service receipt because administrative systems collect data as part of the ongoing administrative process of a program. The time frame that can be covered by such data corresponds, in principle, to periods of being enrolled in a program. Within that time frame, the timing of gathering information tends to coincide with program transactions. Thus, for public assistance programs like AFDC or TANF, information is collected monthly, corresponding to the timing of benefit determinations. Alternatively, for service provision programs, data are gathered when a service is requested or provided. At these transaction points, one can learn about an assistance unit's participation status, the benefits or services received, as well as information relevant to determining a unit's eligibility for services, such as earned income, family composition, or employment status. Because information is obtained at the time transactions occur, it is less likely to suffer from the recall errors often associated with survey data-gathering.

At the same time, the business purposes that dictate the collection of information for ongoing programs once again imply certain shortcomings. Important changes in the case outside the purview of the system can remain invisible because administrative data do not provide any information on the status of individuals or households when they are not enrolled in a program or during the intervals between transactions. The consequences of these limitations vary with what one wishes to measure. If the interest is in analyzing the time recipients spend enrolled in a program, the frequency with which they receive program services, or the amounts or types of services or benefits they receive, the time frames and timing of measurement pose no obstacle. But, if the interest is in examining the timing of non-program activities, such as employment status or childbearing over a person's life, administrative data are less useful because the occurrence of such events will not be measured during periods when a household is "off the program."

Turning to surveys to gather longitudinal information also has its advantages and disadvantages. A key disadvantage is, again, the cost of gathering such information. Unlike administrative data, where the marginal cost of gathering a piece of longitudinal information is low because it is gathered as a by-product of program administration, the corresponding marginal costs for obtaining longitudinal information with surveys is that of locating the individual or household and conducting an interview. Such costs typically dictate that the intervals between interviews are longer than those between program transactions and, as

such, are likely to introduce problems of recall as one gathers information on more distant events.

However, if one is willing to bear these costs, several benefits can be derived from gathering longitudinal information from surveys. The most obvious advantage is that surveys do not require a transaction to occur or a person to participate in a program in order to obtain information. Individuals or households can be followed and interviewed over time, whether they continue to participate in a program or, for that matter, have never participated in a program. Longitudinal surveys that sample more general populations, such as the SIPP, can be used to monitor both program entry and exit. Administrative records from a program's caseloads, in contrast, can be used to analyze the timing of exits but are of limited value for studying entry behavior because of the restrictive nature of the population that is covered by such data, namely those who entered a program some time in the past. The use of surveys also allows one to obtain longitudinal measures of events or outcomes irrelevant to the business purposes of a program, information that generally is not found in administrative systems or, when it is, tends to be inaccurate and out of date.

4.1.4 Obtaining information on program parameters and context

To understand the effects of and trends in social programs for policy purposes, data are not only needed on the behavioral and program outcomes of individuals or households, but also on:

- The rules, or parameters, of state and local programs;
- The processes and procedures that characterize how they operate; and,
- The environment, or contexts, in which programs function and individuals reside.

Examples of program parameters include:

- The maximum payments and services—such as child care, basic or job training—a state's welfare program provides;
- The “replacement ratio” for income under the state's Unemployment Insurance program; and,
- Medical procedures, such as abortions or CAT scans, covered by a state's Medicaid program.

Measures of the rules and procedures by which a program's bureaucracy operates might include:

- Measures of how time-consuming an agency's intake procedures are;
- The extent to which an agency's service delivery is coordinated or fragmented; and,
- Indicators of the quality of a program's professional staff.

Finally, examples of local conditions, or contextual factors, would include:

- Local unemployment rates;
- The types of jobs and industries available in a locality;
- The quality, cost, and availability of housing; and,
- Local crime rates.

Data on program parameters and contextual variables are crucial for policy-relevant research. Measures of program parameters characterize the “treatments” to which “subjects,” namely households or individuals in a particular locale, are exposed. Measuring local context or conditions is an important component for conducting program evaluations because, as discussed below, failure to control for exogenous factors when analyzing program effects is likely to result in “omitted variable” or “selection” bias. Some of the same problems that arise from using data from administrative records and information systems arise in developing measures of the program parameters and contextual variables.

Technically speaking, the rules or parameters of programs are determined by legislation or by the agencies that administer particular programs. The information necessary for constructing measures of some program parameters can sometimes be obtained from the statutes or agency regulations. But such measures may not always be appropriate, given potential differences between stated rules and what is actually implemented. Better measures of the effective rules faced by program participants often can be deduced by monitoring an agency's caseload and what services or benefits its members receive. In some instances, this monitoring amounts to examining the information contained in the administrative records of a program in a particular locality.

Studies of the effects of state AFDC programs on participant behavior have constructed measures of effective maximum benefit levels and benefit reduction rates based on payments by members of state AFDC caseloads.⁶ Alternatively, information from administrative records has also been used to construct measures of, for example, the average waiting times for receipt of services by program recipients. Finally, measures of program administrative practices and effective procedures often have been obtained by conducting process analyses, in which observers monitor the actual operation of agencies or interview key informants about agency practices.

Constructing measures of local context also often involves administrative or survey data. Measures of local unemployment and employment rates can be constructed using data from state Unemployment Insurance records and claims. Alternatively, data from the battery of questions on employment and job search in the Current Population Survey (CPS) are used to form state-level unemployment rate measures. Indices of crime can be constructed from court and law enforcement records on criminal activity, as well as from responses on crime and victimization surveys.

Although the construction of both program and context variables make use of administrative and survey data, the resulting measures are aggregates that index program structure and local conditions where the "units of analysis" are jurisdictions, states, counties, cities, or administrative regions, over which programs and location conditions differ. This raises implications in the use of administrative data for evaluation research, discussed further below.

In conducting evaluation research of social programs, especially those with nonexperimental evaluation designs, it is useful to link jurisdiction- or location-specific measures of program parameters and context variables with outcome data for individuals and households. A key issue in performing such links, whether with individual- or household-level survey or administrative data, is the adequacy of the information on the place of residence in either of these data sources. Typically, administrative records contain the addresses of program participants so that such matches are feasible and, in general, such information is available for linking with survey data. The resulting data sets, however, contain substantial amounts of information that could be used to determine the identity of individuals or households. Therefore, making such data sets publicly available can compromise confidentiality, and a key practical issue that arises is controlling and limiting access to these linked data sets.

4.2 The Strengths and Weaknesses of Administrative and Survey Data in Alternative Types of Research

The differences between administrative and survey data regarding the types of populations they can cover, and the information that can be obtained from each, have implications for their utility in different types of research. In the remainder of this chapter, some of these implications are discussed for three types of research: (a) descriptive and trends analyses,

⁶ See Fraker, Moffitt and Wolf (1985) for more on the methods for deriving such effective measures from audit data of state samples of AFDC administrative records.

(b) monitoring program performance, and (c) evaluation of the (causal) impact of social programs and policies.

4.2.1 **Descriptive research and trend analysis**

One of the most fundamental and important purposes for gathering data is to describe what has happened and to project what is likely to happen in the future. Data from administrative records have long provided much of the information used for the descriptive and trend analyses presented in annual reports issued by state agencies or compiled in compendia such as the *Green Book*, issued annually by the Ways and Means Committee of the U.S. House of Representatives. Because the information in administrative data is event-based, it is well-suited for deriving estimates of the size, composition, and trends in program caseloads.

However, information from administrative data is typically insufficient to estimate rates of program participation because it only provides information on program participants and not the entire population of those “at risk” to participate (the “denominator problem,” noted above). Although one can use administrative data from a state’s Medicaid program to determine the *number* of children who are covered at a point in time, one must obtain an estimate (or count) of the total number of children in a state in order to determine the *proportion* of children covered. Data for the total number of children, the denominator of such a rate, must come from either survey or census data. Administrative data are also of little use in determining either the number or proportion of a state’s child population that would be eligible for a program like Medicaid. Estimating either levels or rates of program eligibility requires not only data that are representative of a state’s population, but also detailed information on those variables, such as income and household headship, on which eligibility is based. Thus, while administrative data can be an important building block for deriving many statistics used to describe program trends, it typically must be supplemented with data that are representative of more inclusive populations.

Another increasingly important set of statistics used to analyze public assistance programs is the lengths of time receiving aid and the speed, or *rate*, at which participants exit the welfare rolls. Since the influential work by Mary Jo Bane and David Ellwood (1983) on duration of welfare spells, there has been growing attention in monitoring the extent to which a program’s caseload is “dependent” on welfare, i.e., spends long periods of time on a program. Trends in such duration statistics are sure to become even more closely watched under welfare reform; in fact, reporting of such is required by states under PRWORA as the federal government and states seek to reduce time on aid and increase the rate of transition from welfare to work. The latter is an especially important phenomenon to monitor, given the time limits placed on a household’s receipt under TANF. Data from administrative records are ideally suited to generate such duration statistics, including the distribution of “time left” for assistance, among different demographic groups or regions of a state.

At the same time, administrative data are not very well suited for analyzing the rates at which households do or do not enter, or enroll in, state programs under TANF. Again, this is because administrative data only contain information on those individuals or households who receive public assistance; they provide no information on persons who have yet to enter the welfare rolls. Although PRWORA does not require states to report on program entry rates, these statistics will be important for monitoring an implicit, if unstated, objective of welfare reform. As Robert Moffitt has argued, states have become increasingly interested in reducing the likelihood that a household (or individual) ever goes on welfare (1992 and 1993). Estimating program entry rates requires data on populations that are at risk to enter a program, and this information typically must be gathered with surveys. Moreover, such surveys must either be longitudinal, following individuals for long periods of time, or must rely on retrospective questions to determine a person’s past participation in such programs.

4.2.2 **Data for use in causal inferences and evaluation of the impact of social programs and policies**

The objective of program evaluation research is to find out what “works” and what does not. More precisely, evaluation research estimates the effects of programs, as measured by their parameters, on the subsequent outcomes (including rates of employment, labor market earnings, continued dependence on public assistance, and so forth) of individuals and households on and off program caseloads. At a minimum, conducting program evaluations obviously requires data on parameters and outcomes, where, as with more descriptive types of research, larger samples and higher quality data improve the quality of research. But, unlike descriptive research, evaluation analysis must devise strategies to isolate the causal effects of program “treatments”; in other words, ways of separating the influence of programs on outcomes from that of other, confounding factors. In technical terms, evaluation research requires strategies, or methods, for minimizing the intrusion of selection bias. There are a variety of strategies one might use to deal with selection bias, and they differ with respect to the amount and types of data required for their “success.” The remainder of this section provides an assessment of the ways in which administrative and survey data facilitate or detract from a variety of evaluation methods often used to estimate program effects.

Evaluation methods fall into two categories: *experimental designs* and the use of *non-experimental*, statistical adjustments. Over the last 20 years, an increasing number of evaluations of social programs have used experimental designs. In such studies, “subjects”—individuals or households who are eligible for or enrolled in a program at some point in time—are randomly assigned to different program treatments (different services or service provision strategies). Subjects who receive these new treatments under study are the “experimental” cases, while those either receiving the status quo program or no services or benefits at all are referred to as the “control” cases.

The virtue of using an experimental design is its *ex ante* validity. In the “normal” operations of a program, it is highly unlikely that treatments are randomly allocated to subjects. If left to their own discretion, program participants likely choose treatments most advantageous for them, resulting in a very nonrandom allocation of program benefits. Random assignment of treatments, if properly implemented, ensures that the resulting experimental and control groups are, before assignment, statistically equivalent. Moreover, if subjects comply with the experimental regime by always “taking” the treatments they are assigned and never taking those that are not, the difference in the mean outcomes for the experimental and the controls will produce unbiased estimates of the impact of the program treatment(s). These estimates of program impact are not only free of selection bias, but are also simple to explain and understand. Many of the evaluations of the early state welfare-to-work demonstrations, including all of those conducted by MDRC, have been based on random assignment of treatments.

One of the other virtues of experimental evaluations is that they have relatively minimal data requirements. In principle, all one needs are data on the outcomes for the experimental and control groups; there is no need to have measures of many, if any, background variables or local contextual variables to obtain estimates of program effects. Relatively speaking, such requirements favor administrative data sources rather than gathering data with surveys. Given the potential cost advantage presented by administrative data, it is not surprising that during the 1980s, these data were the mainstay of welfare-to-work evaluations and the National Job Training Partnership (JTPA) Study funded by the U.S. Department of Labor.

The other category of evaluation methods—the non-experimental methods—includes different statistical adjustments. Some control for confounding factors by means of multiple regression techniques; these techniques to eliminate selective differences in those who receive the treatment under investigation and those who do not. Other procedures exploit longitudinal data on individuals, measuring outcomes before and after individuals participate in a program or have access to a particular set of treatments. Each of these methods is justified if a particular set of assumptions about how the treatment selection process works or about how program participants and nonparticipants would differ with respect to their outcomes in the absence of the treatment.

The inherent problem with using nonexperimental methods is that the analyst typically does not, and cannot, know with absolute certainty which set of assumptions apply to a particular set of data. It is because of this inherent uncertainty that many statisticians and program evaluation specialists strongly advocate the use of experimental designs based on random assignment when evaluating social programs or public health interventions. It is also the case that many of the nonexperimental methods available entail substantial data requirements. Samples sizes typically must be larger than when using experimental methods and require much more data, either in the form of many control variables or longitudinal data on individuals, than do experimental designs. Thus, it might appear that the use of random assignment unambiguously dominates the use of nonexperimental methods designs. During the Bush and continuing through the Clinton administration, there has been a strong preference for random assignment to evaluate waivers from federal requirements under the AFDC program

Although advantageous, reliance on experimental designs to evaluate social programs is not without its problems. First, as noted above, the validity of evaluations based on random assignment rests on the compliance of subjects with the experiment's protocols. But, as Heckman (1992) and Heckman and Smith (1997) have argued, such evaluations in practice are inherently susceptible to any number of types of noncompliance. Individuals drop out of treatments that they do not want, or they seek other means to obtain the treatments they were randomly denied. Such forms of noncompliance compromise the validity of inferences about causal effects.⁷

Second, implementing experimental designs to evaluate ongoing programs is often difficult. Conducting random assignment intrudes on the normal operations of a program. Program administrators tend to resist their use when it requires them to deny services, even at random, to clients seeking such services. They often are quite resistant to the use of random assignment for evaluating nonentitlement programs. A key feature of non-cut entitlement programs is having discretion over who is served. Administrators of these programs often guard such discretion zealously for several reasons. One is the inherent value of having been given control and not being required to serve everyone. Another is the view that this discretion reflects the comparative advantage that program staff have in identifying those individuals most suited for a particular program. A final reason is that programs are subject to performance standards and objectives; giving up control of the process by which program participants are selected is viewed as tantamount to giving up control of their fates.

For whatever reason, it is often quite difficult to get discretionary programs to agree to subject their programs to experimental evaluations that require random assignment of treatments. For example, the vast majority of local JTPA programs that were asked to participate in a random assignment evaluation under the National JTPA Study during the 1980s refused to participate. (See Traeger and Doolittle 1990, and Hotz 1992 for accounts.)

It is likely that resistance to the use of experimental designs for program evaluation will increase under PRWORA. Under this act, states no longer have any requirements to conduct evaluations of any kind, let alone ones based on random assignment. Moreover, households and individuals are no longer entitled, by law, to benefits under TANF. These structural changes in the U.S. welfare system may mean that states will be less likely to use experimental designs to evaluate their welfare programs, if they conduct any impact evaluations at all.

4.2.3 Implications for relative merits of administrative data

What are the implications of these alternative evaluation methods and the likelihood of their future use in program evaluation for the relative merits of administrative versus survey data? As already indicated, administrative data are quite more than adequate for evaluating

⁷ The dropouts do not present a problem, per se, in experimental evaluations if the causal effect of access to a particular treatment is the effect of interest.

program effects when using experimental evaluation designs. Given their potential cost advantages relative to surveys, it is not surprising that they have been the “data of choice” in the experimental evaluations of welfare-to-work and waiver demonstrations over the last 15 years. These cost advantages are likely to persist, as computer technology and software advances make assembling administrative databases less costly in the future. At the same time, it is likely that the resistance to experimental evaluations will continue, if not increase, over the next several years as state and local governments are given greater discretion over their programs.

In this context, there is likely to be a desire on the part of states and local governments to make even greater use of administrative data for whatever evaluations are conducted. The adequacy of administrative data for conducting high-quality evaluations (other than experimental evaluations) will depend on several factors.

The first factor is the ability to link information from multiple systems to create databases containing: (a) more background and demographic variables; (b) more longitudinal data on individuals or households; and (c) data on program participants and nonparticipants. Each of these improvements can increase the range of nonexperimental methods that can be used in impact evaluations. Under certain assumptions, more control variables can help account for the pre-existing differences between program participants and nonparticipants. Access to longitudinal data, in conjunction with a richer set of control variables, makes possible the implementation of more reliable forms of before-and-after evaluation methods. Devising ways to obtain data for comparison groups (albeit nonexperimental ones) of individuals or households who do not receive program services expand the options for nonexperimental methods. All three enhancements of administrative data will likely be needed for conducting reliable program evaluations.

Second, the capacity of administrative data for conducting reliable will hinge on the measures of program parameters and contextual factors discussed above. Under TANF, program parameters will be needed to characterize the differences in programs across states and possibly across counties within states. Measures of contextual factors will be important in controlling for the direct, and possibly confounding, influences that local conditions have on the outcomes. To be sure, such measures will be needed to conduct program evaluations regardless of whether administrative or survey data are used. But, their availability will be crucial if future impact evaluations can make use of less costly administrative data.

Finally, program evaluations based on nonexperimental methods will be greatly enhanced if data are available for multiple policy or program jurisdictions. In the absence of experimental designs, evaluations will need to rely on the variation of the parameters of social and public assistance programs (i.e., treatments) across states or, within states, across counties and local governments. Such variation will be crucial in identifying the relative effects of any given set of parameters. Although having access to data across jurisdictions will be important regardless of the source of individual- or household-level data used, the need to exploit cross-jurisdiction variation presents important challenges to using administrative data. In particular, it will necessitate that comparable measures of outcomes (and other variables) can be constructed in the jurisdictions used in such analyses. As noted in previous chapters, one of the frequently encountered features of administrative data is that the same phenomena (for example, “working”) are either measured in different ways or are not measured at all. Thus, being able to rely on administrative data for cross-jurisdiction evaluations will hinge on developing a common set of definitions and standards for measures in the future.

4.3 Data for Performance Monitoring and Accountability

Monitoring a program’s performance to determine whether it is meeting its objectives or mandates is largely viewed as a management, not a research, issue. But, as already noted, many of the criteria for judging the suitability of data sources for conducting research apply equally to their use in performance monitoring. The threats to determining what differences

in outcomes can be attributed to a program in evaluation research are also applicable to monitoring the program's performance. Discerning whether welfare caseloads declined because of program changes or because of improvements in local employment opportunities is not just important for the validity of a program evaluation; it is also essential for accurately assessing the soundness of a program's management structure.

Because of this overlap between what is good for evaluations and what is good for performance monitoring, the points raised in the previous section with regard to the relative strengths of administrative and survey data apply equally to designing high-quality and valid performance monitoring systems. Consider, for example, the monitoring of employment status of families on state TANF caseloads contained in PRWORA. The act requires that states report on the work activities of adults in TANF assistance units. Moreover, the failure of states to meet increasing rates of work participation can result in financial penalties, namely, reductions in the size of the block grant a state receives. It would seem that an important criterion in developing operational measures of work activity is *that the measures be comparable across states and that they be comparable across time*. To do otherwise leads to inequitable treatment of states. In essence, comparability of data measures and populations to which they are applied is the only way to run a fair accountability system.

Having comparable data across time or jurisdictions is also important for assessing trends across states and over time. But, under TANF, states will be designing very different strategies for encouraging self-sufficiency and will likely exploit their prerogative of defining eligibility for their TANF programs—which will likely result in differences across states in the information that is collected. This reality will present significant challenges for both developing useful monitoring systems for purposes of accountability and for conducting program evaluations.

The increasing reliance on the output of performance monitoring systems in government (as in the private sector) to judge a program's success or failure also has important implications for the quality, or potential biases, in the outcome measures garnered from administrative data. Consider the JTPA system. The act that created this system mandated that local JTPA jurisdictions, or service delivery areas (SDAs), monitor and report to the federal government the employment rates and initial wages of "graduates" of each local program. Moreover, these outcomes were used to allocate a percentage of federal funds across these programs. As has been noted by analysts of the JTPA, local programs frequently responded to the incentives created by this scheme. For example, programs would delay the official enrollment of applicants into the program, not entering individuals into the program's administrative system until the individual had shown evidence of being a "good prospect" for gaining employment. Some programs would put applicants through an initial set of program activities, such as having them attend orientation meetings, to determine which applicants were likely to be conscientious about regular participation in the program. In one sense, these responses represent natural reactions on the part of program administrators to the financial incentives they face. At the same time, they compromise the ability to accurately measure the average impact of a program's services because the data available from the program only represents a selective subset of individuals who were exposed to the program.

This does not suggest that programs should not devise and use performance-monitoring schemes to manage their programs. Good management and accountability to taxpayers certainly can more than justify the use of such systems. Furthermore, the potential for the way data are gathered to impart biases in the measurements produced is not unique to administrative data. Surveys can be subject to their own set of biases, including "house" effects, or the reluctance of individuals to accurately respond to questions about sensitive subjects. Rather, the Advisory Panel wishes to caution that such systems may present important challenges for the interpretation of the data that administrative systems may produce and to keep such potential biases in mind when interpreting research findings.

4.4 The Potential for Linking Administrative and Survey Data

Throughout this chapter, the implicit assumption has been that the choice between administrative and survey data is an either-or proposition. In fact, there is potential for linking administrative and survey data on individuals or households, at least in some localities and states. For example, it is feasible to match the records on households in the SIPP with data from state administrative records, as SIPP obtains the Social Security number of respondents. Linking information from state administrative databases with survey data on individuals or households has the potential for improving the measurement of program participation and of program services actually received. The latter are phenomena that are often hard to accurately measure in surveys, as people may not recall exactly what programs they were in, when they were in them, or what they received. Measuring these phenomena accurately with national surveys is likely to become even more acute under TANF, as state programs under PRWORA will be very different.

Although linking survey and administrative data has some potential payoffs for descriptive and evaluative research, it is not likely to be a panacea for several reasons. First, linking administrative data with national surveys will be of limited value for research on states with small populations because the sizes of state-specific samples will generally be too small to produce reliable estimates of most outcomes of interest. Second, the threats to confidentiality and privacy become much greater with matched data, given that the Bureau of the Census, the collector of most national surveys, is subject to strict mandates for safeguarding the identity of survey respondents. As such, these linked data could not be made available to researchers outside the Census Bureau. However, such linked data might be accessible by researchers at census facilities in Washington, D.C., or in one of the bureau's secure Research Data Centers. Currently, there are only two such remote Research Data Centers, in Boston and at Carnegie-Mellon University in Pittsburgh, but it is likely that a number of such secure sites will be opening in the coming years. This expansion of secure data sites holds promise for the creation and use of linked survey and administrative data, at least for data in large states.

CHAPTER 5

Examples of Ongoing Research Capacity

Earlier in this report, the Advisory Panel identified an emerging trend in the use of administrative data—instances where an ongoing capacity to make data available and ready for research has been developed, improving the research potential of administrative data.

In this chapter, examples are provided of efforts in five states. Certainly there are other efforts, such as those in Florida and Maryland.¹ The examples offered here differ in scope, concentrating on one area of social service programs to the full range of public assistance programs. They represent both more mature, in-use systems and those that are in the planning stages. They also demonstrate how developing an ongoing research capacity is a natural outgrowth of the success that state agencies and researchers have experienced working together on individual projects.

- Section 5.1 summarizes the lessons common to these examples and comments on some important ways in which the efforts differed.
- Section 5.2 offers case histories of efforts in:
 - **California** where the collaboration of state and university researchers on data collection and evaluation for California’s federal AFDC waiver, the 1992 Work Pays Demonstration Project, has led to the creation of five ongoing databases and the construction of research-quality analytic data sets;
 - **Illinois**, where a collaborative effort between university researchers and a single state agency in the early 1980s has evolved into a multiservice, integrated research database, constructed from administrative data gathered by numerous public agencies serving children and families in Illinois;
 - **Massachusetts**, where efforts to link tax administrative databases with other agency databases for enforcing child support payments demonstrated the potential of administrative data for research and evaluation and earned support for the creation of a longitudinal database for research on social service programs, which is now under development;
 - **Texas**, where the availability of administrative data for research has been largely facilitated by the establishment of performance measures by the state legislature in the early 1990s, evaluation of the Texas JOBS program, implemented in 1990, and the multi-agency data collection and data sharing that both required; and
 - **Oregon**, where an integrated database project, still in the planning stages, was mandated by the state legislature in 1993 in order to provide a database for future program evaluation.

5.1 Common Lessons

Different as the efforts outlined below are, they nevertheless offer some lessons.

¹ The work of Jay Pfeiffer, who manages Florida’s Workforce Education and Outcome Information Services Program, and others provides a good example of statewide incremental progress, as does that of Catherine E. Born and researchers at the University of Maryland, School of Social Work in Baltimore. There are even instances of multiple efforts in the same state, as in California, where the Child Welfare Research Center (CIRC) of the University of California at Berkeley, School of Social Work, led by Richard Barth, is linking child welfare administrative data from 58 California counties.

- There is not just one model for success. Different organizational forms and different data architectures have proven to be useful.
- Those interested in developing administrative data for research purposes must find ways to protect the interests of agencies while providing data to researchers. In every successful case the panel knows of, the success depended heavily on researchers who understood the problems of agencies and on agencies who were convinced of the usefulness of administrative data:
 - Giving researchers access to agency data poses risks and is costly; agencies have reason to fear the all-too-often oversimplified interpretations of evaluation research by politicians, the press, and the public. They often need answers “now.” Sharing data places unavoidable burdens on agency employees who will be called on to provide the data, explain it, and fix the problems.
 - Researchers tend to be extremely cautious about the implications of their findings until they have been challenged and confirmed by their peers. They generally are accustomed to releasing their findings in the absence of consequences that pertain to agencies and policymakers.
 - Mutual understandings developed over time, as researchers proved their data could be useful to the agencies (which often requires finding projects that answer a number of agency questions all at once), and as agencies came to trust researchers to provide honest interpretations of the data without embarrassing them.
- Entrepreneurs with a strong social science background and a strong policy focus, either on the academic side or agency side, have been vital to the process. They have provided the vision, the leadership, and the stamina to get things done.²
- Each organization has encountered ups and downs as it has tried to develop trust with policymakers and as has struggled to solve the substantial technical problems involved in constructing useful analytic data sets.
- Technical problems with the data are an ongoing source of challenge and irritation on both sides.
- The process of developing large, linked administrative data sets is incremental. Most of the large efforts began on a much smaller scale, growing out of a partnership between researchers and agency staff working together on a particular project.

The states have also differed in some important ways.

- Most states have kept the data confidential and promised agencies that analyses would not be undertaken without their permission. California, however, has created several public use data sets that, by definition, can be analyzed by anyone who acquires them.
- These efforts have also varied a great deal in the way they construct administrative databases. Texas has moved to an as-needed approach. Chapin Hall has integrated data from many different sources on an ongoing basis. Likewise, in California, University of California Data Archive and Technical Assistance (UC DATA) has worked with the state Department of Social Services (CDSS) to integrate welfare data from a number of sources on an ongoing basis.

² Interestingly, in these cases at least, none of the entrepreneurs have been data processing professionals. This may have implications for the current enthusiasm for data warehousing in the information technology community as a way to develop the research utility of administrative data.

- The states have also differed in the degree to which administrative databases are constructed by those close to, or even within, the state government. Massachusetts has developed a capacity within state government, and Oregon, though at the planning stage, is also doing so within state government. The effort in Texas is located at a state university, as is UC DATA, where there is a strong relationship between the analytical efforts and the state agencies. The Chapin Hall effort in Illinois is at a private university. Also in Illinois, a consortium of university researchers from private and state universities have been selected by the Illinois Department of Human Services (IDHS) to use its administrative data to report on welfare reform in the state.³

5.2 CASE HISTORIES

5.2.1 California Work Pays Demonstration Project

The Research Branch of the California Department of Social Services (CDSS); University of California Data Archive and Technical Assistance (UC DATA); the Survey Research Center (SRC) at the University of California, Berkeley; and the Welfare Policy Research Group at UCLA have collaborated on data collection and evaluation for California's federal AFDC waiver, the California Work Pays Demonstration Project (CWPD).

When initial implementation of the California waiver went into effect on December 1, 1992, UC DATA was asked to design and implement a series of data collection strategies for an experimental evaluation of the work incentives feature of the waiver. The central feature of this strategy was the designation of 15,000 cases on AFDC in four counties (two in southern California—Los Angeles and San Bernardino—and two in northern California—Alameda and San Joaquin) as research cases. Choosing these cases was greatly simplified by the existence of the statewide MEDS file, which has records on all Californians who are signed up for Medi-Cal. MEDS delineated the universe of potential research subjects because all AFDC recipients automatically appeared on MEDS through their eligibility for Medi-Cal.

A supplemental group of 4,000 pregnant and parenting teens has been drawn to evaluate the Cal-Learn component of the CWPD. The aim of this program is to encourage teens to complete high school. Each of the 4,000 teens is being assigned to one of four cells in a two-factor experimental design. One of the two factors is case management services, and the other is monetary actions and incentives. Teens assigned to one cell receive both case management services and monetary sanctions and incentives; those assigned to another cell receive neither; and those assigned to one of the other two cells receive one or other of the treatments but not both. The impact of case management services and monetary sanctions or incentives and their combination can be evaluated with this design.

To develop the best possible tools to evaluate the CWPD, a number of data sets have been constructed based on the MEDS file and the over 15,000 research families in the four counties. These data sets include:

- **Longitudinal Databases (LDB):** Ten percent and one percent samples of cases and individuals have been taken from the MEDS file from 1987 to 1997. These samples

³ Researchers from Northwestern University, Northern Illinois University and the University of Illinois at Chicago have formed the University Consortium on Welfare Reform in response to the Welfare Reform Research and Accountability Act passed by the Illinois General Assembly (see appendix III). The consortium is seeking funding for a project, "Evaluating Welfare Reform in Illinois: A Panel Study of Recipients and Their Experiences," that will link IDHS administrative data with survey data and track 3,000 adult recipients of TANF from July 1998 through June 2004 regardless of whether they stay on aid or not (Lewis 1998).

include all Californians who are enrolled in Medi-Cal, and they are constructed to be continuously updated, rolling cross-sections with continuous monitoring of families once they begin receiving assistance. This continuous follow-up provides the longitudinal component to the data. Data on quarterly earnings from the state Unemployment Insurance data files have been added to confidential versions of these files. These files have a small number of variables (e.g., program status, age, sex, race, county of residence, quarterly earnings, industry of work), but they have a very large number of cases, and they are continuously updated over time.

- **Research Sample Longitudinal Database (Sample LDB):** The MEDS files have also been used to construct a longitudinal database for the 15,000 research cases. This file has the same information as the LDB described above.

- **County Welfare Administrative Database (CWAD):** The CWAD provides information derived from monthly dumps of county AFDC and food stamps databases on the 15,000 research subjects. This monthly information has been formed into four longitudinal databases that run from December 1992 to at least 1997. The initial plan had been to create a truly uniform set of codes and variables across all four counties, but the county AFDC and food stamp case management systems were simply too different to make this possible. In efforts to create comparable data systems in its 58 counties, California has created four data systems consortia; each of the four research counties belongs to a different consortium so that the data systems are substantially different from one another. This has made it especially difficult to create a uniform database across the four counties. Nevertheless, there are many comparable variables, and there is a great wealth of information available in the CWAD.

- **Panel Surveys:** Two waves of in-depth telephone interviews with a 15 percent sub-sample of the 15,000 original research cases, or about 2,000 female heads of assistance units who speak English or Spanish, have been conducted. In addition, two waves of a parallel foreign language survey of 1,350 people who speak Armenian, Cambodian, Laotian, or Vietnamese have been finished. These four language groups were chosen because of the 15,000 research cases, each of the language groups constituted 1 percent or more of the sample. The English-Spanish and foreign language surveys ask basically the same questions, but the foreign language survey includes some additional items about refugee status, including ESL classes and camp experiences.

The panel surveys include background and outcome information that is almost never available from administrative data systems, including questions about education, AFDC history, work history, housing quality and stability, economic hardship, hunger, respondent and child's health and disabilities, labor market activities of partner or spouse, income, child support, child care knowledge and use of child care, and knowledge of work incentives. The rate of interview refusal is extraordinarily low; the greatest problem with conducting the interviews has been locating respondents.

- **Cal-Learn Studies:** A number of data sets are being developed for the Cal-Learn study. Administrative data from a variety of data sources, including AFDC, GAIN (the California JOBS program), and Adolescent Family Life Programs, have been added to the Cal-Learn administrative database. This process has involved linking the four different AFDC systems to several different GAIN computer systems and to the Lode Star data system used for the Adolescent Family Life Programs. Linking these very different systems has presented some very tough problems. In addition, a survey of those in the Cal-Learn program that provides detailed information about knowledge of the program, educational achievement and attainment, family circumstances, and child well-being.

5.2.1.a RESEARCH USING THE CWPDP DATA SETS

A series of reports on the CWPDP have been produced based on these data, including a major interim report of the first 30 months of Work Pays (Becerra et al., 1996). In addition, many researchers have used these data to answer specific research questions. Other important questions such as disability and AFDC, the use of the Earned Income Tax Credit (EITC) by welfare recipients, and the role that job availability plays in exiting from welfare have been approached using the CWPDP data sets.

Disability and AFDC. Although TANF allows states to exempt up to 20 percent of their caseload from the five-year limit on assistance for conditions such as disability, surprisingly little is known about the actual impacts of disability on the receipt of welfare. The likely reason for this is that there are very few data sets that link information on disability to information about AFDC receipt. Henry Brady, Marcia Meyers, and Samantha Luks (ongoing) are using the CWPDP data to investigate the impact of child and adult disabilities on the duration of welfare spells.

The CWPDP surveys were designed to ask detailed questions about disabilities of children and mothers. This alone, however, would not have been particularly useful because the surveys reliably indicate AFDC status only at a point in time. Linking the panel surveys with the research sample Longitudinal Database (LDB) provides reliable retrospective AFDC history extending back to 1987. Linking the panel surveys with the County Welfare Administrative Database (CWAD) provides reliable information on AFDC history after the interview dates. In addition, the MEDS file has been used (after overcoming some difficulties) to check SSI status for each of the families in the CWAD (and in the panel surveys that are nested within the CWAD).

Although there are few studies that directly examine the relationship between disabilities, work, and welfare receipt, Acs and Loprest (1994) have used 1990 SIPP data to show that mothers with severe or multiple limitations are less likely than others to leave welfare for a job. They find little consistent evidence that the disability status of children affects these transitions. These findings may occur because it matters what transitions are being studied. Using the CWPDP data, Brady, Meyers, and Luks (1996) show that the disabilities of mothers and children do not seem to predict exits from AFDC, but they do predict the kind of exits that will occur. Simply put, disabilities of both mothers and children appear to either lead the case to exit AFDC to SSI or to increase the time receiving AFDC if the family does not move to SSI. These two competing effects cancel one another out when one looks only at exits from AFDC because families can exit in two ways to SSI or completely off AFDC and SSI.

Take-up of the Earned Income Tax Credit. There have been very few reliable studies of take-up rates of the EITC among poor people. The CWPDP databases provide a very large sample of poor people along with detailed information on their incomes, making it possible to do a detailed study of this subject. Hill, Hotz, Mullin and Scholz (1997) are studying the EITC using data on AFDC recipients, which links most of the CWPDP databases with IRS tax records, to get a better understanding of how many welfare families take-up the EITC. This study relies on the detailed income and earnings data available in the CWAD and in the LDB after it has been linked to UI data.

Job Availability and Exits from Welfare. One of the most important questions confronting implementation of the new welfare reform is whether there will be jobs for those who seek to make the transition to work. In a recent paper using CWPDP data, Hoynes (1996) has demonstrated how transitions from welfare are facilitated by strong demand for labor and impeded by weak labor markets. The Hoynes study uses the LDB linked to local area data through zip codes and county of residence. The LDB provides a large enough sample to determine whether exits are affected by local labor market conditions.

5.2.1. b KEY LESSONS FROM THE CDSS / UC DATA EXPERIENCE

The experience of CDSS and UC DATA has demonstrated the possibility of linking administrative data files to construct research quality analytic data sets; it has also made

clear the added benefits to be gained by conducting surveys that can be linked with the administrative data. The experience also demonstrates that administrative data files become more useful as they are:

- Extended in time to create longitudinal data sets;
- Linked together to provide more variables; and,
- Cleaned and documented to make them readily accessible.

The CWPDP data sets have been designed so that they can be linked, complement one another, and provide information on important policy issues, such as teenage parenting, quality of life for welfare recipients, disabilities, job preparation, and employment. They provide the basis for monitoring many aspects of the welfare system and for answering very diverse research and evaluation questions.

Although CDSS and UC DATA were (and are) committed to protecting privacy and maintaining confidentiality, some agencies and human subjects committees still balked at making data available. In most cases, these problems were overcome as agencies became convinced that confidentiality could and would be maintained. The matching of records also led to problems. Substantial effort has been devoted to solving these difficulties.

The creation of public use files has required careful thinking about what types of information can be released without compromising people's privacy. Public use has also required substantial efforts to develop documentation that is useful to researchers. The vagaries and oddities of administrative data sometimes cannot be cleaned up and must be explained in a way that makes it possible for researchers to use the data. This often requires detailed investigations into how the data were actually obtained and entered.

5.2.2 The Integrated Database on Children's Services in Illinois

The Integrated Database (IDB) on Children's Services in Illinois is a prototypic multiservice, integrated research database constructed from administrative data gathered by public agencies serving children and families in Illinois. Researchers at the Chapin Hall Center for Children at the University of Chicago have been working on the database since the early 1980s. The initial goal was to construct a fully longitudinal foster care database that would permit the study of foster care duration, exit, and re-entry. That work grew out of a project to create a longitudinal database on foster care using data from the computerized child-tracking system maintained by the Illinois Department of Children and Family Services (DCFS).⁴

At the time, acquisition of the child welfare data by Chapin Hall was made possible through a collaborative research project undertaken with DCFS. The initial database contained the records of the entire population of foster children from 1976 to the present, documenting services provided, children served, outcomes, and costs. By linking together the records of each client over time, the database offered a rare view of children's contacts with foster care over long periods. Unlike the cross-sectional data used in most child welfare research, longitudinal data permit the study of temporal dimensions of care (such as sequence, duration, and outcome of services) and reveal what proportion of children re-enter the system—questions that have important program management and policy implications.

The IDB project grew more ambitious as the staff became adept at working with administrative data and more curious about the other service experiences of foster children. To study the entire human service histories of children and families over time and across service programs, it became obvious that the goal of the database project had to be extended beyond the core foster care data. Data were extended to those on children from other public agencies, especially those providing services to children and maintaining vital statistics on children, and those providing financial assistance to families. With strong support from

⁴ The multi-state foster care data archive has been created at Chapin Hall through funding from the U.S. Department of Health and Human Services (Goerge, Wulczyn, and Harden, 1995).

DCFS, Chapin Hall was able to obtain administrative data from additional agencies to augment the database. The resulting database, which is now operational, documents all child and family contact with the following programs and services: foster care, child abuse and neglect, special education, mental health, family case management, Women, Infants and Children Nutrition program (WIC), juvenile justice, Medicaid, Food Stamps, and AFDC/TANF.

When any child has had contact with more than one agency, the various records belonging to the child are linked, resulting in a particularly rich array of information in these cases. The databases extend back at least a decade and represent complete service populations. Linking all records in the database has yielded the first unduplicated count of Illinois children receiving state services. The database is being updated on an ongoing basis for the study of children's policy and services in order to be responsive to policy changes across the spectrum of child and family issues.

The process of constructing the database overcame many obstacles, the first of which was data acquisition. Agencies were understandably wary of releasing their data but at the same time were interested in supporting and benefiting from such a significant information resource as Chapin Hall had to offer. The Illinois Governor's Office was particularly supportive of the effort, though skeptical that it could be accomplished. To facilitate access to data, Chapin Hall agreed to give agencies the opportunity to review research findings before they were made public. Formal agreements with the agencies' established principles governing the proper use and ownership of data, ensuring the confidentiality of client data, and data access and security. Formatting and documenting the source data presented another sizeable impediment to constructing the multi-agency database.

The IDB project's biggest technical challenge has been to link the records of individual children reliably. Another technical challenge has been to conceptualize and develop a general database structure capable of ordering and subordinating the massive quantity of data and the wide range of variables from the constituent databases into an integrated database. The data received from administrative systems contain much redundant information and tend to be poorly structured for anything beyond the specific reporting requirements of the database.

5.2.2.a RESEARCH USING THE IDB

The Illinois database has been used for a wide range of research activities to support the information needs of Illinois human service agencies. For example, through the DCFS "Quick Response" Quantitative Indicator Analysis Project, the database has provided DCFS management staff with data and research results, including information about special needs of foster care children, placement characteristics, kinship care, and agency performance indicators.

In addition, the database has been used to support the Governor's Task Force on Human Services Reform by providing data and information on human service users to inform a state-level reorganization of human services. The database also has been used to gain a better understanding of the characteristics of the human services population and their experiences across the components of the system. These types of research activities include:

- Analysis of the demographic characteristics and human service use of all Illinois children and adolescents identified as having a disability;
- Study of children in psychiatric hospitals and the mental health needs of children in foster care;
- Analysis of the effects of teenage childbearing on child protective and child welfare services;
- Analysis of child welfare service histories; and,
- Study of the interaction between child welfare and income maintenance programs.

Researchers obtain access to the data by gaining permission directly from the agency (or agencies) providing the data. Thus, any researcher who can get permission from the agency and can cover the costs to Chapin Hall of providing the data, may have access to the data. At present, Chapin Hall has provided mainly Illinois child welfare and AFDC data to other researchers responding to two or three requests for data per month.

5.2.2.b KEY LESSONS FROM THE IDB EXPERIENCE

Just as with any research in public policy, the researchers must become familiar with the settings that are relevant to the conduct and topic of the research. For the Chapin Hall researchers, this meant understanding the way in which the data are collected and the context surrounding that collection. It also meant understanding service bureaucracies and the political environment in which they operate.

Although the research cannot be driven by computer scientists and programmers, using their skills is necessary in building a database that can be efficiently used and updated. Good design and programming practice can save hundreds of thousands of dollars in the database creation phase.

Finally, confidentiality concerns are usually overcome by both discussing the interests of the agency leaders and the researchers and through written agreements. When they are not overcome, it is usually not a matter of law, but a matter of trust or politics and not necessarily at the individual level between the agency leader and the researcher. For instance, a public agency may simply not want to have a relationship with any university.

5.2.3 **Massachusetts Department of Revenue, Child Support Enforcement Division**

In January 1992, the Massachusetts Child Support Enforcement Division (CSE) of the Department of Revenue (DOR) began linking DOR tax administrative data files with CSE administrative files in an intensified effort to prioritize cases, locate income and assets of absent parents, and estimate policy effects.

As the state tax agency, DOR maintains line-item information from tax returns and wage earnings reported by employers. For several years, the tax research unit of DOR, the Bureau of Analysis, Estimation, and Research (BAER), had been extracting information from the tax administration computer systems for research purposes. Because Massachusetts law allows the CSE division access to the DOR tax administration data for child support collection purposes, there were no legal (and only minor political) objections to linking tax data with CSE data. Under the mandate of senior officers of the DOR and of the CSE division, several divisions within DOR (the Information Systems Operations Division (ISO and BAER)) and within the CSE (the Information Research Bureau and Research Unit) collaborated to meet the technical challenges of linking the data.

Tax administrative data are an excellent source of information on income and assets. However, the CSE was also interested in the interaction between the Child Support Enforcement Program and other income-maintenance programs, and the impact of current and proposed policies. CSE already had in place interagency agreements regarding data sharing for child support collection with the departments of Transitional Assistance, Medical Assistance, Employment and Training, and Corrections, and with the Registry of Motor Vehicles. CSE had primarily used these agreements to obtain data on a case-by-case basis. In 1992, the use of these agreements was broadened by requesting data in electronic format on the entire CSE caseload or, failing that, the entire caseload of the agency providing data.

5.2.3.a RESEARCH USING CSE DATA

Initially, data were obtained and merged with CSE administrative files at the micro-level on an "as-needed basis." Need was determined by the questions decision-makers asked. For example, in October 1993, administrators of the Child Support Enforcement Program proposed that employers be required to report all new employees to DOR within the first week of employment. To determine the impact of this policy on child support collections, a

longitudinal data file that linked administrative files of the Child Support Enforcement agency and the wage-reporting agency was developed from the data archived at DOR. Results from econometric analysis performed on this data file provided crucial support for legislation mandating the immediate reporting of new hires to DOR. New-hire reporting has substantially increased child support collections in Massachusetts, and national welfare reform now requires that all states implement the new-hire reporting requirement.

Another example is the Massachusetts cost-avoidance study (Luttrell 1994). Quantifying cost-avoidance is an important aspect of evaluating the performance of child support enforcement programs and developing incentive mechanisms. This study employed a data file that linked administrative files of the CSE, the wage-reporting agency, and the Department of Transitional Assistance to quantify the savings in undistributed welfare benefits that can be attributed to efforts of the Child Support Enforcement agency.

5.2.3.b THE MASSACHUSETTS LONGITUDINAL DATABASE FOR RESEARCH ON SOCIAL SERVICE PROGRAMS

It became apparent that computerized administrative data were an underused source of information for research and evaluation of welfare and child support enforcement programs, as a result of the success of ad hoc projects. Therefore, the CSE Research Unit began regularly collecting data from state agencies, regardless of current data needs. However, files were still linked only on an as-needed basis. Although these data files constitute a potentially rich source of data, they needed to be completely linked at the individual client level both longitudinally and across agencies before their full value could be realized.

At the beginning of 1996, CSE increased dedicated resources and secured relevant expertise from Chapin Hall and UC DATA to design and develop the Massachusetts Longitudinal Database for Research on Social Service Programs (LDB) (Massachusetts DOR, 1995). CSE negotiated contracts with Chapin Hall and UC DATA that covered, among other issues, data security and permissible data use. Chapin Hall staff created the database design, using their system as a model (see section 5.2.2, above). Department of Revenue staff, with assistance from Chapin Hall, are loading the data onto a platform located at DOR. The LDB became operational in the spring of 1998. DOR staff maintain and will expand the LDB.

The usual problems of accurately linking the records of individual clients across agencies over time were encountered and addressed in this project. DOR conducted extensive data cleaning and documentation of the files, which has involved extensive communication with the source agency to learn how each data item had been originally defined and whether any changes had occurred in the process. The primary effort in creating the LDB focused on ascertaining the reliability of the key variables, while documenting and correcting some inaccuracies.

Maintaining data confidentiality is a key issue in this project. The core antipoverty program administrative data files were provided to the DOR through mutually devised confidentiality agreements with the Department of Transitional Assistance and other human service agencies. Once the data are sent to the DOR, extensive procedures to ensure data security and to control access to data are implemented, maintained, and enforced. These procedures include inventorying confidential records when received, storing data tapes in a locked facility, and maintaining passwords.

Access to the data is limited and strictly monitored. Access is restricted to two units of the CSE division, the Research Unit and the Analysis and Reporting Unit, as well as agents under contract with the DOR. Individuals who are neither on DOR's staff nor under contract with DOR do not have access to the LDB. All DOR employees and agents under contract with DOR are required to strictly adhere to department guidelines for protecting confidentiality of all individual level records. DOR employees and agents under contract with DOR are also subject to relevant federal and state disclosure laws.

Full exploration of the data will require an extensive research agenda. The CSE Research Unit and Analysis and Reporting Unit are collaborating with academic and policy research

organizations. Public use databases extracted from the LDB will be made available to individuals who are not DOR staff or under contract with DOR. They will be created by applying the following disclosure limitation steps to the LDB:

1. Remove identifiers;
2. Release only a random sample;
3. Limit geographical details;
4. Categorize continuous variables; and,
5. Examine the LDB and the sample for unique and eliminate population unique from the release sample.

The integrity of the results of analysis using the LDB depends as much on the methods of analysis as on the data. Therefore, use of both the LDB and public use databases will be limited to individuals with the expertise required to do quality research on projects that are consistent with the needs and goals of the Commonwealth of Massachusetts and its Department of Revenue. The LDB Steering Committee and the LDB Research Advisory Group control the use of the LDB and the extracted public use databases. Access to the data must be approved by both the Research Advisory Group and the Steering Committee.

Construction, maintenance, and expansion of the database has been and will continue to be funded primarily by the Department of Revenue. The federal government provided some funds for the initial construction of the database through a grant.

5.2.4 Continuing development of archived data by Texas

Although statewide systems in Texas are the norm rather than the exception today, until the early 1990s the collection, extent, and contents of Texas administrative data systems varied widely by agency. Many statewide administrative data systems had been developed for federal reporting purposes or internal program management, while a handful were developed with performance management or evaluation as the primary goal. Some areas of state government—most notably education—relied on locally controlled data systems that made comparison across different parts of the state difficult. Even when statewide data systems existed, archiving of data files was spotty, and the quality of individual variables was inconsistent.

A number of factors have contributed to increasing the availability of Texas administrative data for research purposes over the past 12 years. Although early research efforts (described below) required researchers to archive administrative data files themselves, several events have accelerated Texas agencies toward recognizing the value in developing and maintaining archives of their administrative data.

- **Strategic Budgeting.** In 1991, the Texas legislature enacted the integration of complementary initiatives in budget reform, strategic planning, and performance measurement. This legislation authorized the Legislative Budget Board to require each agency to develop a statewide five-year strategic planning process and performance-based budgeting. Further legislative action in 1993 linked each agency's goals, strategies, and performance targets with each agency's appropriations, resulting in the Strategic Planning and Budgeting System (SPBS). Most agencies now maintain statewide administrative data systems and archive at least some of their data to meet the requirements of the SPBS. Other initiatives related to this process—in particular, ongoing efforts of the State Auditor's Office to certify the accuracy of performance measures reported by state agencies—have contributed to continuing improvement in the collection and quality of administrative data. Texas is recognized as one of a handful of leading states in performance-based budgeting.

- **The Texas JOBS Program.** The Texas Department of Human Services (TDHS) implemented its JOBS program in October 1990 as a collaborative multi-agency effort

that required cooperation and data collection across a number of state agencies. DHS's partners in this initiative were: the Texas Employment Commission (TEC), the Texas Department of Commerce (TDOC), and adult education cooperatives organized under the Texas Education Agency (TEA). The detailed data reporting requirements of the JOBS programs caused these agency partners to analyze the ability of their existing data systems to meet the needs of the JOBS program and to modify data collection procedures accordingly.

Currently, sharing administrative data among Texas state agencies is a widespread practice, although confidentiality and cost considerations usually need to be addressed for individual projects between the entities providing or receiving the data. These administrative data files are used for administrative purposes among and within agencies, for calculations of core performance measures required by the SPBS, and for research and evaluation purposes, such as those described below.

5.2.4.a RESEARCH USING TEXAS ADMINISTRATIVE DATA

Some of the earliest research using Texas administrative data records in the areas of workforce development and welfare policy was conducted by the Center for the Study of Human Resources at the University of Texas (CSHR) in the mid-1980s. In response to some early welfare-to-work initiatives in Texas and the publication of national welfare dynamics research, CSHR researchers approached several Texas state agencies and obtained funding to begin studying the dynamics of the Texas welfare population (King and Schexnayder 1988). Their work was conducted by obtaining data files and matching individual client records from TDHS (AFDC demographic and spell history data), TODC Job Training Partnership Act (JTPA participant data), and TEC (employment service and UI wage data) to form appropriate research data sets for the questions being addressed.

Over the past ten years, CSHR researchers have conducted a number of increasingly more complex studies using Texas administrative data. These include:

- Several welfare dynamics studies (Schexnayder, King, and Olson 1991; King and Schexnayder 1992);
- Participation and outcomes studies on several workforce development programs in Texas (Schexnayder, King, and Lawson 1994);
- Multiyear evaluations of the Texas JOBS program and a Food Stamp Employment and Training demonstration project (King, *et al.* 1994; Schexnayder and Olson 1995 and 1997);
- Development of simulation models to estimate the effects of Texas time limits on AFDC caretakers (CSHR 1995); and,
- The effects of a fingerprint imaging pilot on Food Stamp and AFDC caseloads (Schexnayder and Olson 1997).

CSHR has begun a five-year evaluation of Texas's welfare reform waiver, Achieving Change for Texas, which will be their most ambitious data-linking effort to date. Data will be obtained from seven different Texas agencies and eleven administrative programs:

- | | | |
|-----------------|--------------------|-----------------------------|
| • AFDC | • JTPA | • Immunization |
| • Food Stamps | • Higher education | • Child protective services |
| • Child support | • Public education | • UI wages |
| • JOBS | • Child care | |

Data will be linked over at least a five-year period. The data for all these programs will be obtained from statewide administrative data systems, almost all of which are archived by the agency responsible for administering the program.

5.2.4.b ARCHIVING EFFORTS ACROSS AGENCY LINES

In the state of Texas, there have been several efforts to develop and maintain archives of administrative data across agency lines. In particular, the Center for the Study of Human Resources (CSHR) and the Texas State Occupational Information Coordinating Committee (SOICC) have developed archives of various types of administrative data from the UI, job training, and AFDC systems, much of which has been linked across programs.

For several years in the mid-to-late-1980s, CSHR maintained an archive of AFDC spell and demographic data, JTPA and employment service program data, and historical UI wage records for AFDC and JTPA recipients. Because of the level of resources required to maintain these files (particularly UI wage records) on an ongoing basis, CSHR asked the Texas Employment Commission (TEC) to archive UI wage records that could be requested as needed for research purposes. In 1988, TEC began archiving these records, making them available on a limited basis to researchers and state agencies on a fee-for-service basis. The Department of Human Services also began creating annual tapes of AFDC caretakers containing complete spell history and demographic information, facilitating research on that population. In the late 1980s, TDOC began participating in an NCEP [SPELL OUT] study to test the feasibility of using UI wage records to conduct follow-up of JTPA participants and began archiving JTPA files. Because CSHR was able to request specific files from these agencies as needed for specific research projects, they discontinued archiving program administrative files in 1989.

A more recent initiative to develop and maintain links between administrative data files across agencies is operated by the Texas SOICC through the Texas Automated Student and Adult Learner Follow-Up System. Since 1992, member agencies, which include public community and technical colleges, JTPA, and some universities and school districts, have provided client seed records to SOICC so that they could track program exiters. SOICC then links information on program exiters with a number of databases to measure program outcomes. Results are then shared with each of the member agencies.

With the passage of state legislation in 1994 to consolidate a variety of workforce development and welfare-to-work programs, SOICC has recently pilot-tested the use of automated record linkage to gather outcomes data on former JOBS, FSE&T, Job Corps, and adult education recipients. SOICC and CSHR have regularly worked together to test the use of these linked files for various research purposes.

5.2.4.c KEY LESSONS FROM THE CSHR EXPERIENCE

It has become clear that demonstrating the types of research that can be achieved through linking administrative data over time and across programs encourages agencies to think more globally about the types of variables that should be captured in their management information systems, and to archive their program data files.

CSHR researchers learned that the use of administrative data requires a commitment to learn program and operational parameters within which administrative data are collected and how those circumstances change over time.

The ability to build and sustain strong working relationships over time with agencies who “own” administrative data depends on a number of factors, including:

- Careful handling of administrative data to avoid careless errors in merging data or interpreting results;
- Taking seriously the terms of confidentiality and data-sharing agreements;
- Giving agencies an opportunity to review and comment on drafts of research reports that use their data, particularly if the findings are critical of an agency or program; and,
- Demonstrating to agencies how the research produced from administrative data can lead to program improvement.

5.2.5 Building a shared information system in Oregon

The integrated database project in Oregon, the Shared Information System (SIS), was initiated by the state legislative branch. In December 1990, the Oregon Legislative Fiscal Office distributed a report, "Job Training, Retraining, and Placement: A Program Budget and Review," that cited the need to establish a common database on employment training programs. During 1991, the Oregon Legislative Assembly passed two bills that required education, training, and employment agencies to coordinate services and provide data for planning, evaluation, and performance measurement purposes. A 1991 house bill established the Workforce Quality Council to oversee the process. A 1991 senate bill addressed the need to improve the state's "corporate" data. In January 1992, the Oregon State Government's Strategic Plan for Information Resources Management stated that:

information is a strategic asset of the state that must be managed vigorously, purposefully, and for the benefit of all agencies.

The SIS was enacted into law by a senate bill passed in 1993. The objectives in creating the SIS was to provide a database for future program evaluation. The SIS can provide both inter- and intra-agency performance measures and outcome evaluations, identify duplication of service delivery and data collection efforts, and improve information used in policymaking. Currently, SIS has complete information on over one million individuals who have been served by Oregon workforce agencies.

The following state agencies were mandated to participate in the SIS:

- Employment Division,
- Department of Education
- Department of Corrections
- Bureau Labor of and Industries
- Adult and Family Services Division
- Vocational Rehabilitation Division
- Department of Insurance and Finance
- Office of Community College Services
- Oregon State System of Higher Education
- Job Training Partnership Act Administration

An additional Oregon agency that administers all subsidized health care in the state, the Oregon Health Plan, voluntarily participates in the SIS.

The Employment Division houses the SIS, though SIS staff are not technically employees of the Employment Division. The Workforce Quality Council jointly manages the technical and conceptual aspects of the project with the SIS Advisory Committee, which is made up of representatives from each of the participating agencies, the private sector, local government agencies, and Work Quality Council appointees.

Each agency was required to develop a client information release form and have it approved by the state Attorney General's office. The SIS can accept agency data only for individuals who have authorized the use of their Social Security numbers to perform matches with data from other agencies unless the agency has a federal statute allowing the sharing of information. Agencies provide data to the SIS with Social Security numbers transformed into individual identifiers using an encoder formula. Each agency has a unique encoder formula. The SIS computer is programmed to decode identifiers back into Social Security numbers and match records across agencies. A limited number of SIS computer staff are the only individuals with access to the decoding formulas.

The creation of truly common data definitions for demographics, service delivery, and outcomes across eleven agencies was a central challenge to the creation of the SIS, overcome by involving one representative from each agency and forming two committees. One committee examined data and technical issues and the other focused on common

performance measures. A third committee composed of policymakers from the participating agencies handled conflict resolution and barrier removal.

Because the database is still under construction, problems are still being addressed. For the next year, issues and corrections for selection bias (if it exists) will be examined, and different weighting schemes (because of a different percentage of clients from each agency) to increase the accuracy of the information will be considered. Further, the data directory will be refined to better reflect the services and outcomes of the SIS participants.

Participating agencies have access to all data in the SIS on their own clients. However, confidentiality restrictions prohibit agencies from making decisions about or taking action against an individual based on the information contained in the SIS. Agencies use the SIS to track program participants who leave their programs, following their work and earnings experience and subsequent participation in social service agencies. Access to the complete SIS for analysis is limited to SIS staff and contractors, although anyone can request aggregate statistics and results of analysis on the SIS. The SIS is funded through state general funds and in-kind contributions of the participating agencies.

5.2.5.a RESEARCH USING THE OREGON SIS

Data contained in the SIS have been used for profiling (probability modeling), performance evaluations adjusted for local economic conditions (expected value of agency participation), and evaluations of the Oregon Health Plan's cost-effectiveness. They have been used by agencies in presenting their budgets to the legislature, by the governor's office for evaluation of the overall state workforce and training system, and they have been used in an evaluation of proposed welfare-to-work programs in Oregon.

CHAPTER 6

Developing the Research Potential of Administrative Data: Summary of Findings and Recommendations

The Advisory Panel on the Research Uses of Administrative Data foresees substantial need for data able to support the rich array of research and monitoring needs of states and the nation in the changing climate of welfare reform and the “New Federalism.” In this concluding chapter, a summary is offered of what has been learned about administrative data sources and what appear to be the critical obstacles to increasing their usefulness in research and in monitoring policy effectiveness.

As this report makes evident, the utility of administrative data for policy-relevant research is considerable at present and promising for the future. Several states have developed or are developing structures and institutional arrangements to sustain ongoing state administrative databases to support research on state-level public assistance programs. These databases contain longitudinal information on individuals or households who have participated in one or more public assistance or training programs. The data have been (or can be) linked across a number of different administrative databases to produce a range of measures needed to assess the performance of these programs. Nevertheless, such databases have been developed only in a limited number of states. Further, their sustainability, where they have been developed, remains less than certain.

A number of recommendations for developing the research potential of administrative data emerge from the Advisory Panel findings. Because the experiences with building and using administrative databases for policy research, though quite promising, are still limited, the recommendations tend to be rather modest. The panel offers them to stimulate ways to think about information needs in a new era of social programs and policies that are more local in their orientation and more varied in their objectives. The recommendations are offered in the hope that they will help forge a mutual investment on the part of policymakers and analysts, agency officials and program managers, and researchers inside and outside the academy to pursue the more sophisticated program monitoring of a larger number of antipoverty strategies—imperative for both broad policy interests and for administering of newly emerging programs and policies.

6.1 Findings: Where We Are and Where We Need to Go

This section provides a summary of conclusions the panel drew from its investigation of the present utility of administrative data. Many of these conclusions confirm the increasing understanding of what is entailed in the use of administrative data for research, while others present possibilities for improving their usefulness in the future.

A number of observations shared by the panel about administrative data and its role in future research efforts were important in motivating the panel’s work. They were:

I. Administrative data sources will need to play a greater role in the monitoring and evaluating of the impacts of social assistance programs in the coming century.

As discussed throughout the report, the devolution of social assistance policy and programs to the state and local level will make the degree of reliance on standard surveys, such as the CPS and SIPP, less tenable in the future. The national cross-sectional and longitudinal data sets derived from standard surveys currently provide inadequate state-level sample sizes. To

support separate analyses of poverty-related issues for the majority of states, these surveys will have to develop measures of such phenomena as program participation in an increasingly heterogeneous policy climate. Forming such measures will take time and, even so, may ultimately have limited usefulness as a source of information for comparing programs and outcomes across states, let alone within states. The shortcomings of standard national survey data in the wake of devolution motivate a more serious effort to develop administrative databases that can compensate, in part.

II. To meet the research needs of future evaluations of policy and programs, there will be a growing emphasis on building administrative databases for linking information across time and across programs and agencies.

Several aspects of the changing nature of social assistance policy places great value on the ability to link administrative data across time (for a given case/individual) and across administrative record systems. The value of being able to reliably link case or individual data arises for at least two reasons.

- First, the new state- and local-based social programs generally have multiple objectives. Under PRWORA, for example, states are encouraged to develop TANF plans and programs with a multi-faceted set of objectives, including encouraging self-sufficiency, reducing out-of-wedlock births, improving the well-being of children. To assess each of these objectives, and the extent to which states can meet them as a group, clearly requires monitoring information from more than one administrative domain.
- Second, time-limited aid under PRWORA necessitates a capacity to develop longitudinal databases on individuals who participate in one or more of these programs over their lifetimes.

These fundamental needs for linking information over time and across program databases are clear. However, the ability of states to develop “linkable.” research-ready administrative databases places great strains on existing data systems for the variety of reasons discussed in chapters 2 and 3.

The panel spent considerable time identifying and studying existing efforts to develop administrative databases. Those that appeared to provide an ongoing capacity for supporting multiple types of research were most intriguing. The Advisory Panel became particularly interested in how these centers formed, how they were structured, and what keeps them going, and came to several conclusions.

III. Several lessons may be learned from the experiences of existing efforts to develop ongoing administrative data for use in policy relevant research:

- Ongoing databases are more typically the result of “bottom-up” rather than “top-down” development efforts. That is, they tend to be the result of localized and more idiosyncratic efforts as opposed to the mandates of any one agency.
- A key element in the development of successful administrative databases for research, especially those of an ongoing nature, usually have involved a

collaboration between one or more state agencies and outside academic and independent research groups or institutions.

- A key feature of the entrepreneurial effort that initiated and sustained the existing databases was the presence of someone or some group that held a longer-run perspective, so that the database was not viewed as useful for only a single project or contract.

The panel was struck by the similarity of these efforts to small firm start-ups. In terms of the energy and ingenuity that comes with such start-ups, they possess the same attraction as developing a new product in one's garage. But, as the statistics on the turnover of small firms indicate, such enterprises also can be unstable and short-lived. Below, the panel offers some suggestions for ways to help sustain the development of these ongoing databases, while maintaining the benefits of the bottom-up approach.

The panel also was struck by the fact that the cases studied for this report involved partnerships of state agencies and outside groups, including a group or institution that is part of a university or an independent research institute. Several features of these partnerships have proved beneficial to these developmental efforts, including: the independence of the research organization; the role that these partnerships have played in establishing sustaining institutional trust; and the fact that these partnerships provide already strapped and overburdened state agencies

The panel's investigation also uncovered a number of important barriers that impede the development of administrative records into analytic data sets capable of supporting high-quality research and accessible to responsible policy and academic researchers. These barriers can be described as follows:

IV. Key institutional issues representing “make-or-break” factors in the development of administrative databases are:

- The ability to negotiate interagency agreements; and,
- The ability to obtain protocols that protect the privacy of clients and the agency and the confidentiality of data.

V. Key issues in the ability to make such databases accessible to more than in-house researchers are:

- The ability to find adequate safeguards for protecting the privacy of clients and confidentiality of data; and,
- The ability to establish contractual (or less formal) relationships giving researchers freedom to conduct research while protecting the integrity of the agency, the privacy of clients, and the confidentiality of the data.

Two common themes underpin these obstacles. The first is the issue of confidentiality and privacy and the profound implications of this issue for the entire enterprise of using administrative data for research. The panel tried, in the entire chapter devoted to this subject (Chapter 3), to compile and discuss existing principles and standards that provide guidance for safeguarding confidentiality while still enabling responsible researchers to gain access to sensitive data. But, as also has been discussed, using administrative databases for research raises new issues

for which the existing standards, developed primarily in the context of large-scale national surveys, may be inadequate. The combination of the technological revolution in information systems and electronic sources of information and some unfortunate abuses of confidentiality have resulted in a level of anxiety and skepticism on the part of the American public that makes this topic all the more sensitive and in need of further assessment. This issue is addressed below.

The other common thread running through the obstacles outlined above is the element of trust, or lack of it, between various communities that produce, analyze, and ultimately own administrative data. On this score, the Advisory Panel is encouraged by the extent to which the efforts described in Chapter 5 seem to demonstrate that cooperation and understanding emerges and is strengthened through partnerships. Agencies have learned that properly informed researchers can use their data responsibly. Academic researchers involved in these efforts have come to better understand the political and operational responsibilities these agencies bear precisely because the information in their databases exist, first and foremost, to fulfill management and service-delivery needs. What remains is finding ways to replicate and sustain the trust-building experiences that have so benefited the research value of administrative data.

Finally, as discussed in Chapter 5, there are limitations to administrative data that affect their usefulness and appropriateness for many types of research. In particular, the panel has noted that:

VI. Administrative data have a number of limitations that diminish their value in certain types of research. These include:

- The choice-, event-, or participation-based nature of administrative data limits inferences and gives rise to the “denominator problem.”
- Administrative data typically do not contain adequate control variables, e.g., demographics of clients.
- Administrative data do not measure all outcomes, e.g., some types of indicators of well-being.
- Data are only available when client is “in the program”; little is known when a person leaves the program.

It is important to note that the limitations noted above need not be permanent. As discussed in Chapter 4, the capacity to link information across many administrative data sources can go a long way toward reducing both the problems of limited control variables and outcomes that would exist in any one administrative data system. Clearly, the technology exists today to perform those links. But the problems of confidentiality and obtaining the interagency agreements to perform these links represent significant costs (and potential barriers) to efforts to develop more comprehensive data systems.

Similarly, concerns about potential biases in administrative data, due to their use in managing programs and for measuring performance in accountability systems, may turn out to be unwarranted. But, to date, we simply lack sufficient experience in the research community from comparisons across survey and administrative data sources to know the nature or extent of these biases. In the next section, the panel recommends that such data comparisons be encouraged, rewarded, and funded so that limitations can be reduced or removed. But, at the same time, it is important that all parties be aware of these current

limitations and exercise caution, given the relatively early stage of using administrative data for program evaluation and analytic research in the context of social service programs.

6.2 Key Findings

- I. Administrative data sources will need to play a greater role in the monitoring and evaluating of the impacts of social assistance programs in the coming century.
- II. To meet the research needs of future evaluations of policy and programs, there will be a growing emphasis on building administrative databases for linking information across time and across programs and agencies.
- III. Several lessons may be learned from the experiences of existing efforts to develop ongoing administrative data for use in policy relevant research:
 - Ongoing databases are more typically the result of “bottom-up” rather than “top-down” development efforts. That is, they tend to be the result of localized and more idiosyncratic efforts as opposed to the mandates of any one agency.
 - A key element in developing of successful administrative databases for research, especially those of an ongoing nature, usually have involved a collaboration between one or more state agencies and outside academic and independent research groups or institutions.
 - A key feature of the entrepreneurial effort that initiated and sustained the existing databases was the presence of someone or some group that held a longer-run perspective, so that the database was not viewed as useful for only a single project or contract.
- IV. Key institutional issues representing “make-or-break” factors in the development of administrative databases are the ability to negotiate interagency agreements, and the ability to obtain protocols that protect the privacy of clients and the agency and the confidentiality of data.
- V. Key issues in the ability to make such databases accessible to more than in-house researchers are:
 - The ability to find adequate safeguards for protecting the privacy of clients and confidentiality of data; and,
 - The ability to establish contractual (or less formal) relationships giving researchers freedom to conduct research while still providing agencies with adequate controls over what researchers disclose.
- VI. Administrative data have a number of limitations that diminish their value in certain types of research. These include:
 - The choice-, event- or participation-based nature of administrative data limits inferences and gives rise to the “denominator problem.”
 - Administrative data typically do not contain adequate control variables, e.g., demographics of clients.
 - Administrative data do not measure all outcomes, e.g., some types of indicators of well-being.
 - Data are only available when client is “in the program;” little is known when a person leaves the program.

6.3 Recommendations for Developing the Research Value of Administrative Data

In this last section, we outline several recommendations to help foster the construction and research value of administrative databases from public assistance programs over the next few years. It is always tempting for panels such as this to offer bold and expansive recommendations. But, as noted in the introduction and as this report indicates, despite a number of efforts around the country to develop and use administrative data for research on an ongoing basis, the “newness” of these efforts does not support sweeping recommendations. These efforts will, however, benefit from practical, though modest recommendations, which may also support other incremental efforts to develop ongoing administrative databases. The Advisory Panel makes recommendations in three areas:

- Fostering institution building
- Further assessment of confidentiality and privacy concerns
- Assessing and improving the quality and cross-state comparability of administrative data for public assistance programs

6.3.1 Fostering institution building

Across the country, a number of opportunities are emerging for the development of ongoing administrative databases for research of social programs and policies. It is the Panel’s view that states and the nation need to build on these promising efforts and develop permanent, ongoing administrative data capacities to monitor policy changes and their impact on the disadvantaged segments of the population. To help realize that goal, the Panel offers three sets of recommendations to foster the construction of permanent administrative data institutions.

The first recommendation concerns the need to improve interactions between those designing, developing, and using state and local administrative data and their access to information of common interest. One of the messages heard from many state research staff and data managers was the difficulty they encountered in gathering information from other states and learning from their peers. Thus, the panel recommends:

I. Establish (and fund) a centralized and ongoing repository of information on administrative data.

A repository could be as basic as a web site maintained by an existing professional organization or research center. Crucial to its usefulness is that it collect and disseminate information on the following issues and topics:

- Legislative and administrative strategies at the state level for dealing with confidentiality and privacy concerns;
- Up-to-date documentation on federal regulations related to administrative data under TANF and other federal programs;
- Legislative and administrative strategies at the state level for dealing with the establishment of interagency agreements;
- Prototypes of agreements for providing nongovernmental researchers access to data that meet concerns of confidentiality and safeguards the political integrity of agencies and state and local governments;
- Reports on the ways in which different state agencies use administrative data to improve the management and accountability of their programs.

The second recommendation concerns the importance of bringing together researchers and program administrators in the development and maintenance of administrative databases that can and will be used to conduct policy-relevant research. Existing efforts to build such databases (see Chapter 5) required involvement between researchers and state and local program administrators. Therefore, the panel believes that an important way to support the

use and improvement of administrative data is to encourage states to establish such “partnerships” as they develop their administrative databases for research.

II. Encourage states without administrative databases to establish partnerships with independent research organizations, such as those at universities, to help develop and use administrative databases on an ongoing basis.

The development of administrative data in the five cases examined for this report provides several different models for partnerships. Regardless of the particular structure, the panel considers partnerships between states and research institutions an important role if both the development and quality of administrative data for research is to be assured and improved over the coming years.

¶ In considering such partnerships, the Advisory Panel takes seriously the inherent tension between establishing a trusting relationship between researchers and government agencies and the need for researchers to maintain their professional integrity so as not to undermine the credibility of research findings. Researchers need to maintain a fine line between recognizing the political and bureaucratic realities and pressures that impinge on agencies and their administrators when conducting and presenting their research while retaining the necessary independence in the work that they do. This tension was evident in several of the efforts to develop ongoing administrative data research efforts discussed in Chapter 5. For example, some of the reports produced by the Center for the Study of Human Resources (CSHR) at the University of Texas were critical of the procedures and practices of the state agencies whose cooperation they needed to get their data. But, as panel member Dr. Deanna Schexnayder noted, the center was able to establish a credible voice within the state, which is respected and listened to by various parties, precisely because researchers with the CSHR, as well as the center itself, have maintained their independence.

Third, it is important that mechanisms be created to more visibly recognize and encourage efforts to build and use administrative databases in research.

III. National organizations (such as American Public Welfare Association (APWA) or the Welfare Information Network (WIN) or National Governors Association (NGA)) as well as organizations and groups within the academic community (such as Association for Public Policy and Management (APPAM) and National Association of Welfare Research and Statistics (NAWRS)) need to find ways to recognize and encourage the use of administrative data in research.

Rewarding the painstaking and often thankless work that must be undertaken to develop and sustain ongoing administrative databases will help to encourage their development. Organizations such as the American Public Welfare Association or the Welfare Information Network or National Governors Association should identify state or local examples of accomplishments in developing administrative databases, and could even establish awards or grants to encourage these efforts. Organizations such as the Association of Public Policy and Management, the American Statistical Association, and the National Association of Welfare Research and Statistics, might encourage presentations and sessions at their annual conferences devoted to research on the methodologies supporting administrative data, and on research that used administrative data sources and program applications that have redesigned administrative data systems to serve both the ends of research and evaluation and program monitoring. These latter efforts would signal the importance and legitimacy of using and analyzing administrative databases in scholarly research as well as make useful design changes more attractive to program managers.

6.3.2 **Further assessment of confidentiality and privacy concerns**

In its assessment of confidentiality and privacy concerns, the Panel found that many of the existing principles and recommendations of previous initiatives apply to the research uses of administrative data. At the same time, new issues were identified that threaten confidentiality and privacy protections when using administrative data. The panel also found that while federal confidentiality guidelines are well-established, guidelines and legislative acts at the state level are quite diverse and, in some cases, the principles to protect privacy and assure the public (including program participants) that information on them will not be used improperly are potentially inadequate. Therefore, the Panel thinks it wise to re-examine and further assess the adequacy of existing practices, especially in light of growing public skepticism about the privacy of information governmental units possess.

I. **Independent organizations (such as the Committee on National Statistics), as well as professional organizations (such as the American Statistical Association) need to conduct a more thorough assessment of the adequacy of existing principles and practices that will protect the privacy of individuals and confidentiality of the information contained in administrative databases. Special attention should be paid to such questions as:**

- How should informed consent of program participants with respect to the use of information on them for research to be handled?
- What mechanisms and procedures should be adopted that will provide access of these data to responsible researchers while still safeguarding the privacy of individuals?
- What guidance can be provided for crafting interagency agreements?
- What are the proper “disclosure” standards for these databases when reporting on results from research based on these data?

6.3.3 **Assessing and improving the quality and cross-state comparability of administrative data for public assistance programs**

The Advisory Panel’s final two sets of recommendations concern data quality and comparability across units. Great strides have been made in the “science” of developing administrative databases, especially those that contain longitudinal information on program participants and those that consist of data linked across various databases. Nonetheless, it is the Panel’s assessment that many unanswered questions persist regarding the quality and usability of administrative data for many types of research, including uses that will require cross-state comparisons.

The situation today confronting the use of administrative data to monitor and analyze the impact of the nation’s devolving public assistance programs bears a remarkable resemblance to that faced in the 1980s with respect to monitoring and assessing the condition and performance of the nation’s primary and secondary educational systems. As with the new emphasis on state and local control in the recently enacted welfare reform legislation, the primary responsibility for financing and governing schools and education programs has long rested with state and local governments. But over the last three decades, there has been a growing concern about the quality and success of the nation’s schools, sparked in part by comparison of American students in the mastery of math and science skills with students in other countries. These concerns were first chronicled in *A Nation at Risk*, a report of the National Commission on Excellence in Education, which was released in 1983. Furthermore, the debate about appropriate curriculum and methods for organizing the delivery of primary and secondary education in the U.S. gave rise to an increasing need for high-quality and comparable data on student performance and what was happening in our nation’s schools.¹

¹ For a more complete discussion of these developments, see Elliott and Ralph (1997).

Although the National Center for Education Statistics, within the U.S. Department of Education, had responsibility for providing such data, serious concerns existed about its adequacy. As described in a report issued in 1986 by the “Panel to Evaluate the National Center for Educational Statistics (NCES),” created by the National Research Council Committee on National Statistics,² the quality and comparability of data on educational performance for the nation’s elementary and secondary schools was inadequate. As noted by one of the commentators on this report, “if the data continue to be as inaccurate in the future as they have been in the past, all other issues are moot.” The panel summarized their findings about the quality and comparability of national education statistics and data as follows:

The poor quality of the data is generally attributed to the fact that data are collected, in large part, from administrative records maintained at the local level, which record “official” rather than “real” behavior; that the data are the product of diverse record-keeping systems that lack comparability in definitions and time periods; that the data provided to the center [NCES] are at such gross levels of aggregation, such as for a state as a whole, as to seriously limit anyone’s ability to check them for accuracy, consistency, and reasonableness; and that the data as published are at some summary levels of geography, such as a region, as to seriously limit their analytical usefulness (Source, P. #).

This criticism was accompanied by a series of recommendations for improving the quality and comparability of educational data collected and compiled by NCES. While still in progress, NCES has adopted many of these recommendations. For example, it has sponsored more research on the quality of its data and appropriateness of its various measures of performance. And, reflecting the political and historical realities of the local control of educational data, it has worked in partnership with states and professional organizations (such as the American Statistical Association and the American Educational Research Organization) to develop standards to improve the quality and comparability of common data across states.

In an effort to avoid repeating some of the problems confronted in developing education statistics, as well as to learn from the strategies for dealing with them, the panel offers several recommendations for ways to improve the quality of administrative data and to promote greater comparability of data elements derived from administrative data for the state- and local-level public assistance programs emerging under PRWORA.

6.3.3.a. ASSESSING THE QUALITY AND VALIDITY OF ADMINISTRATIVE DATA

The concerns outlined above about the quality and comparability of state-based educational statistics derived from administrative records in the 1980s offer parallels to the situation facing human service program administrators and researchers interested in poverty and policy today. As discussed in Chapter 4, when it comes to using administrative data in evaluating the impact of emerging state- and county-based welfare programs under PRWORA, a number of important questions remain unanswered. There have been a few studies of the comparability of variables such as income and program participation status across administrative and survey data (several of which are cited in this report). But the Panel strongly believes more research on the comparability of administrative and survey data needs to be done if administrative data are to become a trusted and appropriately used source of data in high-quality research. Therefore, the Panel recommends that funding agencies and foundations, as well as professional and research organizations, give more attention to and expand their support for research that validates and assesses the relative quality and adequacy of administrative data for all types of research, especially evaluation research.

² See Levine (1986).

- I. **Funding must be provided by agencies (such as the National Science Foundation), private foundations, and government agencies themselves to further research and analysis on such questions as:**
 - Quality of administrative data;
 - Comparability with other data sources, such as survey data;
 - Methodological strategies for dealing with analytic issues, such as the denominator problem, which affect the range of data use; and,
 - The interactions of research and management objectives and how this affects the structure and quality of data.
- II. **Research organizations (such as the Joint Center for Poverty Research) and academic publishers and journals must encourage and help legitimize research on these questions by creating outlets for it, including convening conferences and supporting volumes or special issues of journals on these topics.**
- III. **Those working on the management side of the equation, including professional organizations for the public sector, must collaborate and help support efforts to develop higher quality administrative data.**

For example, the Panel noted in Chapter 4 the importance of having high quality data for implementing “results-based accountability” systems developed by government agencies. Such groups have a direct self-interest in the improvement of administrative databases for the types of “research” that must be done to assess where praise and blame should be lodged. The panel would urge organizations such as the National Academy for Public Administration and the Council for Excellence in Government to take an active role in promoting the assessment of data quality in information systems used for performance assessment.

6.3.3.b IMPROVING CROSS-STATE COMPARABILITY OF ADMINISTRATIVE DATA

The Advisory Panel’s final recommendation concerns data comparability across states. As noted in the introduction, the current trend in social policy places unprecedented responsibility and control in the hands of state and local governments. Many have predicted that this change in the locus of control is likely to result in even less research and monitoring of program performance than in the past. But, as the initial assessment of several states has indicated, these dire forecasts seem premature. As the examples of welfare reform legislation from Illinois and California illustrate, states are not ignoring the research component in their implementation of welfare reform. Moreover, administrative data are likely to serve as a key source of data for whatever research and program evaluation states *do* perform. But as promising as these commitments to research are on the part of a few states, it is the Panel’s assessment that an important national goal remains—namely, being able to monitor and evaluate the impact of the alternative policies developed by states over the next few years. If cross-state comparisons are to prove useful and informative, data that contain comparable measures and populations at the state level are needed.

One potential source of comparable data will be from surveys of nationally representative populations, including the CPS, SIPP, and the new Survey of Program Dynamics (SPD) for the population of social program participants. But, as also noted in the introduction, the sample sizes from these surveys for conducting research will only be adequate for the largest states. Clearly, administrative databases can play a crucial role in cross-state comparisons. To play that role, however, attention must be paid to achieving greater comparability of information and populations in these databases. Accordingly, the panel offers the following recommendation to highlight this important issue.

I. **Guidelines and standards must be developed to ensure that comparable and high-quality data are gathered across states and across agencies within states.**

Following the model used for educational statistics, the Advisory Panel suggests that the National Research Council Committee on National Statistics be commissioned to establish a panel to assess and make recommendations on ways to foster data comparability. Some crucial issues include:

- (a) the availability of universal identifiers to facilitate linking administrative records across states, which should be addressed in the context of PRWORA;
- (b) review of the comparability of state-provided measures of outcome and demographic variables mandated under PRWORA; and,
- (c) assessment of what other data elements could be made available by most or all states that would be valuable in monitoring and evaluating the impact of social assistance programs nationwide.

In developing their recommendations, this panel will need to be mindful of the current political climate in which calls for stringent and mandatory guidelines or standards from the federal government are not likely to be palatable to the states. Therefore, a panel on data comparability must include in its membership representatives from state and local governments, and seek input from professional organizations such as APWA and NAWRS.

Finally, this recommended panel on data comparability should be asked to assess what institutional and government structures might be put in place to improve the quality of administrative and other data sources used to monitor and evaluate public assistance programs in the United States. Entities such as the National Center for Education Statistics (NCES) or the National Center for Health Statistics (NCHS) provide useful models that could be adapted to the public assistance context. Furthermore, such a panel should consider ways in which an NCES-or NCHS-like arrangement could foster an on-going partnership between states and the federal government, and begin to separate national data-gathering efforts in the area of public assistance from the enforcement and sanctioning that has been the focus of many past and existing federal reporting requirements in this area.

6.4 Recommendations in Three Key Areas

I. Fostering Institution Building

- Establish (and fund) a centralized and ongoing repository of information on administrative data.
- Encourage states without administrative databases to establish partnerships with independent research organizations, such as those at universities, to help develop and use administrative databases on an ongoing basis.
- National organizations (such as APWA or the WIN) as well as organizations and groups within the academic community (such as APPAM and NAWRS) need to find ways to recognize and encourage the use of administrative data in research.

II. Further Assessment of Confidentiality and Privacy Concerns

Independent organizations, such as the Committee on National Statistics, as well as professional organizations (such as the American Statistical Association) should conduct a more thorough assessment of the adequacy of existing principles and practices that will protect the privacy of individuals and confidentiality of the information contained in administrative databases. Special attention should be paid to such questions as:

- How should informed consent of program participants with respect to the use of information on them for research be handled?
- What mechanisms and procedures should be adopted that will provide data access to responsible researchers while still safeguarding the privacy of individuals?
- What guidance can be provided for crafting interagency agreements?
- What are the proper “disclosure” standards for these databases when reporting on results from research based on these data?

III. Assessing and Improving the Quality and Cross-state Comparability of Administrative Data for Public Assistance Programs

- Funding should be provided by agencies (such as the National Science Foundation), private foundations and government agencies to further research and analysis on such questions as:
 - Quality of administrative data;
 - Comparability with other data sources, such as survey data;
 - Methodological strategies for dealing with analytic issues such as the denominator problem, which affect the range of usage of data; and,
 - The interactions of research and management objectives and how this affects the structure and quality of data.
- Research organizations (such as the Joint Center for Poverty Research) and academic publishers and journals must encourage and help legitimize research on these questions by creating outlets for it, including convening conferences and supporting volumes or special issues of journals on these topics.
- Those working on the “management” side of the equation, including professional organizations for the public sector, must collaborate and help support efforts to develop higher quality administrative data.
- Guidelines and standards should be developed to ensure that comparable and high-quality data is gathered across states and across agencies within states.

6.5 Concluding Observations

Social policy is undergoing dramatic changes today, with the responsibility for design and implementation of these policies devolving to state and local governments. Because of the scarcity of resources and the reluctance of the nation's citizens to support "big government," what is at stake today in social policy is higher for states, agencies, and clients than it was 20 years ago. To address the uncertainty about "what works" and "for whom," it is important that the focus and tools of research adapt if we are to accurately and fairly describe, monitor, and evaluate just what these changes imply for our nation's poor and disadvantaged.

The Advisory Panel on Research Uses of Administrative Data is convinced that administrative data can and ought to be one of the important tools in this research effort. Administrative data can provide a cost-effective yet extremely useful source of information with which to monitor and evaluate the impact of changes in social policy at the state and local levels. At the same time, much work is needed to develop administrative data systems that can routinely provide information for this research on an ongoing basis. It is the Advisory Panel's hope that this report stimulates and encourages policymakers, program managers, researchers, and funding agencies and foundations to join in the effort to strengthen administrative data and to ensure that administrative data play an expanding role in monitoring the well-being of the nation's disadvantaged.

GLOSSARY OF TERMS

The following are definitions of some of the key terms used throughout this report:

Administrative Data This report is concerned with data and information about those at risk of needing public assistance and about programs for the poor. Consequently, use of the term administrative data strictly refers to [all] information collected in the course of operating government programs that involve the poor and those at risk of needing public assistance.

Confidentiality and Privacy The definitions of confidentiality and privacy, the widely used terms in the discussion of protecting personal information, are not universally agreed on. Confidentiality is used to mean restricting the pool of persons who have access to individually identifiable information. Information privacy, referred to here as “privacy,” is used to refer to one’s right or privilege to set the conditions of disclosure of personal information, including not disclosing the information at all.

Data Warehouses A data warehouse takes data from the day-to-day transactions of an organization and organizes it (usually in a relational database) to undertake informative, analytical processing over a longer time. A data warehouse should provide support for management decision-making, policy analysis, and longer-term research on the organization’s programs and policies.

Data Sets, Data, Databases and Analytic Data sets The terms *data sets* and *data files* are usually used interchangeably to mean collections of information on entities (often cases or people) at one point in time or over time. Database often refers to a large number of interlinked data files. Analytic data sets, as used in this report, refer to data sets and data files that have been organized with appropriate documentation to make them useful for decision support, policy and program analysis, and research.

Linked, or Merged, Administrative Data Linked and merged are usually used interchangeably, meaning that data on the same case are linked over time (e.g., AFDC records from one month to the next), over places (e.g., AFDC records from one location to another), or over data sets (e.g., AFDC data with UI data.) Linked is used with this meaning in the report. Some may distinguish between linked and merged, in that linked data are such that you can connect information about a case in one file with that in another, whereas merged files are those in which all the information is put together.

Management, or Administrative, Information Systems Management Information Systems are those systems that keep track of day-to-day information about cases. They go by different names or acronyms in different agencies, though typically are referred to simply as “our system” by agency personnel.

Research The term research covers a rather broad set of activities and products. As discussed throughout the report, all of the following activities are forms of research because of the common methods and data sources they use: descriptive research, trend analysis, program monitoring, program evaluation, and analytic research.

Survey Data The term survey data is used to mean data gathered via survey interviews, particularly the national surveys such as the U.S. Censuses of Population, the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP) that have traditionally been used for analytic research.

REFERENCES

- Acs, G. and P. Loprest. 1994. *Do disabilities inhibit exits from AFDC?* Washington, DC: The Urban Institute.
- America's Labor Market Information System (ALMIS), U.S. Department of Labor. 1997. State unemployment insurance wage records: Access and use issues. Washington, DC.
<http://ecuvax.cis.ecu.edu/%7elmi/frames/prjects/report/index.html>
<http://ecuvax.cis.ecu.edu/%7elmi/frames/projects/report/append.html>
- Baldwin, J.A., Acheson, E.D., and Graham, W.J. 1987. *The textbook of medical record linkage*. New York: Oxford University Press.
- Bane, M. J. and D. Ellwood. 1983. Report to U.S. Department of Health and Human Services Assistant Secretary for Planning and Evaluation. *Dynamics of dependence: The routes to self-sufficiency*. Cambridge, MA: Urban Systems Research and Engineering, Inc. June.
- Barth, R. and B. Needell. 1997. From the simple to the sublime: Using performance indicators with child welfare policy managers. Working Paper. (No. 2, September). Chicago: Joint Center for Poverty Research.
- Becerra, R., A. Lewin, M. Mitchell and H. Ono. 1996. California work pays demonstration project: Interim report of first thirty months. Unpublished manuscript. UCLA: School of Public Policy and Social Work.
- Born, C. 1997. Working with administrative data: Lessons from the field. *Poverty Research News* 1 (no. 4 Fall). Chicago: Joint Center for Poverty Research.
- Bouman, J. 1997. Letter to advisory panel on research uses of administrative data. (September 25).
- Brady, H.E. and S. Luks. 1995. Defining welfare spells: Coping with problems of survey responses and administrative data. Paper presented at the American Political Science Association Meeting, August 30, 1995 in Chicago.
- Brady, H. E., M. Meyers, and S. Luks. 1996. The impact of child and adult disabilities on the duration of welfare spells. Paper prepared for the Eighteenth Annual Research Conference of the Association for Public Policy and Management, October 31-November 2, 1996 in Pittsburgh.
- Brady, H. and B. W. Snow, 1966. Data systems and statistical requirements for the personal responsibility and work opportunity reconciliation act of 1996. Unpublished manuscript. University of California: UC-DATA.
- California Bill Number AB-1542 (Welfare Reform)
- Caudill, P., and C. Born. 1997. Who is at greatest risk of reaching cash assistance time limits? University of Maryland, SSW, February.
- Council of Europe. 1997. *Recommendations on the protection of personal data collected and processed for statistical purposes*. Forthcoming.
- Elliott, E.J. and J. Ralph. 1997. Quality education data: Unprecedented opportunity for a decade to build. Paper prepared for the National Research Council, Committee on National Statistics, Panel on Performance Measures and Data for Public Health Performance Partnership Grants, National Center for Education Statistics, July in Washington, DC.
- Federal Committee on Statistical Methodology. 1980. Report on statistical uses of administrative records. Statistical Policy Working Paper 6, Office of Federal Statistical Policy and Standards, Washington, DC: USGPO.
- Fraker, T., R. Moffitt, and D. Wolf. 1985. Effective tax rates and guarantees in the AFDC program, 1967-82. *Journal of Human Resources*. 20: 251-63.
- Freedman, S. 1997. Evaluating welfare-to-work programs with administrative data. *Poverty Research News*, 1 (no.4 Fall). Chicago: Joint Center for Poverty Research.
- Goerge, R. M. 1990. The reunification process in substitute care. *Social Service Review*. 64 (3), 422-457.
- Goerge, R. M., T. E. Sommer, B. J. Lee, and A. G. Harris. 1995. Uses of multiple human services by Illinois children and families: The use of administrative data for research purposes. (June). Chicago: Chapin Hall.

- Goerge, R.M., Wulczyn, F. and Harden, A. 1995. Report to the U.S. Department of Health and Human Services. *An update from the multistate foster care data archive: Foster care dynamics 1983-1993*. Chicago: Chapin Hall.
- Gula, A., and C. King. 1989. Final report: Pilot study results—a post-program, follow-up approach for post secondary vocational education students. The University of Texas at Austin, Center for the Study of Human Resources.
- Heckman, J. 1992. Randomization and social program evaluation. *Evaluating Welfare and Training Programs*. Manski, C. and Garfinkel, I. (eds.). Cambridge, MA: Harvard University Press, 201–230.
- Heckman, J. and J. Smith. 1997. The sensitivity of experimental impact estimates: Evidence from the national JTPA study. Forthcoming in NBER volume.
- Hill, C., V. J. Hotz, C. Mullin and J. K. Scholz. 1997. EITC eligibility, participation and compliance rates for AFDC households: Evidence from the California caseload. Unpublished manuscript, University of Chicago, May.
- Hotz, V. J. 1992. Recent experience in designing evaluations of social programs: The case of the national JTPA study. *Evaluating Welfare and Training Programs*. Manski, C. and Garfinkel, I. (eds.). Cambridge, MA: Harvard University Press, 76-114.
- Hoynes, H. 1996. Local labor markets and welfare spells: Do demand conditions matter? Unpublished manuscript, UC-Berkeley, June.
- Illinois Public Aide Code (305 ILCS 5/124.33 new), Section 12-4.33.
- Information Technology Support Center (ITSC) (1997), State of Maryland, U.S. Department of Labor. www.itsc.state.md.us
- ISI Declaration. *International Statistical Review*. 54 (no.2, August 1986): 227-242.
- Jabine, T. B. 1993. Procedures for restricted data access. *Journal of Official Statistics: Confidentiality and Data Access*. 9 (no. 2): 537-589.
- Jaro, M.A. 1985. *Current record linkage research*. Proceeding of the Statistical Computing Section. American Statistical Association. 140-143.
- Jaro., M.A. 1989. Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *Journal of The American Statistic Association*. 84 (406) 414-420.
- King, C. T., and D. T. Schexnayder. 1988. *Welfare dynamics in Texas: An exploratory analysis of AFDC turnover and program participation*. The University of Texas at Austin, Center for the Study of Human Resources and Bureau of Business Research.
- King, C. T., and D. T. Schexnayder. 1992. *Coordination in Texas pre-jobs job training/ welfare programs: Final project report*. The University of Texas at Austin, Center for the Study of Human Resources and Bureau of Business Research.
- King, C. T., D. T. Schexnayder, D. O'Shea, J. A. Olson, D. Tse-i Pan, R. A. Roche, and R. D. Trunkey. 1994. *Texas jobs program evaluation: Final report*. The University of Texas at Austin, Center for the Study of Human Resources.
- Levine, D., editor. 1986. *Creating a center for education statistics: A time for action*. Washington, DC: National Research Council, National Academy Press.
- Lewis, Dan. Evaluating welfare reform in Illinois: A panel study of recipients and their experiences. Funding proposal. www.jcpr.org/consortium.html
- Luttrell, C. A. 1994. Simulating Cost Avoidance Using Results from a Labor Supply Model. A paper presented at the 36th Annual Conference of the National Association of Welfare Research and Statistics, Jackson Hole, Wyoming.
- Massachusetts Department of Revenue 1995. Massachusetts longitudinal database for research on social services. A proposal submitted to the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services, July 31.
- McBroom, P. 1997. News and events: www.berkeley.edu/news/features/11_24_97_clinton.html
- Moffitt, R. and D. A. Wolf. 1987. The effect of the 1981 Omnibus Budget and Reconciliation Act on welfare recipients and work incentives. *Social Service Review*. (June): 247-260.
- Moffitt, R. 1992. Evaluation methods for program entry effects. In *Evaluating welfare and training programs*. pp. 231-52. Edited by C. Manski and I. Garfinkel. Harvard University Press.
- Moffitt, R. 1993. The effect of work and training programs on entry and exit from the welfare caseload. IRP Discussion Paper No. 1025-93.

- National Academy of Science, Committee on National Statistics. 1993. *Private lives and public policies*. Washington, DC: National Academy Press.
- National Commission on Excellence in Education. 1983. *A nation at risk: The imperative for educational reform*. Washington, DC: Department of Education.
- National Institute of Mental Health. 1989. *Data standards for mental health decision support systems*.
- Newcombe. 1988. *Handbook of record linkage: Methods for health and statistical studies, administration, and business*. Oxford: Oxford University Press.
- Oregon Department of Justice. 1993. Opinion No. 8226, August 4.
- Oregon Employment Division, Department of Education, Department of Corrections, Bureau of Labor and Industries, Adult and Family Services Division, Vocational Rehabilitation Division, Department of Insurance and Finance, Office of Community College Services, Oregon State System of Higher Education, and Job Training Partnership Act Administration. 1992. Shared information system. Submitted to the Oregon Workforce Quality Council, May.
- Organization for Economic Cooperation and Development. 1980. *Guidelines governing the protection of privacy and transborder flows of personal data*.
- Schexnayder, D. T., and J. A. Olson. 1995. *Texas JOBS program evaluation: Second year impacts*. The University of Texas at Austin, Center for the Study of Human Resources.
- Schexnayder, D. T., and J. A. Olson. 1997. *Texas food stamp employment and training/JOBS conformance demonstration: Impact evaluation final report*. The University of Texas at Austin, Center for the Study of Human Resources.
- Schexnayder, D. T., C. T. King, and J. A. Olson. 1991. *A baseline analysis of factors influencing AFDC duration and labor market outcomes*. The University of Texas at Austin, Center for the Study of Human Resources and Bureau of Business Research.
- Schexnayder, D. T., C. T. King, and L. O. Lawson. 1994. *Patterns of participation in Texas welfare and training programs: How Hispanics differ from other race/ethnic groups*. Prepared for and also available from the National Commission for Employment Policy, Washington, D. C. The University of Texas at Austin, Center for the Study of Human Resources.
- Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget. 1994. Statistical policy working paper 22: Report on statistical disclosure limitation methodology. NTIS Document Sales, Springfield, VA: May.
- Stevens, D. W. 1994. Research uses of wage record data: Implications for a national wage record database. University of Baltimore.
- Stevens, D. W. 1996. Toward an all purpose confidentiality agreement: Issues and proposed language. University of Baltimore, The Jacob France Center.
- Texas State Occupational and Information Coordinating Committee. 1997. Automated student and adult learner follow up study final report. Austin, Texas.
- Texas State Occupational and Information Coordinating Committee. Annual reports. 1993-1997. Austin, Texas.
- U.S. Bureau of Labor Statistics. BLS commissioner's order number 3-92. Confidentiality of BLS records. BLS: August 18, 1993.
- U.S. Bureau of Labor Statistics. Administrative procedure number 196. Responsibility for safeguarding confidential information. BLS: April, 26, 1996.
- U.S. Bureau of Labor Statistics. Quality improvement project—Unemployment insurance wage records survey. Forthcoming.
- U.S. Congress. 1996. Personal Responsibility and Work Opportunity Reconciliation Act, Public Law 104-193, 104th Congress, August 22.
- U.S. Department of Health, Education, and Welfare, Privacy Protection Study Commission. 1977. *Personal privacy in an information society*.
- U.S. Government Accounting Office. 1994. *Managing for results: State experiences provide insights for federal management reforms*. GAO/GGD -95-22, December.
- U.S. Government Accounting Office. 1996. *Executive guide: Effectively implementing the Government Performance and Results Act*. GAO/GGD-96-118, June.
- U. S. Privacy Protection Study Commission. 1996. *Personal privacy in an information society*.

- University of Texas at Austin, Center for the Study of Human Resources. 1995. *An analysis of proposed Texas legislation on AFDC time limits*. Austin: UTCHR.
- University of Maryland School of Social Work and Maryland Dept of Human Resources. 1997. *Life after welfare: An interim report*. UMD SSW. September.
- White House, Information Infrastructure Task Force. 1995. *Principles for providing and using personal information*.
- Winkler, W.E. 1988. Using the EM algorithm for weight computation in the Fellegi-Sunter model of record linkage. *Proceedings of the Section Survey Research Methods, American Statistical Association*. 1-5.
- Wulczyn, F. and Goerge, R. M. 1992. Foster care in New York and Illinois: The challenge of rapid change. *Social Service Review*. 66 (2) 278-294.

URL'S OF INTEREST:

American Statistical Association Committee on Privacy and Confidentiality:

<http://www.erols.com/dewolf/pchome.htm>

American Public Welfare Association Public Welfare Directory (including listings for information systems and TANF administrators)

<http://www.apwa.org> (for ordering information)

APPENDIX I
**Examples of Successful Data-Sharing Laws
 and Agreements from:
 “Toward an All Purpose Confidentiality
 Agreement: Issues and Proposed Language”**

BY DAVID W. STEVENS, 1996

Examples of Successful Data-Sharing Laws and Agreements

Florida’s unemployment compensation law, Chapter 443, Paragraph 1715, titled, “Disclosure of Information: Confidentiality,” identifies a particular class of public employees who are authorized to be given access to wage records under a records and reports subheading:

Such information may be made available only to public employees in the performance of their public duties, *including employees of the Department of Education in obtaining information for the Florida Education and Training Placement Information Program and the Department of Commerce in its administration of the qualified defense contractor tax refund program.*

One way to approach the Distributed Wage Record Database interest in drafting a uniform data-sharing agreement would be to make explicit reference to authorized parties as they are identified in each state’s unemployment compensation law. This would leave responsibility for such designations at the state level, which is consistent with the current statutory authority for control of SESA’s administrative records.

Florida’s law, Chapter 443, Paragraph 1715, continues under a disclosure of information subheading, that

Subject to such restrictions as the division prescribes by rule, information declared confidential under this section may be made available to any agency of this or any other state, or any federal agency, charged with the administration of any unemployment compensation law or the maintenance of a system of public employment offices.

This exemplifies the type of state-specific language that can be crafted to give a SESAs executive director discretionary authority to approve or disapprove a particular request for access to administrative records that otherwise conforms to applicable Federal and State confidentiality stipulations. This is similar in intent to Ohio’s rule that appears in footnote 15 on page 11 [of the original article].

Illinois shared data agreements incorporate applicable Federal and State confidentiality provisions directly in the agreement document. For example, a data-sharing agreement between the Illinois Department of Employment Security and a Service Delivery Area (SDA) entity, includes Section 1900 Disclosure of Information (720 ILCS 405/1900, as amended by p.a. 88-435, effective August 20, 1993), which states in part that:

The Director may furnish any information that he may deem proper to any public officer or public agency of this or any other State or of Federal Government dealing with: 1. the administration of relief, 2. public assistance, 3. unemployment compensation, 4. a system of public employment offices, 5. wage and hours of employment, or 6. a public works program.

The Florida, Illinois and Ohio examples of discretionary authority are not enthusiastically endorsed by all SESA administrators for one obvious reason—some shy from having the responsibility to exercise discretionary authority, preferring instead to have authorized uses spelled out in statutory language that leaves no ambiguity of interpretation or reason for appeal.

This year, North Carolina’s General Assembly amended Chapter 96 of the State’s General Statutes to create a Common Follow-up System for State Job Training and Education Programs. The language found here is instructive for consideration of a uniform data-sharing agreement to be used in a Distributed Wage Record Database context.

96.30. Findings and purpose.

The General Assembly finds it in the best interests of this State that the establishment, maintenance, and funding of State job training, education, and placement programs be based on current, comprehensive information on the effectiveness of these programs in securing employment for North Carolina citizens and providing a well-trained workforce for business and industry in this State. To this end, it is the purpose of this Article to require the establishment of an information system that maintains up-to-date job-related information on current and former participants in State job training and education programs.

96.32. Common follow-up information management system created.

In developing the system, the Employment Security Commission of North Carolina shall ensure that data and information collected from State agencies is confidential, not open for general public inspection, and maintained and disseminated in a manner that protects the identity of individual persons from general public disclosure.

96.34. Prohibitions on use of information collected.

Data and information reported, collected, maintained, disseminated, and analyzed may not be used by any State or local government agency or entity for purposes of making personal contacts with current or former students or their employers or trainers.

The new North Carolina follow-up system illustrates how straightforward language can be crafted to accomplish a desired data-sharing goal. North Carolina’s Employment Security Commission retains strong discretionary authority over all aspects of confidential data release. The State’s Employment Security Law contains the following two paragraphs that grant this authority to the Commission:

Subject to restrictions as the Commission by regulation may provide, information from the records of the Employment Security Commission may be made available to any agency or public official for any purpose for which disclosure is required by statute or regulation. The Commission may, in its sole discretion, permit the use of information in its possession by public officials in the performance of their public duties.

Such a sweeping authority to exercise discretionary control over the release and disclosure of confidential administrative records would subject a Distributed Wage Record Database system to a high level of uncertainty about the continuity of any agreement with that SESA.

Oregon’s Labor and Industrial Relations, Unemployment Insurance, statute, 657.665, Confidentiality of information from employing unit records, begins with the typical shall be confidential” phrase, but then continues with the following language.”

(3) Notwithstanding subsection (1) of this section, information secured from employing units pursuant to this chapter may be released to agencies of this state,

and political subdivisions acting alone or in concert in city, county, metropolitan, regional or state planning to the extent necessary to properly carry out governmental planning functions performed under applicable law. Information provided such agencies shall be confidential and shall not be released by such agencies in any manner that would be identifiable as to individuals, claimants, employees or employing units.

This exemplifies a sensible approach to the balancing action described by the authors of *Private Lives and Public Policies*.¹ The Employment Department does not anguish about the release of personally identifiable records to a responsible third-party that seeks to use this information for planning purposes; it simply passes the confidentiality stipulation on to this external entity in its own handling and release of the data. Similar authority will be necessary in some form in each state statute if the Distributed Wage Record Database is expected to provide universal and routine coverage of the SESAs.

Finally, Washington's 1996 amendments of the State's unemployment compensation statute include the following pertinent paragraphs.

Governmental agencies may have access to certain records and information, limited to employer information possessed by the department for purposes authorized in chapter 50.38 RCW. Access to these records and information is limited to only those individuals conducting authorized statistical analysis, research, and evaluation studies. Only in cases consistent with the purposes of chapter 50.38 RCW are government agencies not required to comply with subsection (1)(c) this section [which requires informed consent steps for other uses of covered administrative records], but the requirements of the remainder of subsection (1) of this section must be satisfied.

... The employment security department shall have the right to disclose information or records deemed private and confidential under this chapter to any private person or organization when such disclosure is necessary to permit private contracting parties to assist in the operation and management of the department in instances where certain departmental functions may be delegated to private parties to increase the department's efficiency or quality of service to the public.

Again, this statutory language illustrates how particular third-party uses of wage records can be accommodated by amending state unemployment compensation laws.² So, to date, a

¹ It is important to think about the phrase "to the extent necessary to properly carry out governmental planning functions" that is imbedded in this statute. The author has witnessed many contentious debates over a 25 year period that share a common theme of mistrust in a third-party's ability and/or willingness to use SESA wage records in a responsible manner. Progress toward the establishment and maintenance of a Distributed Wage Record Database capability will be affected by the level of confidence that is reached within the SESA/ui community that external parties understand the characteristics of wage records and will not misuse this administrative information. To date, SESAs have absorbed substantial costs in attempts to accommodate third-party requests for the use of wage record, and other, administrative data.

² No one should conclude that it will be easy, or even possible, to devise similar or identical language that can be successfully incorporated in each state's unemployment compensation law. Timing, personalities and other controllable and uncontrollable factors enter into each state's decision about whether to even attempt to amend current law. What may seem logical and achievable in the abstract must be reviewed in the context of adversarial relationships and attitudes and emotions about individual privacy rights that have nothing to do with the particular issue of third-party use of SESA wage records. Most UI program administrators place a heavy weight on the notoriety and chilling effect dangers, and on skepticism that associated costs of cooperation will be fully compensated, while those who seek access to these records typically pay no

state-by-state review of successful and unsuccessful attempts to acquire and use wage record data leads to the inevitable conclusion that when affected parties want to reach agreement they find a way to do so; and, when one of the involved parties seeks to find a way to disagree, they are likely to succeed in identifying a statutory or rule basis for doing so. Thus far, the passage of time has favored those who want to find a way to reach agreement, because the number of successes and availability of accumulating information about how to proceed strengthens the hand of successors.

3.0 Recommendations

There is an ample number of examples of how amendments of state law, issuance of new rules and exercise of discretionary authority can be used to respond to selected third-party requests for use of wage record information. However, there are complementary examples of denials of other requests, and the population of individuals who truly understand what can and cannot be done with wage records alone remains small.

The following recommendations for action are offered based on the author's current understanding of Federal and state initiatives that affect, or are affected by, access to wage record data.

1) There continues to be an urgent need for some agreed-upon source of reliable information about the current status of state laws, rules, discretionary actions and results. Someone at the Federal level should act to reduce the costs of misinformation and lack of information as the frequency of state and local actions increases.

(2) Some authority should take responsibility for crafting and circulating a proposed all-purpose data-sharing agreement to find out exactly which SESAs are unable, or unwilling, to participate in a Distributed Wage Record Database, and for what explicit reasons. This should then be followed by an attempt to respond to the barriers, so interested parties will soon know whether it will ever be feasible to think about universal SESA participation in such an undertaking.

(3) Some attention should now be given to the results of pioneering data-sharing activities. To date, ad hoc sharing of information has been limited because of goals of each undertaking are somewhat different, and the managers of such programs are busy attending to their own responsibilities. A gulf between the education and employment & training worlds persists, which is unfortunate for all who are involved; each can learn from the other.

(4) The desire to "get on with the practical" should not be allowed to diminish continued interest in and attention to the basic issues of validity and reliability. It is important to keep asking "are we actually measuring what we want to measure?" And "are the measures we use adequate for the intended purposes?"

APPENDIX II
**Amendment to Illinois Public Code (305 ilcs-5/124.33 new)
To Provide Research Access to Information on Applications,
Terminations and Denials of Benefits under TANF**

AN ACT to amend the Illinois Public Aid code by adding Section 12-4.33.

Be it enacted by the People of the State of Illinois, represented in the General Assembly:

Section 5. The Illinois Public Aid Code is amended by adding Section 12-4.33 as follows:

(305 ILCS 5/12-4.33 new)

Sec.12-4.33.Welfare reform research and accountability.

(a) The Illinois Department shall collect and report upon all data in connection with federally funded or assisted welfare programs as federal law may require, including, but not limited to, Section 411 of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 and its implementing regulations and any amendments thereto as may from time to time be enacted.

(b) In addition to and on the same schedule as the data collection required by federal law and subsection (a), the Department shall collect and report on further information with respect to the Temporary Assistance for Needy Families (“TANF”) program, as follows:

(1) With respect to denials of applications for benefits, all of the same information about the family required under the federal law, plus the specific reason or reasons for denial of the application.

(2) With respect to all terminations of benefits, all of the same information as required under the federal law, plus the specific reason or reasons for the termination.

(c) The Department shall collect all of the same data as set forth in subsections (a) and (b), and report it on the same schedule, with respect to all cash assistance benefits provided to families that are not funded from the TANF program federal block grant or are not otherwise required to be included in the data collection and reporting in subsections (a) and (b).

(d) Whether or not reports under this Section must be submitted to the federal government, they shall be considered public and they shall be promptly made available to the public at the end of each fiscal year, free of charge upon request. The data underlying the reports shall be made available to academic institutions and public policy organizations involved in the study of welfare issues or programs redacted to conform with applicable privacy laws. The cost shall be no more than that incurred by the Department in assembling and delivering the data.

(e) The Department shall, in addition to the foregoing data collection and reporting activities, seek a university to conduct, at no cost to the Department, a longitudinal study of the implementation of TANF and related welfare reforms. The study shall select subgroups

representing important sectors of the assistance population, including type of area of residence (city, suburban, small town, rural), English proficiency, level of education, literacy, work experience, number of adults in the home, number of children in the home, teen parentage, parents before and after the age of 18, and other such subgroups. For each subgroup, the study shall assemble a statistically valid sample of cases entering the TANF program at least 6 months after its implementation date and prior to July 1, 1998. The study shall continue until December 31, 2004. The Department shall report to the General Assembly and the Governor by March 1 of each year, beginning March 1, 1999, with the interim findings of the study with respect to each subgroup, and by March 1, 2005, the final findings with respect to each subgroup. The reports shall be available to the public upon request. No later than November 1, 1997, the Department, in consultation with the advisory panel of specialists in welfare policy, social science, and other relevant fields shall devise the study and identify the factors to be studied. The study shall, however, at least include the following features:

- (1) Demographic breakdowns including, but not limited to, race, gender, and number of children in the household at the beginning of Department services.
- (2) The Department shall obtain permission to conduct the study from the subjects of the study and guarantee their privacy according to the privacy laws. To facilitate this permission, the study may be designed to refer to subjects by pseudonyms or codes and shall in any event guarantee anonymity to the subjects without limiting access by outsiders to the data (other than identities) generated by the study.
- (3) The subjects of the study shall be followed after denial or termination of assistance, to the extent feasible. The evaluator shall attempt to maintain personal contact with the subjects of the study, and employ such methods as meetings, telephone contacts, written surveys, and computer matches with other databases to accomplish this purpose. The intent of this feature of the study is to discover the paths people take after leaving welfare and the patterns of return to welfare, including the factors that may influence these paths and patterns.
- (4) The study shall examine the influence of various employability, education, and training programs upon employment, earnings, job tenure, and cycling between welfare and work.
- (5) The study shall examine the influence of various supportive services such as child care (including type and cost), transportation, and payment of initial employment expenses upon employment, earnings, job tenure, and cycling between welfare and work.
- (6) The study shall examine the frequency of unplanned occurrences in subjects' lives, such as illness or injury, family member's illness or injury, car breakdown, strikes, natural disasters, evictions, loss of other sources of income, domestic violence, and crime, and their impact upon employment, earnings, job tenure, and cycling between welfare and work.
- (7) The study shall examine the wages and other compensation, including health benefits and what they cost the employee, received by subjects who obtain employment, the type and characteristics of jobs, the hours and time of day of work, union status, and the relationships of such factors to earnings, job tenure, and cycling between welfare and work.

(8) The study shall examine the reasons for subjects' job loss, the availability of Unemployment Insurance, the reasons for subjects' search for another job, the characteristics of the subjects' next job, and the relationships of these factors to re-employment, earnings, job tenure on the new job, and cycling between welfare and work.

(9) The study shall examine the impact of mandatory work requirements, including the types of work activities to which the subjects were assigned, and the links between the requirements and the activities and sanctions, employment, earnings, job tenure, and cycling between welfare and work.

(10) The study shall identify all sources and amounts of reported household non-wage income and examine the influence of the sources and amounts of non-wage non-welfare income on employment, earnings, job tenure, and cycling between welfare and work.

(11) The study shall examine sanctions, including child support enforcement and paternity establishment sanctions, the reasons sanctions are threatened, the number threatened, the number imposed, and the reasons sanctions are not imposed or are ended, such as cooperation achieved or good cause established.

(12) The study shall track the subjects' usage of TANF benefits over the course of the lifetime 60-month limit of TANF eligibility, including patterns of usage, relationships between consecutive usage of large numbers of months and other factors, status of all study subjects with respect to the time limit as of each report, characteristics of subjects exhausting the eligibility limit, types of exceptions granted to the 60-month limit, and numbers of cases within each type of exception.

(13) The study shall track subjects' participation in other public systems, including the public schools, the child welfare systems, the criminal justice system, homeless and food services, and others and attempt to identify the positive or negative ripple effects in these systems of welfare policies, systems, and procedures.

(f) The Department shall cooperate in any appropriate study by an independent expert of the impact upon Illinois resident non-citizens of the denial or termination of assistance under the Supplemental Security Income, Food Stamps, TANF, Medicaid, and Title I social services programs pursuant to the changes enacted in the federal Personal Responsibility and Work Opportunity Reconciliation Act of 1996. The purpose of such a study must be to examine the immediate and long-term effects on this population and on the State of the denial or termination of these forms of assistance, including the impact on the individuals, the alternate means they find to obtain support and care, and the impact on state and local spending and human services delivery systems. An appropriate study shall select a statistically valid sample of persons denied or terminated from each type of benefits and attempt to track them until December 31, 2000. Any reports from the study received by the Department shall be made available to the General Assembly and the Governor upon request, and a final report shall be submitted upon completion. These reports shall be available to the public upon request.

Section 99. Effective date. This Act takes effect upon becoming law.

APPENDIX III

**Article 9. Evaluation of CalWORKS Program Implementation
California Bill Number AB-1542 (Welfare Reform)
AKA the Thompson-Maddy-Ducheny-Ashburn
Welfare-to-Work Act of 1997¹**

11520. The State Department of Social Services shall ensure that a comprehensive, independent statewide evaluation of the CalWORKS program is undertaken and that accurate evaluative information is made available to the Legislature in a timely fashion.

11520.3. The department shall develop a research design to ensure a thorough evaluation of the direct and indirect effects of the CalWORKS program. Effects shall include, but not be limited to, employment, earnings, self-sufficiency, child care, child support, child well-being, family structure, and impacts on local government. Child well-being shall include entries into foster care, at-risk births, school achievement, child abuse reports, and rates of child poverty.

11520.5. The statewide evaluation shall be conducted by an independent evaluator or evaluators. It shall represent a clear delineation of the research questions and shall, through discrete reports issued at regular intervals, provide information regarding process, impacts, and analyses of the costs and benefits of the CalWORKS program.

11520.7. The department shall ensure that county demonstration projects and other innovative county approaches to CalWORKS program implementation are independently and rigorously evaluated and that findings are reported to the Legislature in a timely fashion. The evaluation of a county-specific program shall be developed in conjunction with the county and other appropriate agencies responsible for the local program.

11521. By July 1, 1998, the department shall revise data collection procedures used for quality control and caseload characteristic studies in order to respond to the data collection requirements of Public Law 104-193 and state law. The department shall develop common data definitions to be used by the counties, design common identifiers, and, to the extent possible, standardize state and county data collection infrastructure. The department shall accomplish the requirements of this section in consultation with experts in monitoring and research, representatives of counties, the Legislature, and appropriate state agencies.

11521.3. Evaluation of CalWORKS program implementation conducted or commissioned by the department shall, to the extent practical, use or build upon existing welfare data archives, including, but not limited to, the data bases and research completed to date as part of the Work Pays Demonstration Project authorized pursuant to Chapter 97 of the Statutes of 1992.

11521.5. The department shall have access and authority to obtain for tracking, monitoring, research and evaluation purposes to data collected by counties on recipients receiving cash aid, in-kind payments, or supportive services. 11521.7. The department shall continue the evaluation of Cal-Learn and issue a final report to the Legislature by July 1, 2000.

Article 9.5. Interagency Data Development and Use

¹ The text of the entire bill can be found at: www.sen.ca.gov/www/leginfo/SearchText.html

11525.(a) The department shall establish procedures to provide timely access to information on CalWORKS families to counties and researchers in a manner that maintains confidentiality of data while making it possible to undertake ongoing monitoring, research, and evaluation.

(b) The department, with the cooperation of the University of California, shall establish a project to link longitudinal administrative data on individuals and families who are receiving benefits under the CalWORKS program, or have received benefits under the program within the last 10 years. (2) All data shall be made available to a university center with the capability of linking it with other appropriate data to allow for ongoing assessment of program impact. (3) The department shall ensure that information identifiable to individuals and families is removed so as to maintain strict confidentiality.

(4) The State Department of Health Services, the Employment Development Department, the Franchise Tax Board, the State Department of Education, and any other state or local governmental agency that collects information on aided families shall provide the department with the necessary data, if legally available.

Article 9.7. Role of the University

11526.(a) The Legislature hereby requests the Regents of the University of California to establish and administer a program or programs to support welfare research and evaluation of the CalWORKS program.

(b) It is the intent of the Legislature that the program or programs established by the University of California:

(1) Establish a sponsored grants program to provide funding for interested researchers to undertake studies on important welfare-related issues. These grants shall be applied only to research projects requested by representatives of state and local government entities.

(2) Establish one or more Bureau of the Census secure data sites to link census and administrative databases for ongoing research purposes.

(3) Use existing data archives to develop data sets appropriate for monitoring and evaluating the impacts of CalWORKS program implementation in California.

(4) Create and maintain public use data sets and make data available to researchers and members of the public to support welfare research and related human services research.

(5) Provide an ongoing capacity for supporting, conducting, and disseminating welfare policy research.

(6) Produce and maintain lists of researchers working with California welfare data or conducting research on public assistance in California.

(7) Review, edit, publish, and disseminate research and evaluation reports to state and local policymakers.

(8) Provide forums for the presentation of research findings and the discussion of research on welfare.

(9) Provide a location for welfare data archives and monitor ongoing funding for their upkeep.

APPENDIX IV
1997 Legislature
Florida
New Shared Information Statute

Section 44. Notwithstanding any general or special law to the contrary, the agencies of one or more local governments may establish a collaborative client information system. State agencies and private agencies may participate in the collaborative information system. Data related to the following areas may be included in the collaborative information system, although the system is not limited to only these types of information: criminal justice, juvenile justice, education, employment training, health, and human services.

Section 45.

- (1) The counties involved in the creation and administration of a collaborative client information system shall form a steering committee, consisting of representatives of all agencies and organizations participating in the system, to govern the organization and administration of the collaborative system. Each steering committee shall determine its procedures for governance of the organization, participation in the collaborative information system, and administration of the data in the system. Each steering committee also must develop a security policy to be followed by all agencies participating in the collaborative system to ensure the integrity of the data in the collaborative information system and to guarantee the privacy, to the extent possible, of all clients served by an agency that participates in the collaborative system.
- (2) Before sharing confidential information with other members of the information collaborative, each member of the steering committee shall sign an agreement specifying, at a minimum, the following information:
 - (a) What information each agency will share with the collaborative;
 - (b) How the information will be shared;
 - (c) How clients will be notified that an agency participates in the collaborative;
 - (d) Who in each agency will have access to the information;
 - (e) The purpose to be served by sharing the information;
 - (f) Assurances from each agency that it will maintain the confidentiality of the information as required by law; and
 - (g) Other information decided upon by members of the information cooperative.

Section 46. Notwithstanding any law to the contrary, an agency that participates in the creation or administration of a collaborative client information system may share client information, including confidential client information, with other members of the collaborative system as long as the restrictions governing the confidential information are observed by any other agency granted access to the confidential information. An agency that participates in a collaborative information system is not required to have a release signed by its affected clients before sharing confidential information with other members of the collaborative system. Section 47. An agency that receives moneys from a federal, state, or local agency is encouraged to participate in any collaborative client information system that is available within the service area of the agency.

Section 48. Except as otherwise provided herein, this act shall take effect July 1, 1997.

- (1) Florida Education and Training Placement Information Program.

- (a) The Department of Education shall develop and maintain a continuing program of information management named "Florida Education and Training Placement Information Program," the purpose of which is to compile, maintain, and disseminate information concerning the educational histories, placement and employment, enlistments in the United States armed services, and other measures of success of former participants in state educational and workforce development programs.
- (b) Any project conducted by the Department of Education or the workforce development system that requires placement information shall use information provided through the Florida Education and Training Placement Information Program, and shall not initiate automated matching of records in duplication of methods already in place in the Florida Education and Training Placement Information Program. The department shall implement an automated system which matches the social security number of former participants in state educational and training programs with information in the files of state and federal agencies that maintain educational, employment, and United States armed service records and shall implement procedures to identify the occupations of those former participants whose social security numbers are found in employment records, as required by Specific Appropriation 337A, chapter 84-220, Laws of Florida; Specific Appropriation 337 B, chapter 85-119, Laws of Florida; Specific Appropriation 350 A, chapter 86-167, Laws of Florida; and Specific Appropriation 351, chapter 87-98, Laws of Florida.
- (c) The department, in consultation with the Department of Corrections, shall utilize the Florida Education and Training Placement information Program to match the social security numbers of inmates with information in the files of local school districts, and state and federal agencies that maintain educational, employment, and United States Armed Forces service records. Upon request, the department shall provide the Department of Corrections with such information as is necessary to identify the educational histories, the city/intra-city area and school districts where the inmate was domiciled prior to incarceration, the participation in state educational and training programs, and the occupations of inmates confined to state correctional facilities.
- (d) The Florida Education and Training Placement Information Program must not make public any information that could identify an individual or the individual's employer. The Department of Education must assume that the purpose of obtaining placement information is to evaluate and improve public programs or to conduct research for the purpose of improving services to the individuals whose social security numbers are used to identify their placement. If an agreement assures that this purpose will be served and the privacy will be protected, the Department of Education shall have access to the unemployment insurance wage reports maintained by the Department of Labor and Employment Security, the files of the Department of Health and Rehabilitative Services that contain information about the distribution of public assistance, the files of the Department of Corrections that contain records of incarcerations, and the files of the Department of Business and Professional Regulation that contain the results of licensure examination.
- (e) The Florida Education and Training Placement Information Program may perform longitudinal analyses for all levels of education and workforce development. These analyses must include employment stability, annual earnings, and relatedness of employment to education.