UC DATA
University of California
2538 Channing Way #5100
Berkeley, CA 94720-5100
FAX (510) 643-0663

THE POVERTY RESEARCH PROGRAMS
OF
UC DATA

BARBARA WEST SNOW, PH.D
RESEARCH DIRECTOR

HENRY E. BRADY, PH.D.
UC DATA Director

SPRING 1996
THE POVERTY RESEARCH PROGRAMS
OF UC DATA
Barbara West Snow, Ph.D.
Henry E. Brady, Ph.D. - Director

Advances in research on poverty in the California and the United States require extended
data sources, broader and more sophisticated research techniques, and better access to and
utilization of research information. Therefore, the poverty research programs at UC DATA
collect, archive, and develop new datasets for poverty research as well as provide access to new
databases for researchers and generate research results for policy makers.

What is UC DATA?

UC DATA is the University of California's principal archive of computerized social
science information. It is also a producer of large-scale social science databases for the research
community. UC DATA utilizes local, state, national, and international data to support graduate
and undergraduate instruction in the social sciences and related fields, aids academic and non-
academic researchers, and helps clients, including business, industry, and government access and
reformat quantitative data. UC DATA serves as the regional center for distributing U.S. Census
data to academic institutions in California as part of the Census Bureau's State Data Center
program. UC DATA’s strength in combining datasets for policy analysis is exemplified by a series
of data books which use archived data to study the demography of ethnic groups. The current
series includes the Latina/Latino Public Opinion Data Book (1992), the Latino/Latina
Demographic Data Book (1993), and the Asian Demographic Data Book (1995).

Since public perceptions of welfare are often key driving forces in legislative and policy
change, UC DATA is the national repository of the California Field Poll, covering California
public opinion data since 1956. UC DATA also has access to the Los Angeles Times Poll data,
including such important polls as their "Prejudice in Southern California" survey.
UC DATA has unique strengths in its evaluation research capability, having worked with two state agencies to design and implement comprehensive program evaluations, the California Department of Social Services (CDSS) and the Division of Workers’ Compensation (DWC).

Two of its evaluation research studies employ random assignment research designs:

- The California Work Pays Demonstration Project (CWPDP) is testing whether or not a set of work incentive reforms improves the transition between welfare and work. A stratified random sample of 15,000 cases (10,000 experimental and 5,000 control) from the Aid to Families with Dependent Children population in four research counties is being followed for six years. Experimental cases are given a new package of welfare reforms, and control cases retain the AFDC rules in effect as of September 1992.

- The Cal-Learn Demonstration Project is testing the importance of two programmatic features to the subsequent birth and educational outcomes of teen parents on AFDC. Employing a factorial research design to test the importance of case management services and/or financial incentives on outcomes, UC DATA and CDSS are collecting data on teens new to AFDC in the same four research counties. Randomization is based on the last digits of teens’ social security numbers.

- In the Workers’ Compensation Information Project, UC DATA helped develop a research and policy information system for the Division of Workers’ Compensation. This was designed to link administrative, survey, and other data sources to answer policy questions. A separate project in the Vocational Rehabilitation area has similar goals.

In addition to social program evaluation, UC DATA also has important expertise in the area of management information systems. Specifically, UC DATA:

- designs data collection forms, screens and system documentation,
- develops comprehensive administrative and research information systems,
- constructs functional interfaces between administrative datasets so they can be used in combination to answer complex policy questions,
- is a leader in social science computing with its on-line codebook project,
- helped developed the largest digital library on the Internet with census CD-ROM’s,
- has the capability to identify at-risk populations at fine geographic detail,
can identify at-risk populations by census tract, labor market trends by county, and
transfer payments (Bureau of Economic Analysis - BEA) by county,
is on the census Survey of Income and Program Participation (SIPP) advisory
committee of the Association of Public Data Users (APDU),
has become a leader in the conversion of administrative databases for research use.

California Welfare Data

Like many states, California has initiated programs to meet the requirements of the Family
Support Act of 1988, and has designed and implemented reforms of its welfare system through a
process of legislative action followed by the procurement of federal waivers. In recent legislation,
California has authorized strategies designed to encourage welfare recipients to work and to
become self-reliant, to discourage welfare dependency and child bearing while on welfare, to
make work pay, as well as to assure that teen parents on AFDC complete their high school
educations and reduce their incidence of repeat births. UC DATA is currently collecting data for
an evaluation of these and other changes in California's welfare rules.

Since 1992 UC DATA has worked with the California Department of Social Services, not
only to design research and evaluations for welfare reforms in the state, but also to document the
dynamics of family poverty and welfare use in California. The California Work Pays
Demonstration Project collects detailed, longitudinal data for several inter-related samples of low-
income families receiving welfare in California between 1987 and 1998. These data are being
linked with other administrative datasets and being made available on CD-ROM's to scholars and
analysts with interests in welfare, poverty, employment and family research. One percent and ten
percent, person and case samples drawn each year from Medi-Cal administrative records, together
with updates of activity from prior years, currently comprise longitudinal public use databases
with over five million records.
As part of the CWPDP, panel studies have been implemented to promote understanding of the experiences of different ethnic and family groups on AFDC and their transitions to marriage, employment, and to different kinds of public assistance. Computer Assisted Telephone Interviews (CATI) are being conducted every eighteen months with English (1731), Spanish (483), Vietnamese (504), Cambodian (383), Armenian (349), and Laotian (115) research subjects. Through these efforts, UC DATA is increasingly becoming a university center for welfare and poverty research.

Interrelated research designs and data collection activities have been devised by social scientists at UC DATA and CDSS to assure that data useful for evaluation (and other) research tasks may be gleaned from datasets which are compiled by administrative agencies in California during the ordinary course of their operations. Methods of dataset triangulation help assure that the differing strengths and weaknesses of each dataset individually can serve as checks on their accuracy and comprehensiveness in combination.

Medi-Cal records are being used to analyze welfare spells, as well as transitions between aid categories over time. Administrative records from the AFDC systems of four research counties have been organized and combined into a “County Welfare Administrative Database (CWAD)” containing payments information (and other variables) on cases involved in California’s welfare reform experiments. Case file records of research subjects are being coded, both to validate the data found in computerized administrative records, and to collect information on key variables found only there (such as the value of resources, automobiles, and restricted bank accounts). Extensive CATI telephone interviews provide self-report information about data found in other sources on the same AFDC case, and also allow in-depth information to be collected on smaller samples. Administrative datasets from the Employment Development Department (EDD) are currently being processed for validation of employment histories, as well as to provide primary information on employers of AFDC recipients. Data from the Franchise Tax Board will be analyzed for recipient participation in the Earned Income Tax Credit (EITC) Program.
Data will be collected for several more years with a longitudinal research design format, augmenting the already powerful multi-source information bank on welfare clients. Datasets, once cleaned for consistency over time, stripped of identifiers, and carefully documented by a research team at UC DATA, are being released to the research community as data sources for additional studies.

**Administrative Data and Academic Research**

The marriage of traditional research data collection methodologies, with the development and use of administrative datasets for research purposes, results in a significant increase in the capacity of researchers to conduct high quality studies, as the cost and time associated with data collection is significantly reduced. If funding for social science research is scaled back, administrative data collection may become one of the few available cost effective methods to bring interesting new datasets to researchers. Administrative data may be the only source for such sensitive data as whether a welfare applicant is an undocumented immigrant (Clune and Peete, 1995). Moreover, because the data are collected by states and local governments, local policy makers have an intrinsic interest in seeing whether their own data can be used to ascertain if programmatic changes are working in their own communities.

The Department of Health and Human Services summarized the strengths of administrative longitudinal databases in its 1995 Request For Applications (RFA) to support state efforts to link case-level administrative data across multiple low-income assistance programs. They state, “Recent state efforts to link longitudinal, administrative data across programs have proven extremely successful. Linked databases have provided a more thorough understanding of many aspects of both program participation and the characteristics of individuals who receive benefits from multiple anti-poverty programs. State-supported efforts have also provided valuable insight into both inter- and intra-state variations in program participation. Much of this information would not have been accessible through national panel data (RFA June 14, 1995).”
Current research on poverty and welfare spells often involves secondary data analysis of four important datasets: the Panel Study of Income Dynamics (PSID), the Current Population Survey (CPS), the National Longitudinal Survey of Youth (NLSY), and the Survey of Income and Program Participation (SIPP). Yet only a small subset of these population samples meet various poverty criteria. By the time AFDC recipients are selected, the number of cases upon which analyses are conducted is usually quite small. For example, the PSID may contain 450 cases on welfare. After seven years, this cohort is reduced to 50 cases.

In contrast, administrative databases provide information about very large samples or an entire universe of program participants rather than about small subsets. Having information about large numbers of program participants is particularly advantageous when researchers desire to stratify their samples by select characteristics (i.e. by two parent welfare families). Although the survey format is methodologically superior in terms of its large number of potential variables, controlled administration of questions, and easier implementation of changes, stratification can be problematic when using survey panel data because of small sample and/or cell sizes. More reliability and accuracy can be established with administrative data than with survey data, particularly with respect to payments, income, and the receipt of services. There is no dependence on participants’ memories or interpretations of questions about income or program participation.

Administrative data are often collected monthly over years, whereas other widely used datasets are updated annually, are cross-sectional, or are collected quarterly for a three year period (the SIPP). Because of the program turnover that may be masked in yearly data, spells on and off welfare can not be fully captured by using datasets based on survey panels. Since poverty research with a focus on public policy is often concerned with analyzing program duration and the movement of recipients to and from social programs, the availability of monthly datasets is extremely valuable.
Usefulness to Policy-Relevant Research on Poverty

The public use datasets developed by CDSS and UC DATA are proving extraordinarily useful for poverty research. What follows are some major research areas that are being addressed through linked administrative databases.

**Tax Expenditures and Social Policy.** Select provisions in the federal tax system provide financial incentives or assistance to some low-income households. In recent years, in particular, provisions in the tax codes such as the EITC have become an increasingly important means of providing the working poor with work incentives. Studies which examine the intended effects of various tax provisions on low income households are in short supply, partially because of the dearth of available data in this area. Merging the California Longitudinal Database from the Medi-Cal administrative file with records from the Employment Development Department will allow researchers to determine the effects of the EITC on work efforts and on the economic well-being of program participants. Studies in this area focus on:

- design and compliance issues (George Yin - University of Virginia Law School),
- patterns of eligibility and participation in the EITC among recipients of different forms of public assistance (Henry Brady - Graduate School of Public Policy and Department of Political Science UCB and Susan Ayasse - UC DATA),
- labor supply issues: Whether employers lower wages in response to higher EITC (Nada Eissa - Dept. Of Economics - University of CA - Berkeley),
- the dynamics of welfare and EITC as the incentives change (Joe Hotz - School of Public Policy, University of Chicago); and,
- EITC participation and transfer program non-compliance (Karl Scholz - Department of Economics, University of Wisconsin - Madison).

All of these studies will use datasets developed in California by CDSS and UC DATA. A key new addition to this series of studies will be the availability of tax information for research
subjects through the California Franchise Tax Board. Also, related questions were added to Wave II of the California Work Pays Demonstration Project AFDC recipient survey on when tax returns were last filed and whether or not a paid preparer was used (EITC receipt is associated with the use of paid tax preparers). Expenses for graduate student researchers and travel are being provided for each of the professors and researchers working on the EITC project through an ASPE grant to the California Department of Social Services.

**Participants’ Characteristics.** In many states, historical data on the characteristics of families on public assistance are spotty and often do not include information on factors of policy concern. In California, for example, the Department of Social Services periodically assembles information on AFDC, food stamps, and related program recipients’ characteristics by selecting a cross-sectional sample of participants to survey. Existing information paints a picture about the parents and childrens’ educational level, family size and share of families reporting earnings. With archival datasets in hand, it will be possible to obtain a more comprehensive and accurate picture of recipients’ characteristics and how these vary across participating states.

Jane Mauldon (Graduate School of Public Policy - UC Berkeley) is using Wave I of the CWPDP AFDC survey information to assess the extent of hunger, the use of food banks, and geographic mobility due to the inability to pay rent among AFDC recipients in CA, through a small grant from the Institute for Research on Poverty, University of Wisconsin - Madison. Subgroup characteristics may be linked to several other datasets with specific program information.

Richard Barth (Department of Social Welfare - UC Berkeley) will use AFDC datasets developed by UC DATA to explore the relationship between infant mortality and poverty. To answer complex questions on teen parenting, foster care, and juvenile incarceration, Dr. Barth is also assembling administrative data from the child welfare and foster care system with data from vital statistics and the California Department of Youth Authority (CYA).
**Patterns of Program Participation.** Historically, much of the interest among researchers has centered on the effect of recipient characteristics, program characteristics such as benefit levels, and labor market conditions on the probability that a family will enter or exit the welfare rolls (Albert, 1988). Frequently researchers have relied on PSID annual data to answer questions about the correlates of welfare entry and exit. Because of the turnover that can be masked in yearly data such as PSID data, spells on and off welfare can not be fully captured.

In the next five years, spells on and off welfare are likely to change dramatically in response to major policy shifts. Under the proposed block grant system, states may be given greater freedom in establishing their means-tested programs and in lowering their administrative costs. Some states may elect to make major changes in welfare eligibility and benefit standards. Cost shifting and movement to and from AFDC, and other social programs commonly used by welfare recipients, are likely to change in new ways. In order to assess the extent to which cost-shifting occurs, and changes in patterns or receipt of means-tested programs, linkages must be made between multiple databases for samples of the target populations. By using monthly data from multiple sources, researchers will be better able to understand the experiences of participants across several programs.

Marcia Meyers (Center for Policy Research - Syracuse University) is using datasets developed by CDSS and UC DATA to explore the dynamics of disability, including transitions between Social Security Income (SSI) and AFDC; and the extent of disabilities that prevent employment among AFDC recipients, including the disabilities of children. Dr. Meyers is using the Medi-Cal Longitudinal Database and the AFDC recipient surveys. Her research has been funded by DHHS and a new grant from CDSS to UC DATA.

CA Work Pays Principal Investigator, Henry Brady, and Samantha Luks (Department of Political Science - University of California, Berkeley) are addressing operational definitions by comparing survey and administrative data on the key outcome measure of welfare spell length. Professor Brady and Samantha Luks are also working with Marcia Meyers on a study of spell
duration among disabled adults and families with disabled children. This work uses a combination of Medi-Cal administrative records and survey results from AFDC recipient panels.

Jane Mauldon and Carol Chetkovich (Graduate School of Public Policy - University of California, Berkeley) are exploring the knowledge, attitudes, and practices of AFDC recipients with respect to birth control and the utilization of family planning services in California. Survey information is being used to stratify focus group participants along specific dimensions to more fruitfully explore patterns of family planning program participation. The Reproductive Health Policy Study is being funded by a grant from the Office of Family Planning and the California Department of Social Services.

Work and Welfare. In the past decade states have used several means to promote work on the part of public assistance participants, which include financial incentives, work requirements, training and education. The anticipated development of JOBS program databases as public use files will enable researchers to determine the extent to which these tools have increased the work efforts of participants. The availability of information about type of services and service mix used to enhance worker’s skills will enable researchers to gain much insight into utilization of these services and their overall effectiveness.

UC DATA produced a specialized dataset with zip code information, constructed from the Medi-Cal 1% Cases Longitudinal Database for Hilary Hoynes (Department of Economics - UC Berkeley), to study the strength of local area labor markets and their relationship to neighborhoods, ethnicity, and the duration of welfare spells. This research is being funded by the National Science Foundation.

Paul Ong (Chair of the UCLA Graduate School of Architecture and Urban Planning), has completed two monographs based on the Work Pays Demonstration Project Wave I AFDC English-Spanish survey: "Subsidized Housing and Work Among Welfare Recipients," and "Work and Car Ownership Among Welfare Recipients." Both papers address job access. Professor Ong
found that both car availability and the geographic freedom engendered by Section 8 housing certificates have strong relationships to the realization of employment. His research is being funded by a grant from the Urban Institute.

Implementation and Results of Welfare Policy Changes. Process studies of the implementation of specific welfare reforms have been conducted by UC DATA in four research counties (Alameda, San Joaquin, San Bernardino, and Los Angeles) since 1993. These studies employ qualitative research techniques to document how statewide legislative reforms result in actual changes at the county level, where AFDC is administered in California, and at the level of "street-level bureaucrats." One special team is assigned to investigate the implementation of "Work Pays" incentives like work rule, resource limit, income disregard, and benefit level changes. Another is assigned to document the implementation of the Cal-Learn program for pregnant and parenting teens.

Impact information collected by UC DATA is being analyzed by the Institute for Child and Family Policy Studies at the University of California - Los Angeles, under a separate Interagency Agreement with CDSS. Professor Rosina Becerra (Department of Social Welfare - UCLA) is Principal Investigator of the CWPDP evaluation. Details of the reforms being evaluated and the research designed to collect evaluation information is appended. An interim report covering the first two years of California's welfare reform effort is expected in April, 1996.

Professor Michael Wiseman (Institute for Research on Poverty - University of Wisconsin, Madison) is analyzing longitudinal administrative data from CA237 forms collected since 1973 under a subcontract with UC DATA. Dr. Wiseman is tracking changes in accessions and terminations from AFDC given major historical changes in welfare policy, especially the Omnibus Budget Reconciliation Act of 1981. OBRA eliminated the $30 and a third income disregard and reduced the need standard for AFDC program eligibility resulting in an increase of terminations from welfare. A summary of his preliminary findings is appended.
Administrative Databases for Research

UC DATA is linking several administrative databases in order to gain information about clients' participation in multiple social programs. The resulting databases in combination can paint a detailed picture of participants' varied experiences over a lengthy study period, providing information much richer than any single database used in isolation.

While earlier exploratory efforts assessed whether or not interfaces between different administrative datasets (Food Stamps, Medicaid, Child Support) already existed, UC DATA has made great strides in constructing such interfaces for research purposes in California. In addition to creating public use datasets of a 1% and 10% sample of Medi-Cal persons and cases records from 1987 to 1994, longitudinal case records have been constructed for the California Work Pays Demonstration Project's sample of 15,000 welfare recipients from San Bernardino, San Joaquin, Los Angeles, and Alameda counties. Because California has a county-operated welfare system, information for sample cases must be drawn from county administrative records in order to link payments information by case. UC DATA has successfully constructed linkages of administrative records from diverse datasets, and is now able to track administrative activity for the CWPDP sample through different programs and employment experiences (using survey with wage and employer data from the Employment Development Department).

UC DATA believes that this model exhibits the characteristics necessary for policy research on a diversity of programs and questions in the next decade. Some important questions can only be answered with multiple administrative datasets, such as whether changes in AFDC benefit levels or a time limit on AFDC is likely to result in cost shifting to other programs such as child welfare, foster care, SSI, etc. In order to assess the extent of cost-shifting, linkages must be made between multiple datasets for samples of the target population.
Public Use Dataset Development Process

UC DATA’s approach to developing, documenting, archiving, updating, and establishing public access to administrative data for research purposes is described below.

**Dataset Creation.** Useful datasets for research and policy analysis can be obtained from many sources including individual level administrative and survey data as well as aggregate administrative data. Data are being collected from state and county administrative data bases and through several ongoing surveys of welfare recipients. Information from the surveys, the county level administrative data, and the state level administrative data is combined into a research database with a large number of observations (cases or persons), with many variables recorded on each case or person, and with repeated observations over time. These datasets are being explored for answers to a broad range of questions about welfare.

For example, data from the California State Longitudinal Database has been used to show how the lengths of welfare spells differ for older and younger AFDC recipients depending upon the definition of a “break” in welfare experience (Ayasse, 1995). This analysis raises interesting questions about the degree to which these breaks represent administrative churning or attempts by recipients to leave welfare. Los Angeles County administrative data was used to compare the AFDC benefits received by families headed by citizens with families headed by non-citizens, but with citizen children. This analysis is part of a larger project to determine the costs to welfare of immigrants and their children (Clune and Peete, 1995). Aggregate and contextual data about welfare, welfare budgets, total employment, population, and migration can also be linked to individual level survey or administrative data or used alone to develop useful analyses of welfare experiences.

**Documenting Data, Sources, and Limitations.** Obtaining data is only the beginning in preparing it for public use and dissemination. Documentation must be prepared on the source of the data, the population it represents, and the definition of each data element. If the data are
collected over time or over different geographic units, then contextual information must be supplied on changes and differences in welfare regulations and programs.

Monthly case files are created by matching the case numbers of each case with the social security numbers of the individuals in the case. Once accomplished, these files contain an ample source of data for longitudinal and cross-sectional analysis, but experience has shown that the files are not yet ready for analysis. Since administrative data of this type is primarily collected for purposes of managing caseload, only in cases where a claim has been made does the institution have an opportunity to examine the validity of the information contained in the records. In addition, because of changing departmental priorities, data codes sometimes change.

Each record is rigorously examined and demographic variables, such as ethnicity, date-of-birth, and gender are recoded to eliminate inconsistencies. The data are compared to the available documentation and discrepancies reconciled. This resolution of coding discrepancies requires a considerable amount of time, interaction, and cooperation with knowledgeable individuals.

For all data elements, values are checked to be sure they are consistent with their definitions; frequency distributions are examined for anomalies; and instances of missing data values are identified and explained. Variables are cleaned to insure the highest data quality. For datasets developed from multiple sources, definitions across these sources are explained and reconciled. (UC DATA is developing a County Welfare Administrative Database across four California counties in which every effort is being made to insure that important variables, such as AFDC earnings, are similar from county to county.) Once the files and their corresponding documentation are prepared, those developed for public use are examined by a review group of outside researchers for ease of use, validity of the data, and relevance of the data elements. Changes in both files and documentation are made based on feedback from these researchers.

Archiving and Preserving Confidentiality. For archiving, a summary of each database is prepared, and the dataset is indexed and catalogued. For datasets whose original form included
confidential information, such as names and social security numbers, anonymous identifiers are prepared and a carefully guarded roster that defines linkages between the two is kept so that information received later (as subsequent aid histories, work histories, or episodes in the criminal justice system) can be matched and added to the existing information.

**Public Use and Access.** Multiple avenues must be utilized in order to widely distribute public use datasets and encourage their use in research and public policy evaluation. In addition to the traditional methods of providing mainframe tape copies and printed documentation, mastering and printing CD-ROMs enables placement of more data into a smaller storage space. This frees researchers from the confines of mainframe computing and allows analysis to be done on personal work stations. The availability of research databases are announced at Research Advisory Committee meetings, on a World Wide Web site developed by UC DATA, and at special presentations to legislative staff and welfare policy-makers in California. Some datasets will be made available over the Internet.

**Summary.** Steps used by UC DATA to prepare administrative datasets for research use are listed below:

- Transfer the data files from 9-track tape to UNIX
- Verify contents of files
- Assign longitudinal tracking identification number
- Develop linked history file
- Clean, recode, and standardize data elements
- Strip confidential information
- Create files in universal file format
- Conduct analysis of files
- Produce and validate subsamples
- Prepare documentation
- Produce public use files on appropriate media
- Conduct data integrity testing of public use files
- Deliver public use files to Advisory Group for review
- Receive comments and incorporate possible changes
- Archive and distribute public use datasets

### Datasets Under Development at UC DATA

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Alpha Version</th>
<th>DSS, Eval. Version</th>
<th>Public Use Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wave I E-S AFDC</td>
<td>November 94</td>
<td>April 95</td>
<td>January 96</td>
</tr>
<tr>
<td>2. Wave I FL AFDC</td>
<td>July 95</td>
<td>November 95</td>
<td>January 96</td>
</tr>
<tr>
<td>4.1 QC Survey I</td>
<td>July 95</td>
<td>November 95</td>
<td>January 96</td>
</tr>
<tr>
<td>4.2 QC U-Earnings</td>
<td>September 95</td>
<td>November 95</td>
<td>January 96</td>
</tr>
<tr>
<td>4.3 QC FG-Earnings</td>
<td>July 96</td>
<td>July 96</td>
<td>October 96</td>
</tr>
<tr>
<td>5.1 CWPDP LDB per</td>
<td>1987-92 Nov. 94</td>
<td>1987-94 Nov. 95</td>
<td>1987-94 Dec. 95</td>
</tr>
<tr>
<td>5.2 CWPDP LDB cas</td>
<td>1987-92 Nov. 94</td>
<td>1987-94 Nov. 95</td>
<td>1987-94 Dec. 95</td>
</tr>
<tr>
<td>6.1 LDB 10% person</td>
<td>1987-92 Nov. 94</td>
<td>1987-94 Nov. 95</td>
<td>1987-94 Nov. 95</td>
</tr>
<tr>
<td>6.2 LDB 10% cases</td>
<td>1987-92 Nov. 94</td>
<td>1987-94 Nov. 95</td>
<td>1987-94 Nov. 95</td>
</tr>
<tr>
<td>6.3 LDB 1% person</td>
<td>1987-92 Nov. 94</td>
<td>1987-94 Nov. 95</td>
<td>1987-94 Nov. 95</td>
</tr>
<tr>
<td>6.4 LDB 1% cases</td>
<td>1987-92 Nov. 94</td>
<td>1987-94 Nov. 95</td>
<td>1987-95 Nov. 95</td>
</tr>
<tr>
<td>7. CWPDP MED Pay</td>
<td>November 95</td>
<td>November 95</td>
<td>January 96</td>
</tr>
<tr>
<td>8. CWPDP EDD - P</td>
<td>November 95</td>
<td>November 95</td>
<td>No Public Use</td>
</tr>
<tr>
<td>9.1 10% LDB EDDP</td>
<td>December 95</td>
<td>December 95</td>
<td>No Public Use</td>
</tr>
<tr>
<td>9.2 10% LDB EDDC</td>
<td>November 95</td>
<td>November 95</td>
<td>No Public Use</td>
</tr>
<tr>
<td>9.3 1% LDB EDD-P</td>
<td>December 95</td>
<td>December 95</td>
<td>No Public Use</td>
</tr>
<tr>
<td>9.4 1% LDB EDD-C</td>
<td>November 95</td>
<td>November 95</td>
<td>No Public Use</td>
</tr>
<tr>
<td>10. Combination CD</td>
<td>November 95</td>
<td>January 96</td>
<td>January 96</td>
</tr>
</tbody>
</table>
1. Wave I AFDC English/Spanish Survey

2. Wave I Foreign Language Survey

3. County Welfare Administrative Databases

4. First 18 Months of Data Validation Case Coded Files:
   4.1 AFDC English/Spanish Survey Respondents
   4.2 Earned Income U Parent Cases
   4.3 Earned Income FG Cases

5. CWPDP Longitudinal Database Medi-Cal Spells Records
   5.1 Original and Replacement Sample Persons
   5.2 Original and Replacement Sample Aggregate

6. Statewide Longitudinal Database Medi-Cal Spells Records
   6.1 Ten Percent Persons
   6.2 Ten Percent Cases
   6.3 One Percent Persons
   6.4 One Percent Cases

7. CWPDP Medi-Cal Payments - 30 months

   (Persons in Cases)

   9.1 Ten Percent Persons File
   9.2 Ten Percent Cases File
   9.3 One Percent Persons File
   9.4 One Percent Cases File

10. Combination CD ROM: County Welfare Administrative Databases, CWPDP Longitudinal
    Samples, and QC Case-coded E/S Wave I Survey Sample
THE WORK PAYS DEMONSTRATION PROJECT

Summary. The California Work Pays Demonstration Project--a collaboration between the Research Branch of the State Department of Social Services (DSS), University of California Data Archive and Technical Assistance (UC DATA), and the Survey Research Center (SRC) at the University of California, Berkeley--collects and integrates data on the individual, household, welfare and employment characteristics of Aid to Families with Dependent Children (AFDC) recipient households, during and after their participation in the welfare system. The project will integrate individual-level data from four separate sources:

1. State-level administrative records for AFDC, Medicaid, Unemployment Insurance, and other state and Federal assistance programs;
2. County-level administrative records for the AFDC and Food Stamp programs;
3. Nonautomated client records maintained at county welfare offices; and,
4. Telephone interviews with AFDC recipients.

By combining these various databases and documenting them for public use, it will be possible for researchers to obtain unusually detailed information about the characteristics and economic behaviors of low-income families and children in California. Researchers will also be able to examine other family outcomes, such as income and family stability, and their relation to household characteristics and involvement in state and Federal assistance programs. Transfer payment and earnings data will be available for researchers interested in cost comparison and cost-benefit analyses of AFDC programs.

Legislation. State legislation under the Assistance Payment Demonstration Program, renamed the Work Pays Demonstration Program, mandated specific statewide reforms in California's welfare program. The legislation provided authority to the Department of Social Services to make programmatic changes in AFDC. The following changes became effective December 1, 1992:

1. the maximum aid payment was reduced by a total of 5.8 percent; and
2. the 100 hour work rule was waived for unemployed parent cases. (It was not in effect for family group (FG) cases). This means that for both family group (FG) and unemployed parent (U) cases, individuals with earned income can work over a hundred hours in any given month and not automatically be disqualified for aid.
Additional changes were made in California's welfare laws in July of 1993 when Senate Bill 35 passed. These included additional work incentives:

(3) rescinding the four month limitation on the $30 and 1/3 income disregard, meaning that people can work on welfare and not have their grants reduced dollar-for-dollar; rather, they can keep the first $30 and a third of their income before aid reductions ensue;

(4) increasing the need standard--the Minimum Basic Standard of Adequate Care (MBSAC) to allow income produced from employment; and,

(5) further reducing the maximum aid payment, to a total of 12.9% (including the cuts made previously).

Other SB 35 changes provided:

(6) a child care supplement to AFDC recipients who work, applicable to everyone;

(7) expansion of the Greater Avenues for Independence (GAIN) program by 71 percent to allow 36.6 percent more AFDC recipients to be served. A "one time through," provision was also added to the program, and a one-time only exemption for having a child under the age of three years;

(8) changes in resource limits: The equity value limit for an automobile increased from $1,500 to $4,500; allowable resources increased from $1,000 to $2,000; and savings accounts for specialized purposes became possible: up to $5,000 for childrens' college educations, for a down payment on a home, or for starting a business. (Resource limit changes do not apply at eligibility determination, for which the old rules still apply);

(9) an alternative assistance feature, whereby people who qualify for welfare do not have to receive a grant check if they don't want one, but rather can receive only the Medi-Cal insurance coverage, child care assistance, and other help instead;

(10) a program for pregnant and parenting teens called Cal-Learn, featuring incentives to return to or stay in school, disincentives to bad grades or dropping out, and the provision of child care, transportation, and ancillary assistance necessary to support schooling; and case management services through a specialized agency known as the Adolescent Family Life Program (AFLP).

A Federal increase in the Earned Income Tax Credit provided an additional work incentive.
Research Design  When the first two welfare changes were made, UC DATA was asked to design and implement a series of different data collection strategies to provide information for an evaluation study with an experimental research design. Accordingly, 15,000 cases on AFDC in four counties (Alameda, Los Angeles, San Bernardino, and San Joaquin) were randomly selected for the study. Ten thousand were placed in an experimental group which, like the rest of the welfare population, was subject to the rule changes, and five thousand were kept on the rules that were fixed as of September of 1992. For them it was as if the welfare law never changed. In September of 1991, prior to randomization, all AFDC recipients had sustained an initial 4.4% cut.

Because of case turnover, though an initial sample is drawn with a certain number of people in it, many leave welfare as time goes by, so the original sample resembles the actual welfare population less and less over time. Therefore, in addition to the base samples, UC DATA adds new sets monthly of freshly drawn replacement cases from DSS. Though the U--or unemployed parent--cases (which are two parent families) are only about 5-7% of the welfare caseload, they have been oversampled for the WPDP experiment because as a group they are expected to be more responsive to the work incentive features of the program. One third of the WPDP sample was drawn with this aid code. People in this type of case tend to be employed more and to leave welfare at a faster rate. Hence, in order to keep up with the attrition in the sample, more cases have to be replaced than expected, given the overall average that after five years, 50% of the cases once on welfare will no longer be on.

The research counties were selected based on geography and population type, with one urban county from the northern and southern parts of the state and one basically rural county from the northern and southern parts of the state. A formula was created to project what happens in these counties to the state population as a whole. (Los Angeles County is weighted as if it was two-thirds of the state.)

After SB 35 passed, data collection for the additional changes were folded into the existing research study. With the exception of Cal-Learn, most of the reforms were of a similar type, involving two major changes in the law. One is a liberalization of the work rules, so that it is easier for recipients to stay on welfare, work at the same time, and keep much of their welfare benefits. The other is that benefit levels were lowered by an additional 4.5 percent. These changes happened statewide, and include the 10,000 WPDP experimental cases. The 5,000 control cases, which remain on the September 1992 level of benefits that were 8.5% higher, have much more rigid work rules. At the end of five years policy makers would like to have a database that can answer the question: Is there some kind of behavioral change that differs between these two groups?

Cost Neutrality  The Federal government pays half of all AFDC benefits incurred by the states and the states pay the other half. Since the Federal government will not pay more money than they were used to paying under the old rules, the outcomes of all welfare changes must be cost neutral. If more money is paid out than was paid under the old rules, the states must make up the difference. A controlled experiment is required in order to be able to determine if the changes are cost neutral overall. People are eligible for higher food stamp benefits under the new
rules, because food stamp amounts are based on the total of earned and unearned income for a family. Discerning whether the Federal share has increased is a complicated matter, since food stamps are entirely paid by the Federal government, and the cost neutrality computation includes Medi-Cal payments, and payments for homeless assistance and child care as well.

All data from the welfare administration data systems is sent to UC DATA for the three thousand experimental and control cases in Alameda, San Bernardino, and San Joaquin counties. In Los Angeles county an extract is prepared which summarizes the information from their six thousand research cases. The identification and documentation of similar variables from county to county results in what is called a Uniform Database (UDB). When county data is collected for youth in the Cal-Learn program, an additional nine different data systems will provide information that must be placed in a uniform language and format, and read across counties. UC DATA acquires monthly tapes from the research counties and they vary in size from 6 to 150 megabytes.

**WPDP Surveys** Of the fifteen thousand cases, DSS is supporting a series of in-depth telephone interviews with a 15% subsample, or about 2,250 female heads of assistance units who speak English or Spanish. The first wave was fielded from October 1993 to August of 1994, and includes English and Spanish-speaking subjects. Also out of the 15,000, all adult female heads of assistance units from language groups that make up more than one percent of the sample are part of a Foreign Language survey. This survey asks basically the same questions of an additional 1350 people in the Vietnamese, Laotian, Cambodian, and Armenian languages. The survey program, which is being conducted by the Survey Research Center at UC Berkeley, employs a prospective longitudinal research design, and anticipates two additional waves of interviews of the same subjects spaced eighteen months apart.

Domains of inquiry include respondent and household information, AFDC, education and work history, housing quality and stability, economic hardship, hunger, respondent and child's health and disabilities, labor market activities of partner/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, and the greatest problem with conducting the interviews is locating the respondents. Each interviewee is paid ten dollars for their time.

**Data Validation - Case Coding.** The Review and Evaluation Branch (REB) of the Department of Social Services checks the validity of the information in case files around the state on a random basis. Consequently, there are staff in that unit trained to pull out case folders, examine case files and extract information. UC DATA researchers want to know whether the information acquired from several different sources actually corresponds to the information in the case files, which is a process of triangulating the data. For the 15% of the sample that comprises the English/Spanish interviewees, case coding is being done to validate survey and case information from administrative data files, as well as to collect certain kinds of information found nowhere else--such as the presence of restricted bank accounts, the number of hours worked for cases with earned income, and the equity value of automobiles. UC DATA designed a series of data entry programs that run on laptop computers to facilitate this data collection effort.
In addition to the approximately 2,000 English-Spanish cases that have been interviewed, REB staff will be case coding all U cases in the total sample which have shown wage and salary income on the administrative databases sent from the counties between January of 1993 and June of 1994. Approximately 7 to 8% of all cases showed earned income in during 1993; REB expects to case code about 2,500 U cases with earned income for the first eighteen months. All U cases with earned income for the life of the project will be case coded, since for the U cases alone, the number of hours worked is routinely kept in the case files. Since one of the welfare reforms being evaluated involves the elimination of the 100 hour rule for two parent families, it will be possible to compare hours worked by U parents both before and after the change went into effect.

Process Evaluation The only type of evaluation for which UC DATA is completely responsible is the process evaluation. All other data collected will eventually be sent to a separate evaluation contractor, selected through a competitive bidding process. The evaluation contractor will be responsible for analyzing the information provided by UC DATA and producing interim and final impact and cost benefit studies. Under the Federal Terms and Conditions, impact findings and cost benefit reports must be interpreted using information collected through a process evaluation.

In order to understand the nature and developmental history of the WPDP welfare changes, UC DATA is implementing a series of qualitative field studies designed to explore organizational, situational, and personal contingencies which promote changes in program operations and which are more or less successful in implementing the new welfare rules in the field. The WPDP process evaluation methodology encompasses a variety of techniques with well-established usefulness in evaluation research, including:

1. in-depth, open-ended interviews with staff, administrators, and others involved;
2. participant observation, which permits events to be recorded as they occur;
3. process evaluation inventories, which identify in detail items of information to be collected in the field; and,
4. regular on-going communication with agency staff and directors in which questions may be clarified and anecdotal data can be reported to provide a more complete picture of the implementation process.

Research Sample Longitudinal Database For the original 15,000 cases, DSS has provided welfare histories from the Medi-Cal persons file: date of birth, race or ethnicity, sex, aid type, and social security number (which allows matching). Currently, data is available from 1987 through 1992 and includes spells of other types of aid as well as AFDC, such as General Relief (GR) and Social Security Income (SSI). Eventually, UC DATA will attach this longitudinal database to variables from the county administrative databases that have been defined as comparable from
county to county, and produce a public use dataset with the combined information. All identifiers will be removed and responses from the in-depth telephone surveys will be added for the cases that have been interviewed.

**Longitudinal Database (LDB)** UC DATA and DSS are collaborating in an effort to document the dynamics of family poverty and welfare use in California. The LDB research component collects, cleans, organizes and documents detailed longitudinal data for representative samples (1% and 10%) of low-income families who received welfare in California between 1988 and 1992 and makes these data available to scholars and analysts with an interest in welfare, poverty, employment and family research. Future plans for this component will include tracking the original samples from the Medi-Cal file through other databases, such as the Employment Development Department (EDD) and Greater Avenues for Independence (GAIN).

**Cal-Learn** Teams of scientists and experts from several different fields developed a separate research design to evaluate the Cal-Learn program. Randomly assigned groups are being created within the Cal-Learn eligible population, which includes pregnant or parenting AFDC recipients who are not yet nineteen years of age or high school graduates. Previous research has suggested the importance of testing separately the impact of the two major Cal-Learn program elements: case management services and financial incentives. These two factors yield four distinct research conditions: case management services only, financial incentives only, both, and neither. Cal-Learn eligibles from all groups qualify for free child care and money for transportation and other school-related expenses.

In the four WPDP research counties pregnant and parenting teens--not currently in AFLP or GAIN teen parent programs--will be randomly assigned on the basis of the last two digits of their social security numbers into one of the four research conditions and observed over the five year period that the demonstration project is in effect. Each group will be followed to determine whether or not its members get higher grades, graduate from high school, have healthier babies, and put off having additional children.

While this design is useful for evaluating the programmatic strategies within Cal-Learn, it can tell us little about why adolescents become pregnant, have children, and enter into Cal-Learn. Consequently, broader studies are anticipated with three additional elements: (1) a retrospective evaluation of the social circumstances of program participants and their assessment of the program; (2) annual follow-up interviews with all Cal-Learn participants; and, (3) a prospective study to determine the etiology of teen parenthood and other important social outcomes among high risk youth. For the prospective study, a series of interviews are planned with the teenaged sons and daughters of the AFDC interview sample.
Overview of WPDP Project: Data Structure and Components

**DHS MEDS file**
- **Sample Size:** All of California
- **Time Interval:** Ongoing
- **Data Source:** Primary Source

**County Records-WPDP Counties**
- **4 Counties:** LA=Los Angeles / SB=San Bernardino / AL=Alameda / SJ=San Joaquin

**LDB (Longitudinal Database)**
- **Sample Size:** 4,500,000
- **Sample:** 10% sample of MEDS persons & cases
- **Data Period:** 1987-1998
- **Time Interval:** monthly

**Uniform Database**
- **Sample Size:** 15,000 cases
- **Sample:** County database
- **Data Period:** 1993-1999
- **Time Interval:** monthly replacement

**WPDP County Database**
- **Sample Size:** 15,000 cases (10K experimental, 5K control)
- **Sample:** stratified random sample of 4 counties (LA, AL, SB, SJ)
- **Data Period:** 1993-1999
- **Time Interval:** 100% monthly replacement

**Cost Neutrality**
- **Sample Size:** 15,000 cases
- **Sample:** County database
- **Data Period:** 1993-1999
- **Time Interval:** monthly replacement

**WPDP LDB**
- **Sample Size:** 15,000 cases
- **Sample:** Stratified random sample of 4 counties (LA, AL, SB, SJ)
- **Data Period:** 1987-1998
- **Time Interval:** Monthly

**Case Coding/Data Validation**
- **Sample Size:** 5,000 cases
- **Sample:** Survey + all U cases w/ earned income
- **Data Period:** 1994-1999
- **Time Interval:** every 2 years
- **Data Source:** quality control program

**Process Evaluation**
- Documentation of State and 4 county implementation of APDP/WPDP; includes Cal-Learn and other SB 35 changes
- **Time Interval:** yearly
- **Data Source:** In-depth interviews and archival records

**Prospective Teen Study**
- **Sample Size:** 1,200 persons
- **Sample:** All teens 11-18 of survey respondents, male and female
- **Data Period:** 1994-1997
- **Time Interval:** every 2 years
- **Data Source:** quality control program

**Foreign Language Survey**
- **Sample Size:** 1,818 persons
- **Sample:** All foreign language speakers > 1% of APDP county database
- **Data Period:** 1994-1999
- **Time Interval:** every 2 years
- **Data Source:** data collected by SRC survey

**English/Spanish Survey**
- **Sample Size:** 2,250 persons
- **Data Period:** 1993-1999
- **Time Interval:** every year
- **Data Source:** data collected by SRC survey

**LDB database size:** 10% 15,000 in 4 counties

100% All foreign language > 1%

100% 15% + U cases

15%
WELFARE POLICY AND WELFARE RECEIPT IN CALIFORNIA

Michael Wiseman - Sacramento, CA
Notes by Barbara Snow - UC DATA

On April 27, 1995 Dr. Wiseman, under a subcontract with UC DATA and the California Department of Social Services to analyze longitudinal data on welfare spells over a twenty year period, presented preliminary results at the Capitol in Sacramento. He used numerous overheads to describe the characteristics of the population and to compare population statistics from the US census (updated by the CA Department of Finance) with CDSS data from the CA237. Except for short term dips and peaks, welfare spells have increased proportionately with the population of women in California between the ages of 15 and 44, the child-bearing years. One half of one percent of women in this age category, who are at risk of being on AFDC, are on aid in every month. Using accession and termination rates from the CA 237 forms, he analyzed interprogram movement between FG and U aid categories to examine the structure of AFDC utilization.

The numbers of families on AFDC increased between 1965 and 1971. Declines in the numbers of families on aid occurred with the Welfare Reform Act of 1971, and flattened out with a slow rise until 1981. In the year following the OBRA changes in 1981, the AFDC accession rate declined by 7.5% and by 14.7% in the third year following the change, but the termination rate remained the same. A slow rise between 1983 and 1990 was followed by a steep rise in the AFDC population through 1994. The AFDC U aid category follows the same structure as the FG aid category over time, even though there are many fewer unemployed parent cases.

Along with a rising population of women "at risk" over the period is a disproportionate increase in the number of Latina women in the population and an increased probability of AFDC receipt within this group. The percent of white families on aid has declined by 14% since 1973. The percent of black families has declined by 6%; and the percentage of Latino families has increased by 15%. The California monthly birth rate rose until 1990, and fell conditional upon the economic downturn. The birth rate is inversely related to the unemployment rate.

In 1988 California saw an increase in the minimum wage. Nevertheless, current prices are 40% higher than they were in 1983. Dr. Wiseman used an index to compare prospects of women on AFDC called the "Net Earnings Capacity." To get this variable in dollars per month, he used the Current Population Survey expected earnings based on a weighted average of the prospects of women aged 25 who are high school graduates and who have gone through the eleventh grade working 35 or more hours a week, which is considered full time by the CPS. He used mixed race and average location, since earnings vary by ethnic group and by locale. The variable was defined after the tax liability was subtracted and the earned income tax credit added, and was expressed in 1983 dollars. As expected, this variable was negatively related to entry onto the welfare rolls, but not to terminations. Trends in the real net wage is important for it will mean an expansion in accessions.
He indicated that it is necessary to control for trends so as not to confound the variable observations with the effects of the policy period. In two economic time series models he developed for the AFDC FG accession and termination rates, several policy changes were taken into account, as well as the net earnings capacity, the unemployment rate, the birth rate, the maximum aid payment, and the proportion of the population which is black, latino, and other. Random assignment, such as in the CWPDP research, is necessary to control for exogenous influences. Yet these models predict actual changes very well.

MAP and food stamps have undergone rapid deflation of benefits levels since 1990. GAIN drives terminations down and accessions up only apparently because of the overall caseload increase. GAIN effects so few people, that it has no overall effect on either model. The Omnibus Budget Reconciliation Act of 1981 had the effect of reducing accessions, because it added a test of eligibility. This finding was surprising, although the effect was attenuated over time. OBRA had no effect on terminations, and no effect after the initial spike; falling to insignificance over time. Terminations act with a lag. It takes time to find a job after the unemployment rate declines. Less capable people find employment, but they are the last to be hired and the first to be fired. An increase in the minimum wage is associated with AFDC termination rates. Case size is inversely correlated with the termination rate, as is the number of child only cases.

The estimated proportion of Latina and African American women in the population between the ages of 15 and 44 who are not on aid is strongly associated with accession rates. The proportion of women in this age range who are from "other" ethnic groups is negatively related to accession rates. The proportion of women case heads on AFDC who are latino is also negatively associated with termination rates. But not the proportion who are African American, which is unrelated. The birth rate variable three months into the future is a better predictor of accessions than the current birth rate. Women within three months of delivery qualify for AFDC.

An OBRA effect simulation was discussed. An effect of OBRA was observed in the accession rate, because this policy involved new standards (if income exceeded the need standard by 150%, a family did not qualify for AFDC). There were also changes in the way grants were calculated and in the fixed deductions. The $30 and 1/3 income disregard was limited to four months. This limitation had no effect on the rates of termination. But in fact, the share of the caseload mixing work and welfare declined. The real impact of OBRA came with differences in the processing of eligibility.

The termination rates are underpredicted by the Work Pays model simulations. More people are leaving the welfare rolls than expected. The Work Pays simulation is Dr. Wiseman's attempt to model the anticipated effects of the reductions in the maximum aid payment over a one year period for the FG caseload. It shows that the actual caseload is below the caseload size expected by either simulation, with or without benefit reduction. Contrary to expectation, the MAP is neither associated with accessions or with terminations from welfare, so this model does not suggest a large anticipated effect from changes in the MAP.