
Upjohn Institute Press

Evaluating Net Program Impact

Terry R. Johnson
Battelle Memorial Institute

Ernst W. Stromsdorfer
Washington State University



Chapter 2 (pp. 43-131) in:

**Evaluating Social Programs at the State and Local Level: The JTPA
Evaluation Design Project**

Ann Bonar Blalock, ed.

Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1990

DOI: 10.17848/9780585243900.ch2

2

Evaluating Net Program Impact

Terry R. Johnson
Battelle Human Affairs Research Centers
and
Ernst W. Stromsdorfer
Washington State University

There are many natural social settings in which the researcher can introduce something like experimental design . . . which makes a true experiment possible . . . Such situations can be regarded as quasi-experimental designs. [We need to] encourage the utilization of such quasi-experiments and to increase awareness of the kinds of settings in which opportunities to employ them occur.

Donald Campbell and Julian Stanley
*Experimental and Quasi-Experimental
Designs for Research*

General Concepts and Methods

In a world of scarce resources and growing federal deficits, it is increasingly important for social programs to document the return that society as a whole receives on its investment. To effectively compete for limited federal, state, and local resources, social programs that have provided services for many years find themselves having to document clearly the benefits and costs of their services.¹ In addition, innovative demonstration projects must properly document their outcomes and costs to compare with existing social programs or alternatives to meet the same goals, so policymakers can make informed decisions on whether the demonstration should be continued in its current form, expanded or reduced in scope, or discontinued. The emphasis on social program accountability applies to all types of social programs, including those that provide health, housing, or income maintenance services to the aged or to children, and has been an integral part of the federal employment and training system for economically disadvantaged youths and adults for the past few decades.

The information required to make a decision on whether to expand, contract, or discontinue a social program begins with a *net impact evaluation*. Numerous factors affect the outcomes of participants served by any social program. These factors include the participants' own characteristics, the program environment, and the services provided. Recognizing that many factors may affect program outcomes, a net impact evaluation attempts to answer the fundamental question, what difference does the program make? This is in contrast to a *gross impact evaluation*, which focuses on whether the outcomes of interest for participants are greater after the program than before the program, and does not determine whether the program services *per se* caused the observed differences in outcomes.

A net impact evaluation examines the changes in outcomes from before to after receipt of social services and, in particular, examines whether any change can be causally attributed to the services received. By comparing the outcomes for participants of a given social program with the outcomes that would have occurred if the program did not exist, a net impact evaluation provides valuable information on program benefits that can be combined with cost information to make informed judgments about the cost-effectiveness of the program. In this chapter, we illustrate a general approach to assessing the net impact of a social program by describing how to estimate the net impact of employment and training programs for economically disadvantaged individuals on participants' postprogram labor market experiences.

The details of the net impact evaluation strategy described in this chapter have been tailored to a specific social program, but the key elements of the net impact approach have general applicability to other social programs. For example, the first element of any net impact analysis involves the development of an appropriate *conceptual framework*. This framework identifies the key research questions addressed, the outcomes examined based on those questions, and the participant subgroups and program services delivered to clients.

The outcome measures examined constitute the *dependent variables* for the net impact analysis. The conceptual framework specifies the key relationships investigated between the dependent variables and meas-

ures of various causative factors, such as participant characteristics, the economic environment, and the program services. These factors that affect the dependent variables are called *independent variables*. By fully specifying the net impact questions of interest, the conceptual framework identifies the key relationships between the dependent and independent variables to be analyzed.

The second component of a net impact analysis is the development of a *research design* that provides valid answers to the questions of interest. To select an appropriate research design for a particular social program, one should be guided by several criteria. These include *internal validity*, *external validity*, *statistical efficiency*, and *feasibility*. Internal validity refers to the ability of the design to yield unbiased estimates of the causal relationship between program services and outcomes; that is, valid estimates of the net impact of the program. External validity refers to the ability of the research design to achieve results that can be generalized to a broader population than the specific samples of individuals upon which the analyses are based. Statistical efficiency relates to the overall sampling strategy and the need to utilize sufficiently large samples to obtain precise answers to the research questions of interest. Finally, feasibility relates to the ability to implement the research design and obtain meaningful results in a timely fashion and within the limited resources available.

Although all of these criteria are important considerations in selecting a research design, the criterion that receives the most attention in the research literature is internal validity, or obtaining unbiased estimates of program impacts. To meet the internal validity criterion, one must be able to compare the observed outcomes for participants with the outcomes these individuals would have achieved in the absence of the program. To do so, one needs a comparable group of nonparticipants to serve as a comparison group.

The key to internal validity is the comparability of the groups being compared. One can obtain valid causal inferences about the net impact of the program only if other factors that affect the outcomes are statistically equal for the two groups. The most effective way to ensure this in a social program evaluation is to use a classical experimental design in

which individuals eligible for program services are randomly assigned to treatment or control status. By randomly assigning some individuals to receive services and other eligibles to not receive services, the two groups should be essentially identical on all dimensions that might affect program outcomes. Thus, any observed differences in outcomes can be reliably attributed to the treatment rather than to pre-existing differences in the characteristics of the two groups.

Although this method has often been used to test the effectiveness of new service interventions, it is not generally used to evaluate the net impacts of social programs.² This is, in large part, because of ethical, legal, and programmatic concerns. As a result, a matched comparison group approach—referred to as a *quasi-experimental design*—is typically used. Such an approach relies on some matching rule based on measured characteristics. A matched comparison design could involve the matching of service providers, in which one provider offered the treatment and a comparable provider did not; the matching of individuals, in which participants within a given program are matched to comparable nonparticipants; or the design could include elements of both. Because the feasibility criterion in many cases eliminates the use of an experimental design, in this chapter we describe a quasi-experimental net impact evaluation approach. This approach should yield a considerable amount of useful information on the effectiveness of a social program.

A third element of a net impact analysis is the *data collection and analysis plan*. A guiding principle for collecting data is to make sure that the information used is measured comparably for both participants and comparison group members. Thus, if certain data items are obtained from different social service agencies for the two groups, one must review the data collection forms and procedures to determine comparability. In addition, it is particularly important in a quasi-experimental evaluation to obtain extensive historical information on all key outcomes of interest to determine whether the participants and comparison groups are similar on these measures before participants receive program services. If there are major differences in outcomes between the two groups before program participation, this makes it more difficult to isolate the true effects of program services.

Once the data are collected, the next step is analysis. Because the general statistical issues inherent in most net impact analyses are quite similar, the plan for analyzing the data collected for evaluations of different social programs includes many common elements. For example, a key element of any analysis plan is a detailed examination of the likely extent of *selection bias*. Selection bias refers to the potential non-comparability of the participant and comparison groups due to the fact that individuals self-select themselves to become participants. In addition, the analysis plan must include a strategy for estimating the net impacts of the program overall, and for major program services and participant subgroups to answer the questions identified in the conceptual framework.

Finally, although the results of a valid net impact analysis can provide useful information on the extent to which participants are better off as a result of receiving program services, they are not sufficient to directly address questions related to whether the program should continue to operate at the same level, be expanded, contracted, or discontinued. To determine whether a social program is an efficient use of public resources, one must sum the benefits to participants and the benefits that may accrue to other segments of society, and compare the total benefits with the total costs of the program. That is, one must conduct a *benefit-cost analysis*. In practice, this involves measuring the benefits to participants, taxpayers, and government that accrue over time, properly discounting the benefit stream into current dollar values, and comparing the total benefits with the full social costs of the program.

The remainder of this chapter illustrates the key elements outlined above of a quasi-experimental net impact evaluation of a social program with an application to employment and training programs for JTPA-eligible disadvantaged adults. Although the details of the net impact evaluation strategy described in this chapter are specific to employment and training programs, we emphasize the broader applicability of this general approach to other social programs.

Application of the General Concepts and Methods: JTPA

The net impact evaluation strategy described here was designed in response to an increased need for reliable evaluation information at the state and local levels concerning the effects of employment, education, and training programs funded under the Job Training Partnership Act (JTPA) of 1982. The JTPA significantly changed the employment and training system in a way that gave states and local service delivery areas (SDAs) much greater responsibility for program accountability compared to the Comprehensive Employment and Training Act (CETA). This increased the need for reliable information on program effectiveness at the state and local level at the same time that the federal government greatly reduced its role in providing such information.

As a result, states have less access to reliable federal-level data and federally sponsored research to assist them in making informed judgments about program oversight and social policy. To fill the policy research needs of the states, this chapter describes step-by-step how a state can estimate the net impact of its JTPA programs on earnings and welfare dependency and provide valid information on the investment return from the JTPA.

In developing a state-level net impact model of JTPA programs, we were guided by several considerations. First, for the model to assist states in meeting their new accountability responsibilities, it must produce scientifically *valid* estimates of the net impacts of JTPA program activities and services on relevant participant postprogram outcomes. Second, the model must provide *meaningful* and *timely* information that can be understood and used efficiently by a relatively nontechnical audience. Finally, the model must recognize the *resource and other practical constraints* that states and SDAs face. The two most important practical realities affecting the recommended approach are that (1) states and local SDAs will not generally be able to implement an experimental design in which eligible applicants are randomly assigned to treatment vs. control status, and (2) states and local SDAs will not generally be able to conduct follow-up interviews with a large sample of participants and comparison group members.

Because the proposed net impact evaluation model relies exclusively on administrative records from several agencies that collect these data as part of the normal operating process, the model is widely usable. This model supports timely analysis conducted within the economic and political resource constraints faced by states and SDAs.³ At the same time, it must be recognized that this approach limits the questions that can be addressed and the variables that can be used to adjust for various statistical problems that could threaten the validity of the analysis. Thus, in evaluating other social programs that face different economic and political constraints, one should consider supplementing the research design described here to include primary data collection activities.

The remainder of this chapter describes a detailed approach for estimating the net impact of employment and training programs for disadvantaged individuals. The chronology of the discussion is as follows:

1. We summarize the lessons learned from the research literature on employment and training net impact evaluations.
2. We describe the key elements of the conceptual framework for a state-level JTPA net impact evaluation by building on previous studies. This conceptual framework includes a description of program outcome measures, the trainee subgroups for which impacts should be separately measured, the program activities (services) to be examined, the types of economic and demographic factors that affect these outcomes, and the data that measure these elements.
3. We describe a research design for analyzing the net impact of the JTPA. The research design describes how to select the most reliable comparison group of otherwise similar nonparticipants to approximate what the labor market experiences of participants would have been in the postprogram period had they not participated in the program. This approach to net impact analysis attributes to program participation the *incremental gain* in labor market experiences that occurs over and above what would have happened had these individuals not participated in the program. This is the appropriate concept for providing information on the return on investment of job training programs. The research design also specifies how to select a representative sample of JTPA participants as well as sufficient

numbers of participants and comparison group members to provide statistically valid net impact results.

4. We discuss the steps involved in obtaining and processing the required data. In particular, we describe the elements of a data collection and processing cycle, and indicate potential problems that may arise. We then describe how to organize the various data sources, the types of data cleaning to be performed, and key features of the analysis files to be constructed.
5. We outline a data analysis plan for estimating the net impacts of JTPA programs on participants' postprogram outcomes. Beginning with a description of methods to determine the quality of any comparison group selected, we describe an approach to estimate the overall net impacts of JTPA and major target groups with respect to the key participant postprogram outcomes indicated in the federal legislation. We also discuss potential threats to the validity of the analysis and indicate possible approaches for adjusting for such problems.
6. We describe how to conduct a cost analysis to determine the costs of JTPA services. We also indicate how to combine this information with the net impact results to provide evidence on the return on investment of job training projects.
7. We indicate how the net impact analysis complements analysis strategies described in other chapters.
8. We conclude with a summary of the general concepts and methods applicable to a net impact evaluation of any social program.

Learning from Past Evaluations: Developing the Context for JTPA Net Impact Analysis

In developing a strategy to evaluate the net impact of any social program, it is very important to understand the various approaches that have been used previously and the nature of the results that have been obtained. In an earlier net impact evaluation guide (Johnson 1986), a detailed review of approximately 20 employment and training net impact evaluations was provided. This section summarizes the results of that literature review to set the context for the conceptual framework described in the following section.

With the exception of several recent net impact evaluations of CETA, most of these studies examine the impact of CETA's predecessors, for example, the Manpower Development and Training Act (MDTA) classroom and on-the-job training programs. Some evaluations concern CETA contemporaries, such as the Work Incentive (WIN) program, the Job Corps, or various employment and training demonstrations. Though none of these studies deals explicitly with JTPA programs, and many focus on pre-CETA programs, they are of interest because the programs examined have many characteristics in common with JTPA programs, and the evaluation elements used are the same as those to be developed for a JTPA net impact model.

The previous net impact evaluations of employment and training programs examine the impacts of services on participants' earnings almost exclusively, which is consistent with the policy objectives of federal employment and training legislation.⁴ Moreover, Social Security Administration (SSA) records have been the main source of earnings data. Although SSA earnings records have several advantages (e.g., they are a cost-effective source of longitudinal data measured comparably for participants and comparison group members), they have several potential disadvantages, including coverage problems, exclusion of earnings beyond the taxable maximum, and delays of up to three or four years in obtaining reliable data. These delays prohibit a state-level evaluation from obtaining policy-relevant results in a timely fashion. In addition, when SSA earnings are the only outcome measures available, evaluation is limited to estimating impacts on an annual basis.⁵

The results of these studies generally indicate large net earnings gains for women, particularly nonblack women; whereas, the net effect of employment and training programs on the earnings of adult men is less clear. Although almost all studies find the net earnings gains of men to be considerably less than those obtained by women, several recent evaluations find that male trainees never regain the earnings position they held prior to training relative to otherwise comparable nonparticipants. If this is true, why did men continue to enroll in these employment and training programs? Perhaps because of program earnings and the substantial training subsidies offered by CETA.

These results indicate that separate net impact models should be developed for men and women because the relationship between earnings and various socioeconomic and demographic characteristics clearly is different for men and women.

Although the earlier evaluations of MDTA programs focused more on separate net impact models for whites and blacks, more recent evaluations of CETA programs focus on estimating separate models by age groups. Because of the difficulties in developing valid net impact estimates for youth, many recent studies have estimated models only for adult men and women. The problems for youth are twofold: earnings are not the single appropriate outcome measure for youth, as the relative mix of schooling, market work, nonmarket work, and leisure evolves rapidly over time for them; and it is very difficult to draw a reliable comparison group for youth with limited and highly variable earnings histories. As such, it may not be feasible to develop a state-level JTPA net impact model for youth that provides valid results.

Another evaluation element concerns the participant groups chosen for analysis and the variables included in the model to measure the service intervention or treatment effects. For the most part, these studies focus on estimating the average impact of program services on earnings for the selected subgroups. Because in many cases subgroups of interest were participants enrolled in specific program activities (services), this resulted in numerous net impact estimates by program activity or treatment. The only other dimension of the treatment examined in some of these studies was length of program participation. The results indicate that net impacts vary by program activity and length of participation. Although fewer programs are generally offered under the JTPA (as compared with CETA), and the average length of stay in JTPA is less than in CETA, it is still important to develop models that examine these potential differences in outcomes.

Finally, because virtually all of these studies rely on large-scale data bases on program participants developed well before the analysis was undertaken, little information was generally available on the content of the services provided, whether the program services were provided as planned, and the extent to which the services provided varied across sites

and over time. To overcome this problem, it is important to conduct a process analysis like the one described in chapter 4.

Developing a Conceptual Framework for Evaluating Net Impact

The conceptual framework is the first major element in the design of a net impact evaluation. This framework answers the question “What is to be learned from the evaluation?” and identifies key research questions to be addressed. These are:

1. Outcomes to be examined and the relationship of these outcomes to
2. program activities or services.
Participant groups and program activities to be included.
3. Specific definitions of the variables that measure outcomes, program
4. activities, and any other variables affecting the relationship among activities and outcomes.

As such, the conceptual framework guides the development of the research design and analysis method. This section describes a conceptual framework for conducting a state-level JTPA net impact analysis that accounts for the lessons learned from previous studies and is based entirely on available state administrative data sources.⁶

General Evaluation Questions

Although employment and training programs funded under the JTPA can affect different groups, such as participants, employers, the government, or society as a whole, in various ways, the primary goal of the state-level JTPA net impact model is to determine the extent to which JTPA program activities or services improve the labor market experiences of participants relative to what their experiences would have been in the absence of the program. The net impact of JTPA program activities on participants’ postprogram labor market experiences provides policymakers with an indication of the overall gains due to these programs.

Although it is important to know whether the mix of JTPA activities on average is effective, for policy purposes it is equally important to identify the relative effects on different target groups. For example, to improve targeting it is important to know whether certain participant

groups benefit more from a particular JTPA activity or service than other groups, and whether the net impact of JTPA differs among program activities, by length of participation, and by local economic conditions.

The general objectives of the state-level net impact model can be summarized by the following key research questions:

1. What is the overall net impact of JTPA program activities (services) on participants' postprogram labor market experiences?
2. Do the net impacts change over time? If so, in what way?
3. Which program activities result in the largest net benefits to participants and society relative to their costs?
4. Which groups gain most from participating in JTPA?
5. For a given program activity, do individuals who remain in JTPA longer experience greater net gains in labor market outcomes? Does this effect vary among activities?
6. How does the net impact of the JTPA vary by local program and environmental conditions?

Table 2.1 details the key elements of the net impact conceptual framework that help to make these questions more specific. Each of these elements is discussed below.

Measuring the Factors Involved in Answering the Evaluation Questions

Participant Outcomes

As table 2.1 shows, the state-level JTPA net impact model focuses on specific postprogram participant outcome measures related to earnings and welfare dependency that are available in automated administrative records in most states. Because unemployment insurance (UI) Wage Records are maintained on a calendar quarter basis, the earnings measures for longer periods can be created as the sum of quarterly earnings amounts. The Aid for Families with Dependent Children (AFDC) grants measures can be calculated as the sum of monthly grant payments. The employment status and the AFDC participation status measures are defined as "dummy" (or indicator) variables. For example, a person is defined to be employed for any period of interest if UI Wage Records indicate positive earnings were received during the period (1 if employed, 0 otherwise).

Table 2.1
Key Elements of the Conceptual Framework
for the JTPA Net Impact Model

Outcome Measures

- Quarterly, semiannual and annual earnings and employment status based on Unemployment Insurance (UI) Wage Records.
- Quarterly, semiannual and annual AFDC grants and AFDC participation status based on public assistance (PA) records.

Participant Subgroups

- Adult men and women. to the extent possible, disaggregate net impacts by sex, age, race or ethnicity, education, and for women, by welfare status.

Program Activities or Services

- Classroom training: when program data are available, estimate separately for remedial education and basic skills versus specific occupational-skills training.
- On-the-job training
- Job search assistance.
- All other activities or services.
- Combinations of the above activities or services.

Labor Market Conditions

- Unemployment rate.
- Urban or rural location.

Data Sources

- JTPA Management Information System (MIS): participant characteristics, program activities, and placement experiences.
 - PA Grants Records: whether received AFDC and size of AFDC grant.
 - UI Wage Records: whether employed and amount of quarterly earnings
 - UI Benefit History: whether received UI benefits in preprogram period.
 - Local labor market information: the local (or regional) unemployment rate.
-

These outcome measures are consistent with the major objectives of the JTPA legislation as specified in the law. As the literature review indicates, these measures typically exceed the nature and extent of

outcomes examined in previous studies of the net impacts of subsidized employment and training programs. They also capture the range of short-term and relatively long-term impacts that could be observed within, approximately, a two-year program analysis cycle.

The administrative data sources for constructing key participant outcome measures have several advantages for evaluating the net impact of social programs for which earnings and welfare dependency are of interest. First, UI Wage Records are not truncated at some taxable maximum; that is, actual earnings are reported, which reduces measurement error. Second, UI Wage Records are available by calendar quarter, with only a three- to six-month delay in obtaining reasonably complete data. The availability of data by calendar quarter allows considerable flexibility in the choice of a postprogram follow-up period. Third, UI Wage Records are not subject to interviewer bias or respondent error. Moreover, these data are not subject to problems arising from some respondents reporting net (after-tax) earnings and others reporting gross (before-tax) earnings. Finally, they are not affected by response-rate problems.

Monthly AFDC grants from administrative records have several advantages relative to data obtained through surveys. These include timeliness of data availability and the absence of respondent reporting errors, interviewer biases, and response-rate problems. Because these administrative data systems are relatively inexpensive to obtain, our net impact analysis relies on them.⁷

It is also important to recognize the limitations of these administrative data sources. First, although the Deficit Reduction Act of 1984 required all states to become wage-reporting by 1988, there are still a few states that do not report at this time, and will not for the most part be able to use the net impact model to examine earnings impacts.⁸ Second, states that do not have automated AFDC Grants Records available at the state level cannot easily use the net impact model to examine JTPA impacts on reducing welfare dependency. Third, because our net impact approach relies heavily on the availability of historical earnings and grants data, states must be able to directly access or retrieve from archives two to three years of UI Wage Records and public assistance (PA) Grants

Records for a given person at any one time. Although most UI Wage Records systems are similar across wage-reporting states and generally contain at least seven quarters of data at any time, there is considerable variation in state and local welfare administration and record-keeping practices, as well as differences in the degree of data automation and retrieval capabilities. These variations present obstacles to implementing this component of the net impact analysis in certain states.

Fourth, UI Wage Records do not generally include employees of federal, state, or local governments, self-employed individuals, or employees in certain other occupations.⁹ Finally, because the system is state-based, it is impossible to distinguish individuals who work across the border in a different state from individuals who do not work in covered employment. Thus, one should be very careful in estimating the net impact of the JTPA on earnings for large service delivery areas located near state borders.

Participant Subgroups

The next conceptual issue concerns the participant groups of interest. As table 2.1 shows, the state-level net impact analysis is limited to adults because earnings, employment, and AFDC dependency are not the appropriate outcome measures for youths or those entering the labor market for the first time in their life cycle. Moreover, the more appropriate of these measures, such as schooling attainment and employment competencies, are not contained in any existing program data sets for both program participants and comparison group members. In addition, as discussed above, there is extensive evidence indicating the difficulty in developing a reliable matched comparison group for youths (Dickinson, Johnson, and West 1987(b); Lalonde and Maynard 1987). Because of these problems, the net impact analysis is designed only for adults. States interested in examining youth issues should consider other approaches for conducting a net impact analysis.¹⁰

A second issue concerns whether separate net impact analyses need be developed for any specific adult groups. Sample sizes and state resources permitting, separate net impact models should be developed for adult men and women because the relationship between earnings and other

demographic characteristics is very different for these two groups. In addition, it is desirable to estimate separate net impact models by race and other participant groups, sample size permitting. If no statistical differences are found among any set of population groups, the groups can then be combined for subsequent analysis.

Finally, it is important to investigate whether the impact of JTPA varies by the following participant characteristics:

1. Age—less than or equal to a particular age (e.g., age 35) as compared to over age 35.
2. Ethnicity—whites as compared to blacks and Hispanics.
3. Educational level—at least high school graduate as compared to nongraduates.
4. Welfare status for adult women—welfare recipients as compared to nonrecipients.

The techniques for conducting such analysis are described later in this chapter.

Program Activities: The Service Interventions

The next important element of the conceptual framework is the determination of the key treatments—program activities or services—to be assessed, and the development of consistent definitions of these variables. *Section 204* of the Job Training Partnership Act sets forth a large array of potential activities. However, the major activities provided under JTPA are classroom training (CT), on-the-job training (OJT), and job search assistance (JSA). Nearly 90 percent of adult FY 1984 Title II-A enrollees participated in at least one of these program activities. Although it is important to include in the analysis those participants assigned to all types of activities in order to assess the full range of JTPA activities, as table 2.1 shows, it is most useful to examine the separate effects of CT, OJT, and JSA.¹¹ A brief description of each of these program activities is provided below.

Classroom training involves basic or remedial education or occupational-skills training to ensure that individuals acquire the ability and knowledge necessary to perform a specific job for which there is a demand. Such programs are usually provided in a classroom or an institutional setting off the job.

On-the-job training emphasizes the development of occupational skills in an actual work setting, ideally in the private sector. The programs are designed for participants who have been first hired by the employer, and the training occurs while the participant is engaged in productive work that provides knowledge or skills essential to the adequate performance of the job.

Job search assistance includes any activity that focuses on the development or enhancement of employment-seeking skills. This service is provided to participants who need practical experience in identifying and initiating contact and interviewing with prospective employers. It is usually conducted in a structured setting and can include approaches such as job-finding clubs or instruction for self-directed job search.

Several complications arise in developing measures of independent variables to represent these program activities. First, there is considerable variation among SDAs in the characteristics of specific program activities, such as length of assignment, occupational category of the training, hours of training per day, and quality of instructional inputs. This makes it difficult to specify variables that represent a uniform treatment. When such a variable is not uniform, it is difficult to interpret its statistical meaning, and biased estimates of net program impact can result. Nevertheless, one must aggregate generally similar activities because it is simply not possible to reliably estimate the net effects of the virtually unlimited variations of a given program activity.

Second, large differences among SDAs in the nature of the program activities provided are likely to occur. For example, even within a state, work experience activities in a particular SDA may resemble OJT programs in another SDA. Finally, the way in which the training activities provided are recorded in the MIS can cause complications. For example, due to the lack of uniform national reporting requirements, some SDAs record participation in a job search workshop as job search assistance, while others record it as classroom training because the sessions are conducted in a classroom setting. Such differences in the content and recording of program activities across SDAs emphasize the importance of conducting a process analysis concurrently with the net impact analysis in order to develop meaningful and consistent measures of program activities.

As indicated above, the ways in which the treatment variables are defined are in large part determined by the structure and content of the SDA Management Information System (MIS) and how standardized definitions are across SDAs. In addition, any variable definition depends on the specific research questions of interest and the population size of individuals who participate in a given program activity. For example, it is desirable to separate classroom training activities that focus on remedial education and basic skills from classroom training activities that provide specific occupational-skills training. However, if the number of individuals participating in each of these programs is too small to produce statistically reliable net impact estimates for separate activities, it may be necessary to collapse these two treatment variables into one that represents classroom training program activities in general.

Thus, although the specific definitions of the treatment or program activity variables depend on several factors, the following variables should be specified to examine overall impacts and impacts by program activity and other characteristics of the services provided:

1. Participant dummy variable: 1 if JTPA participant, 0 otherwise.
2. Classroom training dummy variable: 1 if CT participant, 0 otherwise:
 - remedial education and basic skills dummy variable: 1 if CT program in remedial education or basic skills, 0 otherwise;
 - specific occupational-skills training dummy variable: 1 if CT program in a specific occupational skill, 0 otherwise.
3. On-the-job training dummy variable: 1 if OJT participant, 0 otherwise.
4. Job search assistance dummy variable that includes all employment and placement-related activities: 1 if JSA participant, 0 otherwise.
5. Other activity dummy variable: 1 if not a CT, OJT, or JSA participant, 0 otherwise.
6. Multiple-treatment dummy variable: 1 if a combination of two or more of CT, OJT, and JSA, 0 otherwise.
7. Occupation of training dummy variables: 1 if in specific 1-digit Dictionary of Occupational Titles (DOT) code, 0 otherwise, for each type of occupational skill.
8. Length of program participation in weeks.
9. Total training hours, the product of length of program participation in weeks, and the number of training hours per week.

The Program's Environment

The final conceptual element concerns the program environmental conditions that should be included in the net impact analysis. Here, we refer primarily to characteristics of the labor market(s) within which the program operates, although major SDA characteristics can also be considered.

Little is known about how the net impact of employment and training programs varies by the environmental conditions surrounding a program at a point in time or over time. In addition, because of the nature of local program environmental conditions (i.e., there may be no within-SDA variation on these conditions), it is important to recognize that, at best, it will only be possible to obtain reasonably precise estimates of a few key conditions. However, this can be done only in states that have a large number of SDAs and exhibit considerable variation in the conditions among SDAs.

In conducting a net impact evaluation of JTPA programs, it is most important to control for differences in local *unemployment rates* and *location*, i.e., whether the program participants and their comparison group members are located in an *urban or rural* area. These factors are likely to affect the key outcome measures: the employment and earnings of adult men and women.

The unemployment rate can be obtained from the *Local Area Unemployment Statistics*, published by the United States Bureau of Labor Statistics (BLS). This information is available monthly at the state and county level and for over 1,000 cities with a population of at least 25,000. Aggregate measures of the unemployment rate corresponding to the quarterly, semiannual, and annual outcome periods of interest can be calculated as an average of the seasonally adjusted rates.

In specifying the unemployment rate variable for an SDA, it is also important to recognize that monthly values will not generally be available for the precise area of interest. Depending on the geographic jurisdiction of the SDA, the area it serves may be larger or smaller than the county or the city for which any given information is available. Where the SDA serves multiple counties, one should calculate the appropriate labor market variables by aggregating over the counties

served by the SDA. For example, to calculate the unemployment rate one would simply sum the number of individuals unemployed in the various counties served by the SDA and divide by the total number of individuals in the labor force in those counties. In cases where the SDA serves only part of a given county, and where no value is available for a smaller geographical area such as a city, one is constrained to use the county value.

It may also be possible to provide some information on the manner in which the net impact of JTPA varies by different managerial, organizational, or SDA strategies. The service delivery strategies examined should be based on their policy importance to the particular state doing the analysis. Moreover, to ensure that the strategies of interest are distinct and quantifiable, and that there is sufficient variation among SDAs to support the analysis, it is important to conduct a process analysis. Thus, if states with a large number of SDAs (roughly 30 or more) are interested in obtaining information on how the net impact of JTPA varies by a key service delivery strategy, they should first ensure that significant differences in this strategy exist among SDAs.

If it is possible to quantify these differences (usually by means of dummy variables), one could then use the measures of these variables to determine how the net impact of JTPA varies among SDAs. In states with relatively few SDAs, it is unlikely that such an analysis would provide sufficiently precise estimates of the differential effects of the strategy of interest to warrant the analysis.

Developing and Implementing a Research Design for Evaluating Net Impact

To provide valid estimates of the net impacts of JTPA programs on the earnings and AFDC dependency of adult men and women, the research design must contain several elements. Table 2.2 provides an overview of the key aspects of the research design that are discussed in detail in the next several sections. Although the specific details of this research design are sensitive to the features of JTPA, its basic elements and the issues to be considered in making decisions are applicable to any net impact evaluation of a social program.

Table 2.2
Research Design for the JTPA Net Impact Model

Sample Design

- Samples of JTPA participants (adult men and women) must be chosen so that the results can be generalized to the state level.
- Valid comparison groups must be chosen so that the impact of JTPA can be distinguished from the impacts of other factors that also affect earnings and welfare dependency.
- The size of the participant and comparison samples must be determined so that program impacts can be measured with precision

Data Collection

- Comparably measured preprogram and postprogram data for participants and comparison group members must be obtained from several different sources, processed, and analysis files developed

Analysis

- The comparison groups must be examined in detail to determine their comparability to the participant groups and to identify any adjustments that may need to be made to correct for selection bias
 - A comprehensive strategy must be developed to provide valid estimates of the net impacts of JTPA activities (services) on the postprogram outcomes of participants.
-

The elements of the sample design are discussed below. Data collection and data processing issues are the subjects of the following section. A description of the overall estimation strategy and the specific net impact models to be estimated is provided in the subsequent section.

Sample Design

The sample design is a key element of the overall research design. The sample design involves the selection of the participant samples, a strategy for developing the comparison groups, and the determination of sample size requirements for the analysis. As such, the sample design directly affects the internal and external validity of the analysis, as well as its statistical efficiency. Table 2.3 summarizes the three major elements in the sample design for a JTPA net impact analysis model. In the text we discuss each of these elements in more detail.

Participant Groups

The major issues in selecting the participant group concern (1) the

individuals to be included in the sample frame, (2) potential sample exclusions, and (3) the procedure to select participants from the sample frame for inclusion in the analysis. We discuss these issues below.

Sample frame. The choice of the sample frame is an important determinant of the degree to which the findings can be generalized. The sample frame should represent all JTPA participants so the analysis results can be generalized to the state level. The net impact model includes in the participant sample frame all adult male and female JTPA enrollees during a given time interval, as indicated in table 2.3. In

Table 2.3
Key Elements of the Sample Design for
the JTPA Net Impact Model

Participant Group

- Comprised of samples of adult men and women who enroll in JTPA in each quarter of a given program year
- Individuals will be excluded from the sample frame if they are not from 22 to 64 years of age. Individuals will subsequently be excluded from the analysis samples if they have missing data on key JTPA services received (e.g., program activity, length of participation)
- Quarterly samples of JTPA participants will be selected randomly from the groups of adult men and women enrollees that are included in the sample frame to ensure that the sample is representative of JTPA participants in the state

Comparison Group

- Comprised of samples of adult men and women who are new ES registrants in offices in the areas served by the SDAs in each quarter of a given program year
- Individuals will be excluded from the comparison sample frame if they are not from 22 to 64 years of age, if they are not economically disadvantaged, or if they participate in JTPA.
- Quarterly samples of comparison group members will be selected from the sample frame of new ES registrants using a stratified random process to ensure that ES registrants and JTPA participants are similar on certain key characteristics (e.g., welfare reciprocity, UI reciprocity)

Sample Size

- Because the additional cost of increasing sample size is very small, states are encouraged to include in the analysis as many participants and comparison group members as their staffs and computer resources can handle.
 - As a guideline, a total analysis sample of 12,000 cases—divided equally between adult men and women, and between participants and comparison group members (i.e., 3,000 each)—should be adequate to meet most state's analysis needs
-

particular, adult men and women who enroll in JTPA in each calendar quarter of a given program year are to be sampled separately.

This frame has several advantages. First, it yields a representative sample of JTPA participants in which neither short-term nor long-term participants are oversampled, one which is not sensitive to seasonal differences in the characteristics of participants or program activities. Second, because the time period for selecting each participant cohort within the program year is not too long (e.g., three months), it should be possible to select quarterly samples of comparison group members that closely match participants on the timing of the preprogram decline in earnings. This is particularly important for ensuring valid net impact results. Third, using an enrollee-based sample maximizes the amount of preprogram earnings and AFDC data available for the model. Fourth, this approach accounts for the fact that males and females have qualitatively different labor market experiences.¹²

A participant group comprised of adult men and women who enroll in JTPA in each of the four quarters of a given program year has implications for the timing of project results and the length of the postprogram observation period within an approximate two-year program analysis cycle. With such a sample, one can obtain net impact estimates for the period one year following the calendar quarter after termination only for the first quarter cohort, and only a three-month net impact estimate can be obtained for all four quarterly cohorts in approximately a two- to two-and-one-half-year analysis cycle. Of course, by obtaining additional postprogram outcome records for sample members, one could estimate longer-term impacts by extending the analysis period.

Sample exclusions. Once the sample frame is chosen, one must then determine whether certain types of individuals should be excluded. Although such exclusions reduce the representativeness of the participant sample, exceptions may be indicated for several reasons. It may be desirable to exclude cases that lack data on critical items, or the availability of extremely small samples of atypical treatment or participant groups may engender exclusion.

Most studies incorporate restrictions on participant age. Although there is no universal agreement on the age range to use, participants under

age 16 have been excluded because earnings are not an appropriate outcome measure for individuals who are likely to return to school. Participants age 65 and older have been excluded because participation in employment and training programs among individuals eligible for retirement is rare, and a valid comparison group is hard to identify.

Because the net impact model focuses on adults only, we restrict the participant sample to individuals of at least age 22. Because it is difficult to obtain a valid comparison group for older participants, all individuals age 65 and older should be excluded from the participant sample frame. In addition, if the JTPA programs in the state serve very few individuals over age 55, it is desirable to exclude individuals over this age.

Individuals should also be excluded from the sample frame if they have missing data on key variables.¹³ A more difficult problem arises when information is missing on the treatment provided by JTPA. For example, one cannot estimate the net impact by program activity or by length of stay for individuals with missing information on program activity or for those who have incomplete data on the start and end dates of their JTPA participation. A few problems concerning the omission of program start and end dates may arise, in part, because length of stay information is necessary for adjusting certain performance standards for JTPA Title II-A programs. However, since there are no reporting requirements concerning program activity, some cases will contain missing or unusable program activity information.

Moreover, because the quarterly enrollee samples will be selected on an ongoing basis, one cannot know for sure how many cases must be excluded for missing data problems until after the JTPA MIS data are obtained for the sample selected. To compensate for the resulting sample reduction, an expanded participant sample should be selected initially. Individuals who are subsequently determined to have missing data on key JTPA services must be excluded from the analysis sample, provided the reason the items are missing is not systematically related to the impact of the program.¹⁴

A final issue concerns whether to exclude individuals who participate in JTPA for only a minimal period, such as less than a week. Some studies have included all employment and training participants in the

analysis sample while others have imposed arbitrary restrictions that exclude individuals who participate in the program for some minimal period. Note, however, that more-motivated individuals leaving JTPA early because they have found jobs, excluding them from the analysis would result in a negative bias in assessing JTPA impacts. This is because those participants who would do relatively well on their own would be excluded from the participant sample, while the same types of individuals would still be included in the comparison group sample. It is also possible that short-term participants might consist of individuals who would do less well on their own than other JTPA participants. Because similar individuals would remain in the comparison group, excluding the less-advantaged individuals from the participant group would result in a relatively more-advantaged participant group and a positive bias in the estimated program impact.

Either scenario yields a selection bias that threatens the internal validity of the analysis. As such, we recommend that the sample of JTPA participants be kept as representative as possible and that cases not be excluded based on length of stay in the program. It is then possible to examine whether, and in what ways, short-term participants differ from long-term participants to better understand the nature of any selection bias. This, in turn, will help to determine the degree of confidence to attach to net impact estimates by length of program participation.

Selecting the participant sample. Once the exclusions from the participant sample frame have been determined, the next step involves selecting JTPA enrollees for inclusion in the analysis sample. In some states all enrollees in a given program year will be necessary to provide reasonably precise estimates of the average effect of JTPA programs. The sampling issue primarily arises in states that serve large numbers of JTPA participants. We outline a method below for selecting a participant sample from the sample frame described above.

As indicated in table 2.3, the quarterly samples of JTPA participants should be randomly selected from groups of adult men and women enrollees in the sample frame. However, the proportion of males and females varies considerably across SDAs. Therefore, one must first stratify the participant sample by gender before the analysis samples are

selected, otherwise there may be insufficient numbers of either men or women for analysis purposes.¹⁵ Choosing random samples in this manner also has the major advantage of providing separate representative samples of adult men and women participants, so that the results can be generalized by gender. Moreover, estimates of the net impacts of JTPA by gender and of the differential impacts by program activity separated by gender can be obtained without weighting the sample. Also, by selecting participants randomly, an analysis of program activities assigned to different types of individuals is possible.

In addition to stratifying the sample by gender, states that want to focus on specific groups, such as female welfare recipients or male high school dropouts, may also consider stratifying the participant sample and oversampling the groups of interest. In general, stratification is desirable only when the research questions of interest relate to groups that occur rarely, or that occur so frequently that their nonoccurrence is rare. Depending on the specific research questions, one could stratify on the basis of participant characteristics or by program activities (services).

For example, because of the wide variation across states and SDAs in the use of work experience programs, states interested in examining the net impact of these programs would probably need to stratify and oversample participants. Moreover, because job search assistance generally constitutes a less intensive treatment and is therefore likely to have a smaller net impact, a much larger sample of participants in JSA is needed to precisely measure the smaller expected effect. Thus, states interested in precisely measuring benefits gained from JSA participation, must sample program participants in greater numbers. States interested in stratifying the participant sample and oversampling certain groups should consult a sampling expert to understand the steps in drawing the information and the implications for conducting the analysis.

Comparison Groups

To estimate the net impact of JTPA on participants' postprogram outcomes, a method is needed to gauge the results participants would have experienced had they not participated in JTPA. The ideal research design for measuring the net impact of any social program involves the

use of a classical experimental design to develop a true control group. In such a design, JTPA eligibles would be randomly assigned to either a treatment group that could receive JTPA services, or a control group that could not. With this method, the only systematic difference between the two groups is receipt of program services; any differences in program outcomes are due to JTPA. However, ethical and legal concerns can preclude the use of a randomly assigned control group. Thus, we develop an alternative method that approximates a true control group to the maximum extent possible.

A standard approach for determining the net impact of a program is to compare experiences of persons treated by the program (i.e., JTPA participants) with experiences of otherwise similar persons who are not treated by the program (the comparison group). The comparison group is used to estimate what the experiences of the participants would have been in the postprogram period had they not participated in the program. To ensure that differences between the experiences of the two groups can be attributed to the program, the comparison group must have characteristics similar to participants, particularly in terms of program eligibility. Moreover, available data must be comparably measured for the two groups.¹⁶ One must also verify that individuals in the comparison sample in fact did not receive JTPA services.

As shown in table 2.3, the comparison group should be comprised of new Employment Service (ES) registrants in offices in the areas served by the SDAs. ES registrants have several advantages as a comparison group. First, data are available on several characteristics of interest—including those related to JTPA eligibility—that generally are comparably measured with JTPA MIS data.¹⁷ Second, like JTPA participants, new ES registrants have probably experienced a recent decline in earnings.

Finally, also like JTPA participants, ES registrants are in the labor force at the time they apply for assistance. That is, they are either working or actively seeking work. It is important to ensure that participant and comparison group members are similar in their attachment to the labor force. Otherwise, net impact estimates can be erroneous. To ensure comparability on preprogram labor force involve-

ment, the comparison group should be drawn from new ES registrants in the same calendar quarter that participants enroll in the JTPA.

Although ES registrants have several important advantages as comparison group sources, certain disadvantages exist. First, because of recent reductions in federal reporting requirements related to the ES, states are no longer required to submit information on the number of economically disadvantaged applicants registered and served by the ES. Because being economically disadvantaged is the major criterion for JTPA eligibility, and given the importance of ensuring that the comparison group be similar to participants on all characteristics affecting eligibility, it is important that the economically disadvantaged status variable be available for the net impact model.¹⁸ Fortunately, many states apparently have continued to collect information on the economically disadvantaged status of ES registrants. States that no longer collect this information will have to modify the comparison group sample selection procedures, as described below.

A second potential disadvantage to using ES registrants as a comparison group concerns limitations in procedures for retaining historical data on these individuals. In the past, most states have kept automated data with individual ES records, including registrant characteristics and ES services received, for a period of three to five years. In some states, however, individual-level data are purged after approximately one year, and archived backup tapes are not very accessible. In such states, it will be difficult to draw the four quarterly samples retrospectively at one time, as comparison group members for the first quarter cohort would already have been purged. Thus, such states must either draw the comparison samples on an ongoing quarterly or semiannual basis, alter their purging practices, or retain historical data for 18 months to two years.

A final complication encountered in using ES registrants concerns the possibility that the ES registrant file may be dominated by UI claimants. In states in which ES offices are co-located with UI offices, or in which the policy is to actively monitor job search efforts of UI claimants, a large proportion of these claimants may be entered in the ES registrant file. Because of the markedly different labor market experiences of the two groups, it is inappropriate to compare the outcomes of JTPA participants

with a sample that is dominated by UI recipients. Thus, it is necessary to undersample UI claimants in the ES registrant file in certain states to obtain a comparison sample with a proportion of UI claimants similar to the JTPA population.

In adjusting for this problem, note that UI claimant status, as recorded on the MIS systems for ES and JTPA, may not represent the same concept. In particular, for JTPA participants, being a UI claimant typically means that the person has filed a UI claim and has been determined to be monetarily eligible. The ES claimant status refers simply to the filing of a claim for benefits and does not imply monetary eligibility. Because of this difference, a typical JTPA "claimant" is much more likely to receive UI benefits than a typical ES claimant. To ensure that UI reciprocity is comparably measured for the two groups, the UI Benefit History file must be used to determine whether the person was a UI recipient. A decision on the appropriate rate for sampling UI recipients from the ES registrant file would then be based on this measure.

Despite these potential disadvantages, we believe that ES registrants are the best comparison group source among existing state data bases. We now turn to a discussion of additional details related to drawing a sample of ES registrants.

Comparison group sample exclusions. Prior to selecting comparison groups of adult men and women ES registrants, certain cases should be excluded from the sample frame to maintain comparability with participant samples. In addition, it is desirable to exclude those individuals who are clearly not eligible for JTPA and who are likely to have an earnings potential considerably different from JTPA participants. We discuss these sample exclusion considerations below.

To maintain comparability with the JTPA participant samples, the group of ES registrants should be restricted to individuals over 21 and under 65 years of age. If it turns out that no one in the JTPA sample is over a given age (for example, age 55), then the ES registrant sample should be similarly restricted. Also, ES registrants must be excluded from the sample if they are JTPA participants during either the preprogram, program, or postprogram period. This problem, known as com-

parison group contamination, results in comparing outcomes of program participants with outcomes of other individuals who have participated in the program. Such a comparison yields biased net impact estimates, and results in understating the true impacts of the program. To minimize this problem, one should compare the Social Security Account (SSA) numbers of current and recent JTPA participants with the SSA numbers of ES registrants, and exclude all matches from the comparison sample.

A final issue concerns procedures to ensure the similarity of participants and comparison group members on characteristics related to JTPA eligibility. As indicated above, the primary criterion for JTPA eligibility is that the person be economically disadvantaged. Over the last few years at least 95 percent of adults in Title II-A programs have met this criterion. Moreover, of those who are not economically disadvantaged, or who were not certified to be economically disadvantaged, the act requires that they be persons facing demonstrated employment barriers. Because virtually all adult Title II-A enrollees are economically disadvantaged, the comparison group should also exclude all ES registrants who are not economically disadvantaged at application. This will help ensure that comparison groups are similar to JTPA participants on the key characteristic related to JTPA eligibility.¹⁹

Selecting the comparison group samples. We recommend that the selection of stratified random samples of adult male and female ES registrants have the same distributions as JTPA participants on certain key characteristics. This approach maintains maximum statistical power for the desired sample design, while ensuring that the participant and comparison samples are similar.²⁰

Because of program eligibility considerations and certain practical issues concerning the relationship between the ES, UI, and welfare programs, some of the more important characteristics on which to ensure participant and comparison group comparability are economically disadvantaged status, receipt of UI benefits, and receipt of AFDC benefits. Because comparability between the two groups on economically disadvantaged status will be ensured by excluding from the sample frame for the comparison group all new ES registrants who are not disadvantaged, no additional matching is required on this characteristic.

A simple random sample would probably include relatively too few ES registrants who are AFDC recipients, and relatively too many ES registrants who are receiving UI benefits. For example, nationally, only 9 percent of adult men and 35 percent of adult women JTPA terminees in PY 84 were receiving AFDC at application, and 15 percent of adult men and 8 percent of adult women JTPA terminees in PY 84 were UI claimants at application. On the other hand, it is probable that fewer than 10 percent of all ES registrants are AFDC recipients and, in states in which the Job Service is co-located with UI, the fraction of ES registrants who are likely to be UI claimants could approach 50 percent.

To ensure similarity on these important characteristics, comparison group members should be randomly selected from the sample frames of adult men and women. Thus, for the separate samples of adult men and women, procedures would be used to make certain that the participant and comparison groups are similar on the proportions in the four cells representing combinations of AFDC and UI recipient status. Operationally, for a given total sample size of participants and comparison group members, sampling rates for each cell would be determined to match the two distributions, and then comparison group members would be selected randomly from the cells at the given sampling rates as described in the next section.

Sample Sizes for Participant and Comparison Samples

An important element of the research design is the determination of the appropriate sample sizes for participant and comparison groups. As we indicated earlier, many states will have little choice concerning participant sample size. Because the marginal cost of increasing sample size is very low, even medium to large states should generally use the largest numbers of participants and comparison group members feasible. In states with very large JTPA programs, however, samples should be drawn. This raises the issue of total sample size as well as allocation of total sample among the participant and comparison groups.

The appropriate sample size for the net impact analysis ultimately depends on the size of the impact that is important to detect for policy purposes and the level of statistical accuracy required. With larger

sample sizes, one has greater assurance of detecting small differences in overall outcomes between the participant and comparison groups, as well as detecting differences for major participant subgroups or among program activities. The likelihood of detecting a given difference in outcomes also depends on the allocation of the total sample between the two groups and the unexplained variance of the outcome measure, such as earnings or AFDC grants. Thus, although the sample size requirements will differ for net impact evaluations that rely on other outcome measures, in general the more homogeneous the sample, that is, the smaller the variance of the outcome measure, the smaller the number of cases necessary to detect a given difference in outcomes at a specified level of statistical significance.

Based on a number of considerations, a total *analysis sample* of 12,000, divided equally between adult men and women, participants and comparison group members (that is, 3,000 each) should meet most states' analytical needs.²¹ This sample size recommendation refers to the final analysis samples and, because some cases will be omitted for various problems described above, initial samples should be somewhat larger.

States interested in obtaining more precise net impact estimates for subgroups of adult men or women, should consider larger sample sizes as needed. In addition, states with relatively small JTPA programs (i.e., fewer than 1,000 adult enrollees per year) should be very careful in interpreting the results, as only very large impacts are likely to be judged as significantly different from zero. As a result, such states might consider pooling samples over time to increase sample size and thereby enhance the reliability of the net impact findings.

Finally, although a total analysis sample size of 12,000 should be adequate to meet most states' needs, the appropriate sample size depends on several factors and there is no size that is correct under all circumstances. States that are unsure as to the appropriate sample sizes to use in a net impact analysis should discuss their concerns with a sampling expert.

Data Collection and Processing Plan

A variety of data collection and processing tasks must be conducted in support of the overall research design. Quarterly samples of participants and comparison group members must be drawn. Preprogram, program, and postprogram data must be obtained from JTPA, ES, UI, and PA (MIS) records. These must be merged with individual participant and comparison group records. All data must be cleaned, and certain cases may need to be excluded. Analytic variables must be specified, and procedures must be implemented to deal with missing information. Finally, analysis files must be developed. This section reviews the various data collection and processing tasks that must be conducted. Readers interested in additional detail are referred to the implementation guide on net impact evaluation in the series of evaluation guides listed in the appendix.

Although none of the data collection and processing tasks outlined in this section is particularly difficult, the overall magnitude of the undertaking is considerable. Moreover, there is a major coordination issue, since many of the tasks must be performed by staff of several different agencies or subagencies. The size and breadth of the data-related tasks have two important implications.

First, there must be active and continuing cooperation and support on the part of several state agencies and subagencies. Because these agencies have different policies and priorities concerning issues such as data confidentiality, any issues of concern must be resolved at the outset. A regular data collection and processing schedule must also be established. Lack of support on the part of any of the agencies involved will considerably reduce the value of the net impact results and could render them useless.

Second, one person should be given the responsibility of managing the full range of tasks and the authority to obtain the necessary staff and computer assistance. The magnitude of the data processing tasks and the involvement of multiple agencies make these conditions particularly important for managing the data processing system. Although managing this effort is likely to be a time-consuming activity in the initial stages, once the system is in place and the individual tasks become routine, the management time required will decline considerably.

Table 2.4 shows the various data collection and processing tasks involved, from sample selection to preparing an analysis file for estimating the net impacts of JTPA. To increase clarity, we have chosen to present as separate steps some activities that could easily be combined into one step. In addition, note that there are likely to be unanticipated data-related issues and problems. To minimize such problems, it is particularly important for researchers and key data processing staff members from each of the involved agencies to meet frequently. Frequent communication helps identify idiosyncracies in the systems, which could produce noncomparable data for certain types of individuals. This communication may produce potential solutions for such problems as well.

Table 2.4
Overview of Data Collection and Processing Tasks

-
- Select quarterly samples of JTPA participants and obtain, merge, and process preprogram data from various sources for these participants
 - Select expanded quarterly samples of new ES registrants who are economically disadvantaged to serve as comparison group members
 - Obtain, merge, and process preprogram data from various sources for the expanded samples.
 - Select quarterly comparison groups of adult men and women ES registrants from the expanded samples to match the distribution of participants on four cells comprising combinations of UI recipient status and AFDC recipient status.
 - Merge the quarterly preprogram data files that include all of the data obtained in the above steps for the samples of participants and comparison group members
 - Create separate annual preprogram analysis files for adult men and women from the quarterly preprogram data files. This involves merging the quarterly files, editing the data, creating analytic variables, and implementing procedures to handle cases with missing data.
 - Obtain program and postprogram outcome data (i.e., AFDC Grants Records and UI Wage Records) for all quarterly samples of participants and comparison group members and create appropriate outcome variables
 - Obtain data on JTPA services for participants in each of the quarterly samples and create appropriate activity or service variables
 - Create separate annual net impact analysis files for adult men and women by merging the preprogram analysis files with the outcome and treatment variables.
-

Selecting Participant Groups

The first data collection and processing task involves participant group selection. As noted above, samples of adult men and women JTPA Title II-A enrollees must be selected on a quarterly basis. In table 2.5, we list the steps used to select appropriate participant groups for a particular quarter. Minor modifications may be made to accommodate states that desire larger samples.

Although the procedures listed in table 2.5 could be used to select participant groups in any state, some of the steps may not be necessary in certain areas. Some states may need to alter the procedures slightly to meet their needs. For example, because the statistical precision of the net impact analysis is not very sensitive to moderate sample size differences, given the large sample sizes involved, a state that serves adult men and women in JTPA in approximately equal proportions could omit the step involving sample stratification by gender. However, because minimal effort is required to stratify the samples, and the models are to be estimated separately by gender anyway, prestratification by sex is recommended.

The fourth step in this list concerns the procedures involved in selecting the participant samples. Specifically, we suggest that a range of two-digit numbers (00-99) be specified (with the size of the range dependent on the sample size) and that the last two digits of the person's SSA number be used to select the sample, since it is a random number.

Table 2.5
Steps in Selecting the Participant Sample

-
1. Create a file of all persons who enrolled in JTPA Title II-A programs in any SDA during the quarter that includes SSA number, age, and sex
 2. Exclude those who are under age 22 or over age 64 (or perhaps 55 if serving older persons is rare).
 3. Create separate subfiles for adult men and women.
 4. Select a random sample of adult men and women from the two separate subfiles using the last two digits of the SSA number, which are random numbers. The size of the quarterly samples should reflect seasonal differences in enrollment and be such that the final analysis samples for the program year, after excluding cases for missing data, include at least 3,000 adult men and women each
 5. Obtain preprogram PA Grants Records and UI Benefit History data and create measures of AFDC recipient status and UI recipient status at enrollment
-

Only individuals whose SSA final-digit numbers fall into range would be included in the sample. In addition, to account for potential seasonal differences in JTPA enrollments, one should select a fixed proportion of enrollees in each of the four quarters, using data on enrollments in the prior year to determine appropriate proportions. This is superior to selecting an equal number in each quarter.²²

Given an estimated 5 percent sample loss due to missing data, the recommendation in step 4 translates into the initial selection of approximately 3,200 adult men and women each to yield an analysis sample of 3,000 each. States with more severe missing-data problems would have to select larger initial samples. Finally, in step 5, key preprogram data are obtained to use in developing the matched comparison groups. The data sources to be used and the specific measures to be developed are discussed below.

Selecting Comparison Groups

As described earlier, the comparison groups are developed from new ES registrants in areas served by the SDAs. Table 2.6 provides an overview of the steps that ES data processing staff could use in selecting appropriate comparison groups of adult men and women for a particular quarter of JTPA enrollees. Repeating these steps for the subsequent three quarters yields matched comparison group members for the entire program year.

The first four steps listed in table 2.6 are designed to yield a sample frame that is somewhat more comparable to JTPA participants than a sample of all ES registrants. Steps similar to these would also be used to develop matched comparison groups. These initial steps mirror the first three steps for selecting the participant samples.

The fifth step addresses the need to expand the initial sample of ES registrants to overcome the sample loss expected at steps 7 and 8. Specifically, the size of the initial samples must be large enough ultimately to yield sufficient numbers of ES registrants who have not participated in JTPA in each of four cells defined by combinations of UI recipient status and AFDC recipient status. This assures that a comparison group with a similar distribution of these characteristics can be

Table 2.6
Steps in Selecting the Comparison Group

-
1. Create list of all ES offices located in areas served by SDAs in the state.
 2. Create a file of all persons who were new registrants in these ES offices during the quarter that includes SSA number, age, sex, and whether the person is economically disadvantaged.
 3. Exclude those who are: (a) not economically disadvantaged, (b) age 21 or less, or (c) who are older than the oldest individual in the quarterly JTPA enrollee sample.
 4. Create separate subfiles of adult men and women.
 5. Select an expanded random sample of adult men and women new ES registrants from the two separate subfiles. As a general guideline, approximately 2,500 each of adult men and women each quarter should be sufficient.
 6. Obtain available AFDC Grants Records and UI Benefit History data and create measures of AFDC recipient status and UI recipient status at enrollment for all ES registrants identified in step 5.
 7. Exclude persons who are currently participating in JTPA or who participated in JTPA during the prior year based on a match of SSAs.
 8. Select separate random samples of adult men and women that match the distribution of participants on the four cells comprising the combination of the comparable measures of UI recipient status and AFDC recipient status.
-

drawn. Because the expanded initial sample size is likely to vary considerably from state to state depending on individual characteristics, local economic conditions, and the state policies concerning the relationships among the local ES, AFDC, and UI offices, it is very difficult to provide precise guidelines. As a starting point, we recommend initial samples of 2,500 adult men and 2,500 adult women ES registrants be selected in each quarter.

The next step involves obtaining certain preprogram agency data and developing measures of AFDC recipient status and UI claimant status at enrollment. After excluding current or recent JTPA participants from the comparison group sample frame, based on matching SSA numbers (step 7), the final step involves allocating the remaining SSAs to the four recommended stratification cells comprising combinations of UI recipient status and AFDC recipient status. Sampling rates for each cell would be used to match the quarterly distribution of participants on these key characteristics.

These procedures can be used by most states, but some of the specific steps may need to be modified to meet various states' analytical needs,

data limitations, or specific circumstances. For example, significant modifications to the procedures may be necessary in states that do not have all of the required data elements in their ES MIS system, in particular, data on whether an ES registrant is economically disadvantaged. Because being economically disadvantaged is the primary eligibility criterion for JTPA, it is important to develop procedures to select comparison group members who are also economically disadvantaged.²³

In the absence of specific information on economically disadvantaged status, an alternative approach, used extensively in the literature, involves excluding individuals with very high preprogram earnings and then explicitly matching the remaining comparison group members to participants on the basis of preprogram earnings. That is, instead of excluding all persons who are not economically disadvantaged at step 2, one would first obtain UI Wage Records for a much expanded sample at step 5—perhaps up to five times as large if only 20 percent of ES registrants are economically disadvantaged. Then, exclude all persons with high earnings in the immediate preprogram period, who would certainly not be eligible for JTPA. The precise cutoff level is a matter of judgment and depends on the distribution of preprogram earnings in both samples. As a general guide, however, a cutoff level set at the maximum earnings of participants (separately for adult men and women) in the six months before enrollment or somewhat higher (for example, one standard deviation), seems reasonable.

A final issue in selecting the matched comparison groups involves the development of consistent measures of AFDC recipient status and UI recipient status at enrollment. Because ES and JTPA data bases will not generally collect comparable data on these two factors or maintain the enrollment values in the MIS, one must develop consistent measures of these items from the same data set in order to develop appropriately matched groups. The recommended approach for developing these measures is discussed below.

Obtaining Preprogram Data for Participants and Comparison Group Members

An integral step in selecting matched comparison groups for a specific program quarter of JTPA Title II-A adult men and women participants

involves the processing of several different data elements from agency records. In addition to JTPA and ES enrollment/registration data, AFDC Grants Payment Records, and UI Benefit History Records, it is important to obtain UI Wage Records for the preprogram period soon after the participant group is selected. Timely acquisition of these data is necessary because (1) some states do not retain much historical MIS data, thus, the sooner the data are obtained, the longer the preprogram period covered; and (2) some of these preprogram data items are used to develop a profile of individual characteristics that in turn is used to select comparison group members similar in these characteristics.

As a result, though the steps for obtaining data for the preprogram period for comparison group members are identical to the steps for JTPA enrollees, the timing and magnitude of the task differ considerably. The various data elements to be obtained and merged with the quarterly samples of participants and comparison group members are described below. Some problems that may be encountered in this process are also identified.

As indicated in table 2.7, the first two sets of data elements for individuals in the analysis come from JTPA application information and the ES application form. Although only a few items from these forms are used in support of the steps listed in this table (e.g., age, sex, disadvantaged status), it is useful initially to extract *all* application data from the JTPA and ES MIS systems for those individuals selected into the quarterly samples for some analysis purposes. Although only the data items that can be regarded as comparably measured for participants and comparison group members will be used for net impact analysis, all JTPA application data should be obtained in case states are interested in using the net impact design to estimate gross program impacts, or to examine whether assignment of program activities (services) to individuals depends on other measured characteristics available for participants. Moreover, all ES application data should be obtained to get a better sense of the characteristics of this group and how they are likely to differ from the characteristics of the individuals in the participant group.

Detailed preprogram UI Wage Records and AFDC grants histories are particularly important to a net impact analysis. Ideally, three years of

Table 2.7
Obtaining Application and Other Preprogram Data

-
1. Obtain JTPA enrollment information for each participant: age/sex used in selecting sample (see steps 1-4 of table 2.5); other items used to develop variables for net impact analysis.
 2. Obtain ES registration data for each potential comparison group member. Age, sex, and economically disadvantaged status used in developing sample frame for selecting matched groups (see steps 2-4 of table 2.6); other items used to develop variables for net impact analysis
 3. Obtain preprogram monthly AFDC Grants Records for up to three years for participants remaining after step 4 of table 2.5 and for comparison group members identified in step 5 of table 2.6, and create (a) variables measuring quarterly totals and (b) an AFDC recipient-status indicator, defined as 1 if the person received AFDC grants in the month before, during, or after the month of JTPA/ES enrollment or application, and 0 otherwise.
 4. Obtain UI Benefit History data for the calendar quarter before and after enrollment for each person in step 3 above, and create (a) total UI benefits received during the quarter prior to enrollment or application and (b) a UI recipient-status indicator, defined as 1 if the person received UI in the month before, during, or after the month of enrollment or application, and 0 otherwise.
 5. Obtain up to 12 quarters of preprogram UI Wage Records, and create totals.
-

preprogram data are needed. There are generally six to 12 quarters of UI Wage Records available at any one time, with approximately a three- to six-month lag before these data are complete. As a result, to ensure that wages for the immediate preprogram quarter are complete for the analysis, it is necessary to update the data for this quarter when the postprogram earnings data are obtained.

One must also obtain AFDC grants received by participants and comparison group members. In addition to obtaining preprogram monthly grants records for up to three years to serve as important control variables in the net impact models, in order to define welfare-recipient status similarly for JTPA participants and ES registrants, data on AFDC grants received for the month after enrollment month must be obtained for all individuals. To minimize problems caused by differences in the length of time from JTPA or ES application to AFDC enrollment, or caused by differences in recording practices among different agencies, a JTPA enrollee (or ES registrant) should be defined as an AFDC recipient if the person received AFDC grants during the calendar month prior to enrollment or registration, the month of enrollment or registration, or during the month after enrollment or registration. Similarly, a UI-recipient-status indicator can be developed using the same approach with data from the

UI Benefit History file. Although these definitions may differ from JTPA or ES definitions of welfare-recipient status or UI claimant status, they will enable one to generate a comparison group that is statistically similar to the participant group on this important characteristic.

The PA MIS systems in some states may present obstacles to obtaining accurate preprogram AFDC Grants Records for certain types of individuals. For example, in attempting to develop preprogram measures of grants paid, because of changes in household status and other factors, the data base may not allow one to verify that a specific person was in a particular assistance unit throughout the three-year period. This may be true even though it may be possible to identify up to three prior years the preprogram monthly grants paid to a particular assistance unit for a given individual currently in that unit. As a result, the preprogram AFDC history for that unit may not accurately reflect a person's welfare-recipient status during that period. This is particularly a problem for individuals who experience a marriage or divorce, or who change living arrangements.

To obtain accurate information on the preprogram AFDC status of participants and comparison group members, such problems must be overcome to the extent possible. A potential solution available in Washington State would involve using the "Warrant Roll Extract Files," which contain a record of all AFDC payments made each month, and a list of all SSA numbers in the household that month. These files could be linked over time to determine which assistance unit the person of interest was in, and to obtain the correct preprogram data.

Preparing a Preprogram Analysis File

The next task is to develop preprogram analysis files—created from the annual preprogram data file—for adult men and women. Once the preprogram analysis files are developed, the analytic procedures described in later sections can be implemented to investigate the comparability of the JTPA enrollee and comparison groups in the preprogram period.

Before describing issues involved in conducting these tasks, one important feature of the preprogram analysis file should be noted.

Specifically, because preprogram data elements are defined in terms of their relationship to the *quarter of enrollment*, elements in the same fields on the analysis file will correspond to different *calendar periods* for individuals who enrolled in different quarterly files. For example, data elements for the immediate preprogram quarter, on the file for the enrollees in the first quarter of a program year, will correspond to the period of the second preprogram quarter for individuals who enroll one quarter later. Before comparing dollar amounts in certain preprogram quarters across files, therefore, one must adjust for overall price changes by translating all nominal dollar amounts into real terms.²⁵

Editing data files. Although considerable cleaning and editing will have been performed by the respective ES and JTPA data processing staffs as part of their normal procedures, one must conduct edit checks to become familiar with different files and to check data quality. The first type of edit check compares the results of a simple frequency distribution on all variables in each of the annual files with a range of acceptable values.²⁶ For other data elements, such as UI wages, UI benefits, and AFDC grants, a range of acceptable values should be created that incorporates rough estimates of the maximum amounts that can be received from certain programs in the state.

Although some errors are obvious by inspection of a single data item, other errors may not be apparent except when viewed in combination with another data item. Thus, as a second edit check, limited cross tabulations must be carried out regarding certain items, to identify additional potential data quality problems. For example, cross tabulations of age by education could identify 22-year-old individuals with 19 years of education, an unlikely occurrence. It is also useful to cross tabulate earnings and AFDC grants received in the same preprogram quarters. The presence of individuals with large values for earnings and AFDC payments in a given quarter may be indicative of data errors or other problems.

Specifying analytic variables. The analytic variables specified should be comparably measured for enrollees and comparison group members. For variables derived from a common source, such as AFDC Grants Records, UI Wage Records, and UI Benefit History data, comparability

should not be a problem. However, measures of personal characteristics will be obtained from both the ES and JTPA MIS, and differences in the ways in which questions are asked or answers are recorded can present major obstacles to defining comparable variable definitions. Moreover, even when questions and response codes appear to be the same, the information collected may correspond to slightly different concepts due to differences in staff instructions and training. Because of these potential problems, one must review the application forms to both the ES and the JTPA and the corresponding handbooks that provide instructions for recording answers to each question, and resolve remaining issues through discussions with appropriate agency staff.

At a minimum, one should develop comparable measures of age in terms of years, a set of dummy variables for race/ethnicity and sex, a dummy variable for veteran status, and a set of dummy variables for occupation, based on the first digit of the DOT code. In addition, one should develop limited indicators for educational background that are comparably measured. For example, the ES application form in most states generally collects education information in the form of the highest grade of schooling completed (from 0 to 19 years); whereas, the JTPA application form often records an individual's education status in terms of one of the following four codes: (1) school dropout, (2) in school (high school or less), (3) completed high school or received GED, and (4) currently attending or has attended schooling programs beyond high school. With such information, however, it should be possible to recode values from the ES application form to specify separate dummy variables for whether the person is (a) not a high school graduate (i.e., 0-11), (b) a high school graduate (i.e., 12), and (c) has completed additional schooling beyond high school (i.e., 13 or more). Every effort should be made to implement procedures such as these whenever possible to define comparable measures of variables for both groups.

Procedures for handling cases with missing data. In general, the variables used in the analysis should not suffer from major missing data problems. However, in instances where independent variables (e.g., age, education) have missing values, it is preferable to adjust for the missing variable in question rather than exclude all cases that have missing data

on any relevant variable. Although there are several alternative procedures that one can use to create substitute values for missing data, the gains from using an elaborate system to fix relatively minor problems is not likely to be worthwhile.²⁷

As a result, the mean values of the independent variables—calculated separately for participants and comparison group members, and, of course, separately for adult men and women—should be used for cases with missing data. Thus, as part of the initial analysis task, one should calculate the means of all independent variables separately for participants and comparison group members on the analysis files, and prepare recode statements that set the value of a variable equal to the appropriate mean whenever it is missing. In addition, if differences in the independent variables among the quarterly samples are likely to occur, one should consider using means calculated separately by quarter of enrollment or registration to capture trends in these variables over time. The treatment of missing data in the program participation or service variables is discussed below.

Obtaining and processing during-program and postprogram outcome data. Postenrollment UI Wage Records and AFDC Grants Records must also be obtained for both the participant and comparison groups. Then, appropriate variables must be specified and merged onto the preprogram analysis file. In addition, UI Wage Records must be obtained for the immediate preprogram quarter for all participants and comparison group members to correct for potential measurement error problems due to obtaining data “too early” for that period.

No problems are anticipated in collecting quarterly UI Wage Records in the postenrollment period. The necessary information can generally be obtained from a single request made at the very end of the data collection process. In states that retain only six quarters of UI Wage Records, an intermediate request must be made no later than 18 months after the month in which the first quarter of individuals were enrolled, to ensure that the entire history can be obtained for early enrollees.

With regard to collecting monthly AFDC Grants Records during the program and postprogram periods, it may be preferable to obtain such information on an ongoing basis, rather than only once at the end of the

data collection process. In working with PA data systems in which the Recipient History File does not enable one to identify whether a person is in the particular assistance unit throughout the period of interest, it is preferable to obtain the information on an ongoing basis each month to minimize measurement error. Obtaining information on this basis requires that for each of the subsequent months the SSA numbers of participants and comparison group members must be compared with the list of SSA numbers in assistance units that received AFDC payments during the month. The actual values of monthly grants would be included for the SSA numbers that matched, and zeros would be included for those SSA numbers that did not match. Quarterly and annual values would then be calculated as the sum of monthly values.

Note that by stopping the collection of these agency data at a single point, one will obtain eight quarters of UI Wage Records and AFDC Grants Records for participants and comparison group members who enrolled or registered during the first quarter of a program year, seven quarters of data for those who enrolled or registered during the second quarter, six quarters of data for those who enrolled or registered in the third quarter, and five quarters of data for those who enrolled or registered during the last quarter of the program year. Each set of quarterly earnings and AFDC grants data would include one quarter for the actual quarter of enrollment or registration, two subsequent quarters of data that are likely to include program earnings for some participants, and varying postprogram quarters of earnings records. Because of the usefulness of having rectangular analysis files, a common number of postenrollment quarterly values should be created for all individuals on the file, and missing data codes (e.g., -9s) placed in postprogram quarters for later enrollees for whom data are not yet available. One could, of course, subsequently obtain actual values for these quarters and replace the missing data codes.

Obtaining and Processing JTPA MIS Data

Because individuals can participate in multiple activities, most state JTPA data systems will have a program activity file in which a given individual may have multiple records. To specify consistent analytic

variables, program activity records must be extracted for all SSAs in the four quarterly samples of JTPA enrollees. Depending on the archiving procedures followed in a given state, it may be possible for these data to be obtained from a single request made at the end of the data collection process, and hopefully after all or almost all individuals have terminated from JTPA.

After the program activity records are obtained, the next step involves specification of variables to represent services received by JTPA participants. In general, it should be possible to develop relatively detailed indicators of the services received from JTPA, related to the type of occupation and length of training, and whether the person completed training. These variables are described in the conceptual framework section. Once these variables are specified, they should be merged with the preprogram analysis files for participants, and zeros must be entered for comparison group members for all of these variables.

If participants have missing data on key JTPA treatment variables, they should be excluded from the analysis samples, provided the reason the items are missing appears to be random (i.e., not systematically related to the likely net impact of the program). To make this determination, one must compare the characteristics of participants who have missing data on the variables (e.g., length of participation) with the characteristics of participants who have data on the variables. For example, one should compare the age, race, and education of individuals in the two groups of participants to determine if there are major differences.

Moreover, if in the process of collecting data on program experiences one also obtained information on placement status at termination, it is useful to compare JTPA enrollees on their placement status at termination to judge whether having missing data is systematically related to the impact of the program. If enrollees with missing data on JTPA experiences are equally likely to be placed in jobs following the program as enrollees with complete data, this would provide additional confidence that the validity of the analysis will not be compromised by excluding such individuals.

Once the JTPA analytic variables have been developed and decisions

made on the treatment of missing data, the final task involves merging these variables and the during-program and postprogram outcome variables to the preprogram analysis files for each individual by SSA number. This results in the creation of net impact analysis files for adult men and women. These files will support all of the analysis tasks described in the next section.

Data Analysis Plan

A data analysis plan must be developed to examine the adequacy of the comparison groups selected, and to use the comparison groups to estimate the net impacts of the program on the outcome measures specified in the conceptual framework. This section presents an overall strategy for obtaining valid estimates of the net impacts of JTPA programs on the postprogram earnings and welfare dependency of adult men and women enrollees.

Before describing the details of the plan, we want to emphasize that the recommended approaches should be quite accessible to all states interested in conducting net impact analysis of JTPA or of other social programs. For example, all of the analysis techniques to be used are contained in standard statistical software packages—such as SAS and SPSS—that should be readily available and familiar to state-level analysts. In addition, after some initial data processing on a mainframe computer, it may be possible to download the analysis files to hard disks that can be accessed by minicomputers. This will minimize the computer resources required to conduct the analysis.

An overview of key elements of the data analysis plan discussed below is shown in table 2.8. We first describe an analysis strategy for examining the adequacy of the comparison groups selected in order to get a better understanding of the direction and magnitude of potential selection bias. We then describe a statistical model that can be used to estimate the average net impacts of JTPA and the impacts for important subgroups. The section concludes with a discussion of potential adjustments for certain data and design deficiencies.

*Examining the Adequacy of the Comparison Groups:
Obtaining Evidence on Selection Bias*

If the samples of JTPA participants and ES registrants developed through the steps described in the previous sections are similar on both measured (e.g., age, race, education) and unmeasured (e.g., attitude toward work, motivation) characteristics, then valid inferences about the impacts of JTPA programs can be drawn from such comparisons. However, whether an individual participates in a social program is likely to depend on both individual and agency decisions.

Table 2.8
Overview of the Data Analysis Plan

-
- Examine the adequacy of comparison groups using analysis techniques such as differences in means, differences in distributions, and multiple regression analysis. The adequacy of the comparison groups will be judged in terms of three criteria:
 1. Similarity of participant and comparison groups on measured characteristics (e.g., age, race, education)
 2. Similarity of participant and comparison groups on preprogram earnings and AFDC grants.
 3. Similarity of the relationships between preprogram earnings (and AFDC grants) and measured characteristics for participants and comparison group members
 - Estimate average net impacts of JTPA for adult men and women using an autoregressive model. Net impacts will be estimated for four postprogram outcome measures—earnings, whether employed, AFDC grants, and whether an AFDC recipient—in each of three different postprogram periods: three months, six months, and 12 months.
 - Estimate net impacts of JTPA on the various outcome measures for adult men and women and key subgroups using autoregressive models. In addition to sex, the subgroups of interest include:
 1. Participant characteristics such as race or ethnicity, age, education, and welfare recipient status for women.
 2. Program activities such as CT, OJT, JSA, and all other activities.
 3. Program length of stay.
 - Adjust net impact estimates to the extent possible for data and design deficiencies:
 1. Contamination of the comparison groups.
 2. Uncovered earnings.
 3. Selection bias.
-

For example, JTPA participants must decide to apply to the program, meet certain legislated eligibility criteria, be selected by the agency for program participation and assigned a program activity, and decide to accept that assignment and enroll in the program. Although ES registrants do not have to meet any formal eligibility criteria, certain individuals, such as those receiving benefits from government programs such as UI, are required to register with the ES, and some offices follow selective registration policies. Furthermore, whether an ES registrant receives ES services depends on several factors, including the availability of suitable job openings, and the person's qualifications and persistence. Because of these various selection processes, it is unlikely that the resulting samples of JTPA enrollees and ES registrants who do not receive services are truly equivalent on both measured and unmeasured characteristics. This is the issue of *selection bias*, which is highly likely to be present in evaluations of other social programs as well.

All nonexperimental approaches to evaluating the net impact of a social program will probably contain a certain amount of bias. That is, the formal conditions required to ensure unbiased estimates of program impacts are not likely to be met, even if one had extensive data on the characteristics of program enrollees and comparison group members. This is true for the proposed research design. As a practical matter, therefore, one should not focus on the fact that the two groups are not identical, but identify the major dimensions on which the groups differ and determine the extent to which the net impact estimates are likely to be sensitive to those differences.

As indicated in table 2.8, three different criteria can be used to judge the adequacy of the comparison groups selected:

1. Similarity of the JTPA enrollee and comparison groups on measured individual characteristics.
2. Similarity of the JTPA enrollee and comparison groups on preprogram earnings and AFDC grants.
3. Similarity of the relationships between preprogram earnings (and preprogram AFDC grants) and the measured individual characteristics of JTPA enrollees and comparison group members.

Although these are the traditional criteria for judging the adequacy of

nonexperimentally derived comparison groups, they are necessary, but insufficient, conditions for overcoming selection bias. Even if the comparison groups selected generally meet these criteria, this should not be interpreted as definitive evidence of an absence of selection bias. With this caution in mind, some analyses are outlined below that can be performed for each of the quarterly samples, and for the annual sample as a whole, to see whether these criteria are met. If they are not met, the analysis identifies the types and extent of differences between the groups. These factors must then be kept in mind when interpreting net impact results.

The first criterion (Criterion 1) is the similarity of the two groups on measured characteristics at enrollment or registration. It is particularly important to compare the participant and comparison groups on available measured characteristics known to affect earnings and AFDC grants. For example, it is particularly useful to determine to what extent the two groups differ on age, race, education, occupation, and handicapped status, and other relevant personal characteristics that are comparably measured for both groups.

Using standard software packages, one would compare the means and the distributions of these measured characteristics for participants and comparison group members (separately for adult men and women) in each of the four quarterly samples and in the overall program year sample.²⁸ Because the output from standard software analysis packages normally includes the results of *t-tests* and *Chi-square* tests for formally testing the equivalence of the means and distributions of variables in two samples, it is straightforward to compare the similarity of the participant and comparison groups on all measured characteristics.

Similar analyses should be conducted across JTPA program activities. That is, one should not only compare the characteristics of participants to the characteristics of comparison group members, but also compare the characteristics of participants with respect to program activities received, such as CT, OJT, and JSA. This will indicate any additional selection bias arising in estimating net impacts by separate program activity. For example, if one determined that more motivated or energetic individuals were being sent to OJT, the net impacts of this

program activity would be somewhat inflated because of this assignment process. On the other hand, if there were relatively few differences in measured characteristics by program activity, this evidence would provide some confidence that no additional selection biases would be introduced in deriving estimates of the net impacts by program activity.

The second criterion (Criterion 2) to judge the adequacy of the comparison groups is the similarity of the key outcome measures of participants and comparison group members in the preprogram period. This involves a test of whether a significant difference exists in the preprogram earnings and AFDC grants of the two groups, controlling for measured characteristics. Such a test provides valuable evidence on whether the two are comparable on the basis of the lagged dependent variables or, in other words, whether there are differences in the outcome variables between the groups in the preprogram period that are due to unmeasured characteristics.

If there are any differences in adjusted preprogram earnings or AFDC grants between the two groups, then this analysis will also provide evidence as to the direction and magnitude of the selection bias. For example, the extent to which JTPA participants have larger (smaller) adjusted preprogram earnings than ES registrants provides some indication as to whether they are more (less) advantaged on the basis of unmeasured characteristics. Moreover, the size of the estimated difference is a reasonable estimate of the amount by which the net program impacts could be overstated (understated) if the difference persisted in the postprogram period.

To formally test for differences in the preprogram earnings and AFDC grants of participants and comparison group members, one would estimate ordinary least squares regression equations (separately for adult men and women) with preprogram earnings and AFDC grants as dependent variables. Multiple regression is a technique that estimates the independent influence of each characteristic on a particular dependent variable, controlling for the influence of all other characteristics in the equation. For example, differences in earnings among individuals may result from differences in education and other personal characteristics, such as age or race, as well as differences in local unemployment

conditions. The regression technique controls for the influence on earnings of local unemployment conditions and other personal characteristics, and estimates the independent influence of all these factors as well as program activities on earnings. All standard software analysis packages include multiple regression programs capable of handling the analysis tasks described in this section.

For a given set of outcome measures, the principal task in specifying the regression equations to be estimated is making decisions about which variables to include in a given model. Because the objective of the analysis is to identify whether there are significant differences in the preprogram earnings and AFDC grants of the two groups after controlling for measured characteristics, there are several guidelines that can be used in making decisions concerning the independent variables to be included in the models.

First, include in the model all personal characteristics of the individuals at enrollment or registration who were examined as part of the analysis conducted for Criterion 1 above, such as age, race, education, occupation, and handicapped status. An exception will be those who must be omitted because too few cases exhibit that characteristic, or those who must serve as the “left-out category.” For example, it is likely that in many states there will be too few of certain minorities (e.g., Native Americans) to include them as separate variables in the model. As a result, one may need to collapse the five race or ethnicity group variables into three variables, i.e., dummy variables for white, black, and other race or ethnicity status.

Note also that in estimating the regression model, one of the race dummies must be omitted to serve as the “left-out category” (reference category) for comparison purposes. If the dummy variable for white status is omitted from the equation, then the coefficients of the other two dummy variables would represent the effect of being in that particular group, relative to being white, on the dependent variable. For every set of dummy variables included in the regression model to capture the effects of a certain characteristic, one of the variables must be omitted to serve as the reference category for comparison purposes. Because the effects of the included variables are all measured *relative* to the left-out

category, the results have the identical interpretation no matter which variable is chosen to serve as the omitted category.

Second, it is important to include previous preprogram measures of quarterly earnings and AFDC grants variables in the model. That is, in examining the comparability of earnings in the immediate preprogram quarter, one should include quarterly earnings and AFDC grants from the second through the twelfth preprogram quarters, given data availability. If, however, one were examining the comparability of earnings and AFDC grants in the immediate preprogram year, then the second, third, and fourth preprogram quarterly earnings and AFDC grants variables would have to be omitted from the regression equation. Such variables are, by definition, part of the dependent variable in this case and, as such, cannot independently affect its value.

Third, variables that are “jointly determined” with preprogram earnings and AFDC should be excluded from the model. Specifically, exclude the dummy variables for AFDC recipient status, UI recipient status, and UI benefit payments in the immediate preprogram quarter from all regression equations estimated over a preprogram period. These variables are essentially other measures of low-income status in the same period and cannot independently affect earnings and AFDC grants in the same period.

A final guideline is to define the variables used according to the appropriate time period. For example, if the dependent variable is earnings or AFDC grants in the immediate preprogram quarter (year), then the unemployment rate in the local area should similarly be defined as the rate for the immediate preprogram quarter (year).

By following these guidelines, one can identify a set of independent variables from those specified, using the procedures described in the data processing section of this chapter. These variables should be included in both the preprogram earnings and preprogram AFDC grants equations. The independent variables to include in a regression model to examine the similarity of participants and comparison group members on earnings (and AFDC grants) in the immediate preprogram quarter are listed in table 2.9. Note that the interactions between the white and the age and education variables are optional and need not be included in the final

model. However, they are included in this list to emphasize the importance of controlling for all measured differences between participants and comparison group members.

For each preprogram period of interest, one would estimate four regression equations that included this set of independent variables. That is, separate models would be estimated for adult men and women and for both of the key outcome measures, earnings and AFDC grants. The test for differences in earnings and AFDC grants between the participant and comparison groups in the immediate preprogram quarter would be based on a *t-test* of the estimated coefficient of the JTPA participant dummy variable. On a more intuitive level, because participation in JTPA during a given period cannot have an effect on earnings or AFDC grants in previous time periods, the coefficient of the JTPA dummy variable in each of the regression models described above should not be statistically significant (i.e., should not be significantly different from zero). The extent to which the estimated coefficients are statistically significant and deviate from zero provides evidence on the direction and magnitude of the likely selection bias.

With earnings in the preprogram period as the dependent variable, statistically significant negative (positive) coefficients on the JTPA dummy would indicate that participants were less (more) advantaged than comparison group members in that period on unmeasured characteristics. If this persisted through the postprogram period it would probably result in understating (overstating) the net impact of JTPA on earnings. Thus, if this analysis indicated that after adjusting for differences in measured characteristics the preprogram earnings of JTPA participants were \$200 less (more) than the earnings of the comparison group, then one might consider adding (subtracting) \$200 to (from) the net impact estimate to adjust for differences in unmeasured characteristics. Note, however, that because preprogram earnings and AFDC grants will be included as independent variables in the net impact model, the extent of this bias should be less in the postprogram period. As such, adjusting the net impact estimate for the total difference in preprogram earnings is likely to overcompensate for the bias due to differences in unmeasured characteristics.

In analyzing the preprogram similarity of earnings and AFDC grants between the two groups, one can examine several different time periods. For the most part, one should be primarily interested in examining the immediate preprogram quarter or year and separate regression equations, like the one described above, could be estimated for both periods.²⁹ In addition, one could also estimate a regression equation like the one described above for each preprogram quarter and derive a set of estimated coefficients of the JTPA participant dummy. To the extent that including additional lagged values of earnings and AFDC grants in the equation serves to reduce the differences between the two groups, the coefficients of the JTPA dummy variable should be largest (in absolute value) in the early preprogram periods and tend toward zero as the preprogram outcome is measured closer to the date of enrollment.

The third criterion (Criterion 3) used to judge the adequacy of the comparison groups is the similarity of the relationships between earnings (and AFDC grants) and individual characteristics for JTPA participants and comparison group members in the preprogram period. This criterion, which is considerably stricter than the previous two, is quite important because, if the same model is generating earnings (or AFDC grants) in the two groups, it suggests that program impacts will be less sensitive to other potential statistical problems. This would provide additional confidence in our ability to obtain unbiased estimates of program impacts.

To test for differences in the preprogram earnings (or AFDC grants) equations of participants and comparison group members, one would estimate a modified version of the regression equation described above to provide information on Criterion 2. Specifically, one would estimate an equation that included all of the explanatory variables listed above, *plus* each of the variables (except the JTPA participation dummy variable) multiplied by the JTPA participation dummy variable. The formal test of whether the earnings and AFDC grants equations in the preprogram period are different for participants and comparison group members is sometimes referred to as a *Chow test* and is based on an *F-test* of the hypothesis that the coefficients of the interaction terms (i.e., the JTPA participant dummy multiplied by each of the other variables in the model) are all zero.³⁰

The three criteria and related analyses described provide considerable information regarding the adequacy of the comparison groups in the preprogram period and the probable biases that must be dealt with. It should be emphasized again that these criteria are relatively strict tests of the comparability of the two groups, and one should not generally expect nonexperimentally derived comparison groups to meet all of them. If the conditions are generally satisfied, then the chances of obtaining unbiased program net impact estimates using standard statistical models are improved. If the criteria are strongly rejected (e.g., *F*-statistics of 10 or 20 when approximately 1.5 is sufficient for rejection), then one should be very cautious in proceeding to estimate net impacts with these comparison groups. Instead, one should first double-check to be sure that the data processing and analysis guidelines described earlier were followed. If the criteria are still strongly rejected, one should then consider obtaining assistance from a researcher familiar with these issues. If, as is most likely, the results are somewhere in between (i.e., preprogram differences between the two groups that are sometimes statistically significant, but not exceptionally large), then one will need to understand the implications of these differences for interpreting and adjusting the net impacts results.

Estimating the Average Net Impacts of JTPA Programs

The four general postprogram outcome measures for the JTPA evaluation are earnings, whether employed, AFDC grants, and whether an AFDC recipient. We will discuss the specific postprogram periods for which these outcomes will be measured for different samples of enrollees, and describe the regression model to estimate average net impacts. A subsequent section will describe how to obtain separate estimates of net impacts for major demographic groups, by program activity and by length of program participation.

Choice of Postprogram Periods and Implications for Potential Additional Sample Exclusions

The choice of the postprogram periods to be examined depends on the distribution of length of stay in JTPA. For example, if no individuals

participated in JTPA longer than six months, then for a given quarterly sample of enrollees (e.g., those who enrolled during first quarter of PY 1985), all such individuals would have terminated from the program by the end of the third quarter of PY 1985. As such, earnings and AFDC grants received during the fourth quarter of PY 1985 would be the outcome measures for the first complete postprogram quarter for these enrollees. If, however, there is considerable diversity in program length of stay and some individuals remain in the program much longer, one would have to decide whether to postpone the analysis and wait until all cases have terminated, or exclude such cases from the analysis samples. Although it is generally not desirable to restrict the participant sample to those who have terminated from JTPA by a particular date (because terminees could differ systematically from nonterminees, which could result in additional selection biases), in most cases it will simply not be possible to wait for all participants to terminate from the program and still provide timely net impact results.

To provide timely results, it may be necessary for states to choose a cutoff date that defines the program period. Any participants who are in the program after that point would be excluded from the analysis.³¹ In general, we expect that defining the cutoff date to be six months after the end of the enrollment period for each quarterly sample (e.g., March 31, 1986, for those who enrolled in JTPA during the first quarter of PY 1985) should be adequate to meet most states' needs. This allows a length of stay that is no less than six months for any individual and up to nine months for individuals who enrolled very early in a particular quarter. We expect that such cutoff dates, applied uniformly to participants in all four quarters of the program year, would result in excluding no more than 10 percent of the participant sample in most states. This is unlikely to significantly bias the average net impacts of JTPA, and should not significantly reduce the precision of the estimated impacts.³²

States that operate JTPA programs that tend to have very long program lengths of stay should consider extending the cutoff date to estimate earnings impacts for the same number of postprogram quarters. On the other hand, in states where JTPA services are relatively brief on average, it may be possible to define a cutoff date that allows for a shorter program

period and, as a result, net impacts can be estimated over a longer program period without delaying the analysis.

As described in the conceptual framework, we recommend that the net impacts of JTPA for adult men and women be estimated on each of the four general outcome measures for a three-month, six-month, and 12-month postprogram period. Based on the data collection plan and the strategy to be used to exclude long-term participants (if necessary), the research design enables one to estimate the net impacts of JTPA on these four measures over a three-month postprogram period for JTPA enrollees from all four quarterly samples within approximately a two-year program analysis cycle. It enables one to provide net impact estimates on these outcomes measured over a six-month postprogram period for the first three quarterly enrollment samples. The net impacts for a 12-month postprogram period can only be estimated for participants who enrolled in the first quarter of the program year.³³ Because of the importance of longer-term impacts in making judgments concerning the effectiveness of employment and training programs, some states might consider collecting additional quarters of postprogram information for all individuals for subsequent analysis, and particularly for those who enrolled in the last three quarters of the program year.

Autoregressive net impact models. To estimate the net impacts of JTPA for adult men and women, we recommend that an autoregressive model be used. Using this approach, ordinary least squares regression equations would be estimated for each of the 12 outcome variables, that is, four outcome measures in each of three different postprogram periods, separately for adult men and women. The autoregressive approach is so named because preprogram values of the outcome measures—quarterly earnings and AFDC grants—are also included as independent variables. This approach has the primary advantage of controlling for any differences in measured characteristics between the two groups that remain after the matched comparison groups are selected, which helps to minimize the problem of selection bias.

To control for potential differences in the characteristics of participants and comparison group members to the extent possible, it is recommended that the net impact regression model be a slightly ex-

panded version of the models used to determine whether the comparison groups meet the preprogram comparability Criteria 2 and 3. The only changes in the independent variables to be included in the basic net impact model, as compared to the variables included in the preprogram models discussed above and listed in table 2.9, are as follows:

Table 2.9
Sample Independent Variables to Include in Model to
Examine Adequacy of Comparison Groups

Demographic and Personal Characteristics

Age
 Age squared
 Black dummy
 Other non-white dummy
 High-school graduate dummy
 Post high-school education dummy
 (Age) x (white dummy)—optional
 (Age squared) x (white dummy)—optional
 (High-school graduate dummy) x (white dummy)—optional
 (Post high-school education dummy) x (white dummy)—optional
 Veteran dummy—for men only
 Handicapped dummy—if measured comparably for both groups

Recent Employment Experiences

Set of eight one-digit DOT dummies for example, allowing professionals to be the left-out category, the eight occupation dummies would correspond to clerical and sales; service; agricultural, fishery, and forestry; processing, machine trades; benchwork; structural work, and miscellaneous

Preprogram quarterly earnings—separate variables for preprogram quarters two through 12, data permitting

Preprogram quarterly AFDC grants—separate variables for preprogram quarters two through 12, data permitting

Labor Market Data

Unemployment rate during the immediate preprogram quarter
 Urban location dummy

Program Participation Variables

JTPA participant dummy (or alternatively, separate dummy variables for program activities)

Other Variables

Set of dummies for the quarter of enrollment or registration for instance, allowing the first quarter to serve as the left-out category, dummy variables for whether participants (comparison group members) enrolled (registered) in quarter two, three, or four

1. Quarterly earnings and AFDC grants in the immediate preprogram quarter should be included in the net impact model.
2. The net impact model should also include the UI recipient dummy variable, UI benefits earned in the immediate preprogram quarter, and the AFDC recipient dummy variable.
3. The unemployment rate should be defined according to the postprogram period for which the model is being estimated.

Thus, following these guidelines, one can estimate autoregressive models separately for adult men and women, and the estimated coefficient of the JTPA participant dummy variable represents the average net impact of JTPA on earnings and AFDC grants for the three postprogram periods of interest (i.e., three, six, and 12 months). For dependent variables expressed in dollar terms—earnings and AFDC grants—the coefficient of the JTPA participant dummy variable can be interpreted as the average *dollar* impact on a given outcome measure. Dividing the estimated dollar impact by the mean earnings or AFDC grants of comparison group members results in an estimate of the percentage change in earnings or AFDC grants due to JTPA.

For dummy dependent variables (i.e., whether employed in a particular period, or whether receiving AFDC grants during a particular period), the autoregressive net impact model is equivalent to a linear probability model. The model essentially estimates the effects of various factors on the probability of a certain event occurring, for example, having positive earnings in a given postprogram period. As such, the estimated coefficient of the JTPA participant dummy variable can be interpreted as the average *percentage point change* in the probability of working or receiving AFDC grants due to JTPA. By dividing the estimated percentage point change by the mean proportion of comparison group members, one can obtain an estimate of the percentage change in the probability of working (or receiving AFDC) due to JTPA.

Obtaining Net Impact Estimates for Various Subgroups

The models described above focus on providing overall estimates of the net impacts of JTPA for adult men and women. Determining whether JTPA effectiveness varies by the type of program activity and by

personal characteristics has important policy and planning implications.³⁴ Because JTPA program activities and participant characteristics can change considerably over time, knowledge of how program net impacts vary among them would help interpret time trends in JTPA's impacts. Furthermore, information on which program activities and services work best for given types of participants can provide valuable information for targeting future employment and training programs. Although the approach to estimating net impacts for different groups is formally identical, whether the group refers to the type of program activity or to individual characteristics, additional selection bias is likely to arise. In the next section we describe how to modify the autoregressive earnings and AFDC grants models to estimate the net impacts of JTPA for various groups, and review the additional biases that one must be aware of in interpreting the results.

Net impacts by participant characteristics. In general, specific group effects are estimated by including in the regression equation an interaction term that represents the product of the dummy variable for JTPA participation with the variable for the group of interest. Suppose one is interested in testing whether the net impact varies by a characteristic that is represented by the three dummy variables Z_1 , Z_2 , and Z_3 . One might think of the three variables as representing race or ethnicity categories (white, black, other).³⁵ Then, the only modification required to the autoregressive model described above involves replacing the JTPA participation dummy variable with three variables that each involves the JTPA dummy variable multiplied by one of the three variables representing the particular group (i.e., $JTPA \times Z_1$, $JTPA \times Z_2$, $JTPA \times Z_3$). The coefficients of these three variables are estimates of the net impact for the three groups of interest.³⁶

In attempting to disaggregate JTPA net impacts across groups, it is important that the group characteristics also be included in the model as control variables to account for differences in the general level of earnings (or AFDC grants) across these groups. In our illustration, the three Z_i variables must be in the model separately so that the estimated net impact coefficients only capture outcome differences due to JTPA across these groups and do not also capture the average differences in

outcomes due to the Z_1 's themselves. In addition, it is also important that the groups be mutually exclusive and exhaustive.

For example, suppose the Z_1 's refer to various participant age categories: 22-34, 35-54, and 55-64. Then two types of problems can arise in estimating the net impacts for these age groups:

1. Recoding errors can occur in creating the variables (e.g., ranges of 22-44, 35-54, 55-64) that result in overlapping the age ranges so that individuals age 35-44 would appear in both of the first two groups (i.e., the groups are not mutually exclusive).
2. Individuals in the sample may not fall into any of the three age categories created (i.e., the groups are not exhaustive).

This could occur if some participants were younger than 22, older than 64, or if there were a gap in the age ranges used. If the groups are not exhaustive, then all of the participant observations that do not fall into one of the categories would be treated as comparison group members, which would result in biased estimates of the net impacts of JTPA for the other groups.

At a minimum, we recommend that states examine differential impacts by race, education, age, UI claimant status, AFDC recipient status, and preprogram earnings for those individuals who had preprogram earnings. Because individuals' preprogram characteristics cannot be affected by JTPA, no additional selectivity bias is introduced in disaggregating JTPA net impacts by demographic groups.³⁷ However, this is not likely to be the case when examining whether JTPA effectiveness varies by program activity.

Net impacts by program activity or service. In principle, to probe beneath the average net impacts of JTPA and provide information on the program activities that contributed to the average effects, one would perform an identical interaction analysis to the one described previously, using Z_1 , Z_2 , Z_3 and Z_4 to represent classroom training, on-the-job training, job search assistance, and other program assistance respectively. Then, if c_1 represents the estimated coefficient for the interaction term between the JTPA dummy and Z_1 , then c_1 is the estimate of the average net impact for CT, c_2 would represent the estimated net impact for OJT, c_3 would represent the estimated net impact for JSA, and c_4

would represent the estimated net impact for other JTPA activities. There is, however, potential selection bias that can threaten the internal validity of the by-program activity net impact analysis.

Such bias relates to the nonrandom assignment of JTPA participants to program activities. As described above, the assignment of program activity is likely to be based on the agency's perception of an individual's needs and abilities. To the extent that this assignment process is based solely on the measured characteristics of participants, such as age, race, sex, education, and preprogram earnings, this will not bias the net impacts by program activity, as these characteristics will be included in the net impact model. But if the assignment of program activities is based on unmeasured characteristics, such as motivation and ability, and those unmeasured characteristics also affect earnings, then selection bias results. Thus, one must be very careful in interpreting net impacts by program activity.

Obtaining Net Impact Estimates by Program Length of Stay

To investigate whether the net impacts of JTPA vary by length of stay in the program, one would estimate an autoregressive model like those described earlier, with the overall program participation dummy variable replaced by a JTPA variable that measures length of program stay in terms of total weeks or, more appropriately, total hours. If the effects of length of stay on the outcomes are approximately linear, a convenient specification involves a model with a JTPA participation dummy and the participation dummy interacted (i.e., multiplied) with total weeks (total hours) in the program minus average number of weeks (total hours) in the program. In this specification, the coefficient of the JTPA dummy represents the estimated impact of JTPA at the average length of stay (average total hours), and the coefficient of the interaction term is an estimate of the dollar impact of an additional week (hour) of program participation.

Although the autoregressive earnings model controls for differences in measured characteristics between short- and long-term participants, it is likely that some differences in unmeasured characteristics remain. Individuals who leave the program early may be less motivated or,

alternatively, may have found employment on their own. On the other hand, individuals who stay in the program a long time may do so because they have fewer employment opportunities. Length of stay is also likely to depend on the type of program activity and SDA characteristics. Because of these additional selection bias problems, caution is needed in asserting a causal relationship between services and program impacts by length of stay.

Adjustments for Potential Data and Design Deficiencies

In addition to the problem of potential selection bias, there are some deficiencies in the UI earnings, JTPA, and ES data that may affect results. UI Wage Records are incomplete. They do not reflect earnings from jobs that are not in covered employment, or earnings from jobs located across the border in other states. The JTPA and ES data are deficient because there is inadequate information on whether ES registrants participated in JTPA, which may result in a contaminated comparison group. In this section, we briefly discuss the likely extent to which the basic impact estimates will be affected by these data and design deficiencies and the types of adjustments that may be necessary.

In the earlier data processing discussion, procedures were outlined that could reliably exclude those individuals from the comparison group who were currently participating in JTPA, who had participated in JTPA in the previous year, or who participated during the postprogram periods being examined. If it is not possible to implement these procedures, the comparison group will be contaminated to a certain extent. Such contamination would lead to an underestimate of the net impacts of JTPA, since it would dilute the treatment, as some comparison group members would have also received JTPA activities and services.

Although the ES is one source of applicants for the JTPA program, and one might expect that contamination could be high, existing data indicate otherwise. For example, based on data for the State of Washington for PY 1985, only 0.1 percent of all ES registrants active during the year were recorded as having enrolled in JTPA programs. Only 0.3 percent of those who were economically disadvantaged enrolled in JTPA. Although the figures are somewhat higher for enrollment in any training activity (e.g.,

JTPA, Job Corps, WIN, other)—1.0 percent for all applicants and 3.1 percent for those economically disadvantaged—even these participation rates are small enough to be safely ignored.

In states that have higher probabilities of economically disadvantaged ES registrants enrolling in JTPA, and in which it is impossible to exclude those who participate in JTPA from the comparison group before conducting the net impact analysis, it may be necessary to make some aggregate adjustment to the net impact estimates. Specifically, if p (q) is an estimate of the proportion of the adult men (women) in the comparison group participating in JTPA during the period of enrollment through the postprogram period (i.e., from one to two years), the estimated average program net impacts for adult men (women) should be multiplied by $1/(1-p)$ (or $1/(1-q)$) to adjust for this problem.

The second major data deficiency is that UI Wage Records do not include jobs in uncovered employment, or earnings from jobs in other states. However, the omission of earnings due to these problems biases the estimated impact of JTPA only if program participation causally affects the probability of working in uncovered employment or the likelihood of working in another state. Given the focus of JTPA on employment in the private sector, this should be less of a problem for the state-level net impact model. Also, in order to create a meaningful adjustment, one would need information on interview-reported earnings and UI earnings for both groups in the postprogram period, which will not generally be available. Thus, the best one can do is acknowledge the potential problem and indicate that the net impact estimates are based on the reasonable assumption that JTPA does not affect the probability of working in uncovered employment or working across the border in other states.

Cost Analysis and Benefit-Cost Comparisons

The estimated net impacts of JTPA program activities on participants' postprogram labor market experiences can be used to estimate the benefits of the JTPA for program participants and, under certain assumptions, the benefits of the JTPA to society as a whole. To determine whether the JTPA is an effective use of public resources, however, one

must assess the costs of providing JTPA activities and compare the costs to the benefits of the program. It is a serious conceptual error to assess a social program on the basis of overall costs or benefits alone. Moreover, to make informed decisions about the design of the program, policymakers must know both the costs and benefits of specific program activities. That is, program activities that yield relatively small benefits may yet be very effective when compared to the costs involved, since what matters is the social rate of return on the dollars invested in each participant, just as it is the rate of return on capital that matters for any private sector investment.

When all costs and benefits are accounted for, a benefit-cost analysis judges the *social* efficiency of a program. It determines whether the value of the goods and services available to society and by extension, to the members of society are greater as a result of the program. To make this determination, the benefits are typically assigned a monetary value, and their present value is compared to the present value of the monetized program costs. Assuming that all present and future benefits and costs are identified, appropriate monetary values are assigned and an appropriate interest rate is used to discount future benefits and costs to their present values. JTPA could be considered a worthwhile use of public resources if (1) the present value of the benefits is larger than the present value of the costs, or (2) the rate of return, that discounts the sum of costs and benefits to zero, exceeds the socially specified rate of return.

Although the process of conducting a benefit-cost analysis is straightforward, there is a variety of issues that limit the validity of such an analysis. Given data limitations and other issues, it will not generally be possible for states to conduct a comprehensive benefit-cost analysis to provide a definitive estimate of JTPA's social return on the investment. Nevertheless, the general approach described below is useful in organizing information on benefits and costs, and enables states to obtain some sense of the effectiveness of the JTPA and the conditions under which JTPA can be regarded an efficient use of public resources.

The discussion begins with a brief description of a benefit-cost framework for analyzing the effectiveness of the JTPA. Then we briefly describe the use of the net impact estimates to measure some of the

important social benefits due to the JTPA. We subsequently describe how to conduct a cost analysis to estimate the marginal cost of serving additional JTPA participants and the marginal costs of different program activities. A discussion of the issue of discounting future benefits and costs so that comparisons can be made in present value terms follows. We conclude by considering a few additional comparisons that should be made to determine how sensitive the overall conclusions are to certain assumptions.

Benefit-Cost Framework

The benefit-cost framework presented in table 2.10 lists the major benefits and costs that would ideally be accounted for in conducting a comprehensive benefit-cost analysis. As an aid for keeping track of the different benefits and costs, they are presented from three perspectives: the participant, the taxpayer, and society as a whole. The first class of benefits and costs consists of those benefits received by, or costs borne by, program participants. The participant perspective is important because it sheds light on an individual's incentives and willingness to participate in the program without coercion. The taxpayer perspective, sometimes referred to as the nonparticipant perspective, is important because it reflects the effects of the program on the government budget and the willingness of taxpayers to support the program.

The most inclusive set of program benefits and costs are those accruing to society as a whole. These are simply the sum of benefits and costs received, or borne by participants and taxpayers (that is, all members of society), taken separately. These represent a full accounting of all costs and benefits involved in operating the program. It is important to note that the social perspective ignores transfer payments between segments of society, that is, between participants and taxpayers, and examines instead whether the program results in a net increase in the resources available to society.³⁸ This is the appropriate perspective for a governmental body to take in examining the overall effectiveness of the JTPA.

Table 2.10
A General Taxonomy of the Benefits and Costs
of JTPA from Different Perspectives

	Participant	Perspective Taxpayer	Social
Benefits			
• Increased output			
- Postprogram output	+	0	+
- Program output	0	+	+
• Reduced receipt of income transfers			
- Reduced welfare payments, regardless of whether still on welfare	-	+	0
- Increased tax payments	-	+	0
• Reduced use of alternative social programs	0	+	+
• Nonmonetary benefits			
- Reduced crime	-	+	+
- Improved work attitudes of participants	+	0	+
- Improved mental and physical health	+	+	+
Costs			
• Program operating costs (e.g., rent, staff wages, and fringes, materials and supplies, and overhead administrative costs)	0	-	-
• Participant opportunity costs (e.g., forgone earnings or home production)	-	0	-
• Transfers to participants (e.g., stipends)	+	-	0
• Costs of participation (e.g., transportation, child care, extra clothing, and food)	+	0	+
• Psychic costs (e.g., stress of studying and being tested, separation from children)	-	0	-

Program Benefits

The major benefit of the JTPA from the social perspective is the increase in output produced by participants. Conceptually, two types of gain should be distinguished: (1) the increase in postprogram output, measured by the increase in earnings of the participants, and (2) the increase in output produced while an individual participates in the program. For the most part, the current-program (as opposed to pre- or postprogram) output due to the JTPA is likely to be small, particularly for participants in classroom training and job search assistance-program activities. Only for OJT programs is the value of program output likely to be positive, and even for these programs it is difficult to assign appropriate monetary values. Because of the difficulties involved in measuring the value of program output, as well as in measuring the value of other nonmonetary benefits, such as reduced crime or improved mental or physical health of participants and their families, we recommend that states do not attempt to directly measure these benefits, but recognize their potential importance when discussing the overall results from the benefit-cost analysis. The primary benefit to be measured, therefore, is the increased postprogram output due to JTPA.

The participant-comparison group differences in earnings in the postprogram period are used to measure the increase in output of goods and services available to society during that period due to JTPA. This is a reasonable procedure provided that JTPA participants do not find jobs in the postprogram period at the expense of other disadvantaged persons.³⁹ It is beyond the scope of the state-level model to assess the extent of such job displacement. As a result, the benefit-cost analysis is limited to determining whether the social benefits from receiving JTPA activities are greater than the costs to society of providing those activities, as measured by the change in total postprogram earnings due to JTPA activities.

Two issues arise in translating participant earnings gains into a measure of the increase in output of goods and services available to society. First, one must determine how to extrapolate the postprogram gains observed for the periods from three months to one year following termination, into subsequent periods. For example, if the three-month,

six-month, and 12-month net earnings impacts imply similar gains per quarter, then it may be reasonable to assume that the gains persist over time. However, based on previous studies, it is likely that the gains decline over time, and information through just the first postprogram year may not be sufficient to estimate reliable time trends for the purpose of extrapolating future gains. This emphasizes the importance of augmenting the basic design with the collection of additional postprogram data to more precisely estimate the long-term gains from the JTPA. If this is not possible, at a minimum the benefit-cost analysis should indicate whether earnings gains observed during the one-year follow-up period are sufficient to make the program worthwhile, and if not, indicate whether JTPA would be viewed as a worthwhile social investment if the gains persisted at their current level for up to five years.⁴⁰

A second issue concerns adjustments that should be made to earnings gains to account for fringe benefits. That is, if the increase in output is equal to the increase in compensation paid to those who participate in JTPA, then although this compensation is primarily in the form of monetary earnings, adjustments for nonmonetary earnings should also be made. Fringe benefits include pensions, health and other forms of insurance, and payments on behalf of the worker for unemployment insurance, workers' compensation, and FICA. As a rough approximation, insurance and pension benefits for workers served by JTPA are estimated to be approximately 10 percent of monetary benefits (Woodbury 1980), and payments to government programs are approximately 10 percent as well. Thus, we recommend that the net earnings gains be multiplied by 1.2 to adjust for fringe benefits in deriving a measure of the social benefits due to JTPA.

To summarize, the benefits to be measured and included in the benefit-cost analysis include only the increase in postprogram output due to JTPA. This is approximated by the increase in postprogram before-tax earnings, as measured by the estimated coefficient of the JTPA dummy variable in the net impact equation—adjusted for potential data and design deficiencies as described in the previous section—and subsequently adjusted for fringe benefits. Procedures will be developed to determine how the increase in earnings over the first year should be

extrapolated to yield estimates of increases in postprogram output in subsequent years. These steps will yield an estimated stream of future benefits for both adult men and women. Individual values in these benefit streams will then be weighted by the proportion of men and women served by the JTPA in the particular program year, to generate an estimate of the aggregate benefit stream due to the JTPA.

JTPA Costs

As indicated in table 2.10, there are several different cost components in a benefit-cost analysis. The major cost categories include (1) program operating costs, (2) participant opportunity costs, (3) transfers to participants, (4) costs associated with participating in the program activities, and (5) psychic costs to participants of participating in the JTPA.

The major costs from the social perspective are the program operating costs. These include direct operating costs, such as rent, salaries for instructors, and costs of materials and supplies, and indirect or overhead costs, such as those involved in managing and administering the program overall. Because no fee is charged for program participation to those eligible, the operating costs are not considered as costs from the perspective of program participants. However, operating and administrative costs do involve the use of resources that have alternative uses. They represent real costs from the perspective of the taxpayer and society as a whole. Thus, in table 2.10, such costs are represented as a zero to program participants and as a minus in the other two columns.

A second important component of cost concerns the earnings opportunities and home production that participants forgo while participating in the program. These forgone earnings and home production are clearly costs to participants and, to the extent that less output is produced because workers were participating in the JTPA, the forgone output (as measured by forgone earnings or the value of forgone home production) is a cost to society as well. Although previous studies have recognized forgone earnings and lost home production to be important elements of program cost, because of data limitations these components are almost always excluded from the final benefit-cost comparisons. We also recommend that this cost component be formally omitted from the

benefit-cost analysis. Although the social program costs are understated by this exclusion, given the relatively brief length of participation in JTPA and the questionable nature of the assumption that considerable output was forgone when previously nonworking individuals participated in the program, we believe this is justified. In interpreting the results of the benefit-cost comparisons, however, one should indicate to what extent the overall assessment is likely to be sensitive to omitting the social cost of forgone earnings and lost home production.⁴¹

Other potentially important costs of JTPA from the taxpayers' perspective are transfers to participants in the form of the money value of classroom materials, stipends, transportation, child care, and food or clothing allowances. Although such costs are much less important under JTPA than under CETA, they could be considerable in some cases. Note, however, that such transfers are a cost from the taxpayers' perspective, and a benefit to the participants that receive them. As such, transfers do not affect the cost-benefit analysis from the social perspective because the loss to the taxpayer is cancelled by the gain to the participant.

Other potentially important costs are the direct costs participants incur in participating in JTPA activities as well as any psychic costs. These psychic costs are inherently unmeasurable, and are included in the conceptual framework only for the purpose of completeness. The costs incurred by participants in attending classes or participating in job search activities require data that must be obtained by interview from individual participants. Because of the large expense involved in acquiring such information, and given the fact that these costs are likely to be a small share of the total cost of the program to the individual and to society as a whole, as a practical matter these costs are omitted from the final benefit-cost comparisons.

To summarize, the costs of JTPA to be measured and included in the benefit-cost analysis will be limited to those involved in operating the program, that is, the sum of rent, staff, materials and supplies, and administrative costs. In interpreting the benefit-cost analysis comparisons, however, it is important to recognize that many of the social costs of JTPA have not been measured, and that these unmeasured costs could affect the overall assessment of whether or not JTPA is an efficient use of social resources.

Estimating Program Costs

The benefits from JTPA will be expressed in terms of incremental dollar gains per individual adult participant. Therefore, the cost analysis must similarly estimate the incremental (i.e., marginal) outlay in dollar terms per individual adult participant. Many obstacles exist in deriving reliable estimates of the marginal costs of serving JTPA participants: problems of data omission, inconsistent aggregation, difficulties in allocating input costs among joint outputs, and ambiguity involved in imputing prices of existing agency or SDA resource inputs. However, statistical methods that overcome several of these problems and provide useful information on the marginal costs of employment and training programs are available.

The primary data source for the cost analysis is the *JTPA Annual Status Report (JASR)*. The JASR provides for each SDA the characteristics of program trainees and information on program outcomes and costs for Title II-A and Title III programs funded under the JTPA⁴². Fortunately, these data are provided separately for adult and youth participants in Title II-A programs.

The JASR data have several advantages. In addition to being in a standardized format with unambiguous definitions of all information items, the JASR contains data on total federal expenditures (but not total social costs) in operating the JTPA, as well as some information on the socioeconomic characteristics of adult program trainees that can be used to standardize the cost analysis. The more important variables are the number of trainees by sex, age, education, race or ethnicity group, welfare-recipient status, limited English language proficiency, and handicapped status, and the average number of weeks participated.

These participant characteristics can be thought of as inputs that enter the employment and training production process and have obvious instructional and resource implications that affect costs. For example, those participants with limited English language proficiency will probably require more program resources to complete training successfully. As such, these characteristics can be used to standardize the relationship between total costs and participants served to obtain estimates of marginal costs as described below.

The JASR contains two major limitations. The JASR does not contain (1) data on the number of terminees by type of program activity, or (2) information on administrative and other costs incurred at the state level in operating JTPA programs. Without information on the number of individuals served by type of program activity, one cannot identify the marginal costs of specific JTPA activities. This precludes comparing the marginal benefits and marginal costs of different program activities and services. As a result, one cannot identify those activities and services that are relatively most effective. One can only evaluate JTPA as a whole. In addition, without information on the costs incurred at the state level, the marginal costs of serving JTPA participants as derived from JASR data are understated.

To overcome these problems, we offer two recommendations. First, every effort should be made to obtain data on the number of adult terminees that participated in various program activities and services during the program year. At a minimum, it is useful to have data on the following: CT-only terminees, OJT-only terminees, JSA-only terminees, terminees who only participated in some other activity, and terminees who participated in multiple activities. This information must be obtained from each SDA for the same period in which the terminnee characteristics and program costs on the JASR are reported. One could implement the following steps to obtain the necessary information for PY 1989:

1. Create a working file (on tape or disk) of all persons who terminated from JTPA Title II-A programs in any SDA in the state during PY 1989. The file should include the person's age, data on all program activities and services participated in, and an SDA identifier.
2. Exclude from the file all persons who are 21 or younger.
3. Create variables that represent each type of activity of interest and that may have different cost structures. For example, as indicated above, it is important to differentiate the costs by type of activity as well as costs for those who participate in only one activity vs. multiple activities. This can be accomplished by creating five variables, the first four of which would simply be dummy variables indicating whether the only activity the person participated in was

CT, OJT, JSA, or other, and a fifth variable indicating whether the person participated in any combination of these activities.

4. Create a separate subfile of adult terminées for each SDA.
5. Create counts of the number of individuals in each SDA in each of the five program-treatment types, and merge these counts with the JASR data for each SDA.

Our second recommendation concerns how to handle costs incurred at the state level in the operation of JTPA programs. Conceptually, the actual or imputed JTPA expense incurred at the state level should be added to annual program-year SDA total costs to obtain a better estimate of the overall social costs of JTPA. Provided information is available on the total costs contributed by the state to the operation of JTPA, one can allocate these to the various SDAs. For example, one method is to assume that the overhead costs incurred at the state level in support of various SDAs are proportional to the number of adult terminées in each SDA. Thus, to allocate state-level costs in operating JTPA programs to the different SDAs, one could multiply total state costs by the ratio of the number of adult terminées in a given SDA to the total number of adult terminées in all SDAs. Such a procedure would, in part, overcome the limitation of the JASR data described above. If it is not possible to obtain estimates of costs contributed at the state level to the operation of local JTPA programs, this limitation has to be recognized in interpreting the results of the benefit-cost analysis.

With the basic data set and adjustments described above, one can estimate a program cost function that provides information on the marginal cost of serving JTPA participants. Using ordinary least squares regression techniques, one could estimate a regression equation with total federal expenditures plus allocated costs incurred at the state level (if possible) for the SDA as the dependent variable, expressed as a function of the following variables:

1. Number of adult men terminées.
2. Number of adult women terminées.
3. Number of adult terminées who are:
 - black
 - Hispanic

other nonwhite
 students—high school or less
 high school graduates
 aged 22-54
 welfare recipients
 single household heads with dependent children
 UI claimants
 limited English language proficiency
 handicapped

4. Average number of weeks of participation.

With these independent variables in the regression equation, the coefficient of the variable “number of adult men terminees” would represent the marginal cost of serving additional male adult participants in JTPA on average, and the coefficients of the other variables in the model would capture the extent to which the marginal cost varied for serving persons with specific characteristics.

If the procedures outlined above are followed so that data on the numbers of terminees by program activity are obtained for each SDA, then one would estimate a second regression equation like the one above except that the “number-of-adult-men-terminees” and the “number-of-adult-women-terminees” variables would be replaced by the following four variables: the number of CT-only terminees, the number of OJT-only terminees, the number of JSA-only terminees, and the number of terminees that participate in multiple activities. In this formulation, the coefficients of these four variables would represent estimates of the marginal cost for each of the different types of program activities. If there are enough observations, the numbers treated can again be separated by gender, forming eight categories so that costs by activity can be estimated as a function of gender.

These estimates of the marginal costs of serving adult JTPA participants (either overall or by program activity) are then compared to the marginal benefits from the program in terms of increased postprogram output (either overall or by program activity) to state whether JTPA is an effective use of public resources.

One additional potential limitation of the cost analysis should be noted. This concerns the issue of sample size. Since the analysis is based

on SDA-level data, the number of observations available in an annual cross-sectional analysis equals the number of SDAs in the state. In relatively small states with few SDAs, working with annual data may yield insufficient observations to estimate a cost model like the one given above and inhibit ability to obtain reliable estimates of program marginal expenses. A solution is to pool quarterly data over a few years and include dummy variables for different quarters to account for seasonal cost differences and other lumpy costs.

Benefit-Cost Comparisons

Three data elements are required to conduct a cost-benefit analysis: estimates of the benefit stream over time, estimates of program costs over time, and the interest rate used to discount future benefits and costs into present dollars. In this section we indicate how to discount the future benefit stream so that the present value of benefits can be compared to the current program costs, and indicate the criteria to be used to measure the net effectiveness of JTPA as a social investment. The discussion concludes with examples of comparisons that should be made to determine how sensitive the overall conclusions are to alternative assumptions.

Because the benefits of an employment and training program occur over time, one must translate this stream into a common reference period. Conventionally, this involves discounting future dollars into their present value, using an interest rate that approximates the alternative costs of the funds invested. The two interest rates that have been used most often in such processes are the rate of return on investment in the private sector—historically averaging around 10 percent before taxes—or the long-term rate of growth of the economy—historically, around 3 percent. We believe that the lower rate is preferred for evaluating an investment in human capital from the point of view of society as a whole. Because there is much disagreement about which is the more appropriate interest rate to use, however, we also recommend that states examine how sensitive the main results are to using a higher figure such as 10 percent.

Using a 3 percent discount rate, one would measure the net effectiveness of JTPA by calculating a benefit-cost ratio, where the numerator is the present value of the incremental benefits due to the program (i.e., \sum

$B(t)/(1.03)^t$), the denominator is the present value of the costs incurred (which requires no discounting, since all costs are incurred in the current period), t refers to the postprogram years in which benefits are realized, and n is the last year in which benefits are realized. Within the numerous limitations described earlier, the JTPA would be regarded as an efficient use of public resources whenever the benefit-cost ratio exceeded 1.0.

In addition to obtaining the main benefit-cost results described above, we believe it is important that benefit-cost ratios be calculated to demonstrate how sensitive the conclusions are to alternative assumptions. In particular, alternative ratios should be calculated for the following:

1. A 10 percent discount rate.
2. Benefit estimates that do not include adjustments for selection bias or for potential contamination. Since each set of estimates rests on a different set of inherently untestable assumptions, it is important to know how sensitive the overall conclusions are to the size of these adjustments.
3. A range of program benefits and costs that reflects the fact that the main estimates are subject to statistical imprecision. For example, one could construct an upper and lower bound of a 90 percent confidence interval for the net impact of JTPA on postprogram earnings by adding and subtracting 1.96 multiplied by the standard error of the JTPA dummy variable to the estimated JTPA coefficient. By adjusting both the upper and lower bounds for the fringe benefits, one would then obtain an estimate of the upper and lower bounds for the increase in postprogram output due to JTPA. Upper and lower bounds for the marginal cost of JTPA can also be obtained by creating a 90 percent confidence interval around the appropriate regression coefficient (i.e., adding and subtracting 1.96 multiplied by the standard error of the estimated coefficient of the number of adult men/women terminees in the cost equation). Then, by choosing different combinations (e.g., upper bound for benefits and lower bound for costs, lower bound for benefits and upper bound for costs), one can provide useful information on how sensitive the benefit-cost ratios are to alternative assumptions.

These sensitivity analyses, in combination with the main benefit-cost results, should provide useful information on the effectiveness of JTPA programs.

Relationship Among Evaluation Models

Although much could be learned from implementing only the net impact model, considerable complementary information can be obtained by implementing the process analysis and gross impact models described in the companion chapters. In this concluding section, we briefly indicate how the net impact model relates to the other analytic approaches developed to assist states and SDAs in understanding the operations and impacts of their JTPA projects.

The net impact and gross impact analyses are quite complementary. Although both models are designed to address program effectiveness questions, they differ in terms of the types of evaluation questions that can be answered, the range of outcome measures of interest, and the types of comparisons being made. For example, the net impact analysis is limited to adults only, and because it relies exclusively on administrative data sources, there are relatively few outcome measures to examine, and only a limited number of personal characteristics can be included in the analysis.

On the other hand, the gross impact analysis can include youths as well as adults, an expanded set of labor market outcomes, and additional personal characteristics. As such, the gross impact model can be used to address certain relative effectiveness questions for youths served by the JTPA and can possibly provide information on the mechanisms through which the JTPA affects adults' earnings and welfare dependency. Because of the additional outcomes available, a gross impact analysis may be able to provide some evidence on whether the earnings changes are due to changes in wage rates, changes in hours worked per week, or changes in weeks worked per year, although a comparison group is necessary to provide definitive evidence on these issues.

In addition to providing complementary information on different subgroups and outcome measures, information from the gross and net

impact models may shed light on important methodological issues affecting the validity of analyses of social-program impacts in general. For example, the gross impact model can use interview-reported earnings, whereas the net impact model relies on UI Wage Records to create measures of earnings. There are advantages and disadvantages to both approaches. By estimating gross impacts, using the net impact design and with the same samples of participants and comparison group members, one can provide evidence on the extent to which the impact results are sensitive to the use of the different data sources. In addition, because the gross impact model has an expanded set of independent variables available, by implementing both models using the same analysis samples it is possible to get some idea of whether the net impact estimates are sensitive to these omitted variables—an important statistical issue. Such comparisons provide important information on the limitations of the different analyses and indicate other independent variables or outcome measures that would be important for subsequent program analysis.

An SDA process analysis is a very important source of information for adding flesh and conceptual relevance to the net impact model. Because of the inherent limitations of the nonexperimental approach in estimating program net impacts, an SDA process analysis is a necessary first step to a valid net impact analysis. In particular, because the validity of the net impact results rests on the similarity of the participant and comparison groups selected, it is critical to understand the JTPA participation-selection process, the factors that govern the assignment of participants to program activities, and differences in the content and recording of program activities across SDAs. A process analysis offers the following information for the net impact model:

1. It will provide a detailed description of the criteria (explicit and implicit) used by SDAs and their subcontractors in screening JTPA applicants to choose individuals for program participation. As such, the process analysis will yield important insights into the type and extent of “creaming” that occurs and the likely differences that may exist between participants and comparison group members that are not possible to control for in the net impact model.
2. It will include a detailed description of the procedures followed in

assigning participants to program activities. It will reflect whether more-advantaged participants are assigned to specific program activities, whether all participants are first assigned to JSA, and whether only those who are not immediately placed are subsequently assigned to CT or OJT. This information helps in determining whether the estimated net impacts by program activity are likely to accurately reflect the relative effectiveness of different activities, or merely represent the fact that more-advantaged individuals are assigned to certain activities, while less-advantaged individuals are assigned to other activities.

3. It can identify major differences in the content of program activities across SDAs, as well as differences in the ways in which similar program activities are recorded in the JTPA MIS. This information is useful in developing meaningful, consistent measures of program activities across SDAs.
4. It will identify variables to include in the model. For example, it can identify SDA characteristics, such as service delivery strategies, which are quantifiable and differ across SDAs, so that they can be included in the model to test whether the net impact of JTPA significantly differs across these dimensions.

In addition to benefiting from the SDA process analysis, note that the net impact model may also produce information that would be of interest to a process analysis. For example, the net impact model may indicate that after adjusting for differences in participant characteristics and local labor market conditions, the net impact of JTPA is considerably different in some SDAs than in others. The process analysis can then examine in detail what it is about the specific SDAs that accounts for such differences. If measures of specific SDA attributes that are potentially responsible for the different net impacts can be developed, they can be included in subsequent net impact models to determine whether they account for the different net impact estimates across SDAs. Such ongoing interaction between the process and net impact analyses highlights the complementary nature of the two analytic approaches and should result in an improved understanding of the factors that affect program effectiveness.

Summary

We have described a general approach to examining the net impact of a social program and illustrated this approach with an in-depth description of how to estimate the net impact of employment and training programs for economically disadvantaged individuals funded under Title II-A of the Job Training Partnership Act. Although the details of the net impact evaluation model have been tailored to a specific social program, the key elements of the approach have broad applicability to other social programs. In particular, any net impact analysis approach must include a conceptual framework, a research design, and plans for data collection, processing, and analysis to answer questions posed in the conceptual framework. Moreover, because a major issue in any social program evaluation is the comparability of participant and comparison groups, it is likely that the concern over selection bias will always be present and the statistical methods described in this chapter are useful in dealing with this potential problem.

Although the results of a valid net impact analysis can provide very useful information on the extent to which participants overall are better off as a result of receiving program services (and, potentially, which participant subgroups benefit most), additional information is required to determine whether a social program is an efficient use of public resources. Specifically, one must use the results obtained from the net impact analysis and other analyses to develop measures of total program benefits, and compare the total benefits with the costs of the program. Although the costs involved in conducting such analyses as part of an ongoing program evaluation effort may be high in the initial stages, once the system is in place the costs should decline considerably and the benefits from the evaluation should be substantial.

NOTES

1. For example, one of the major reasons that the Job Corps education and vocational training program for disadvantaged youth has been able to avoid budget cuts during recent years, despite its extremely high cost per participant, is the availability of considerable research information indicating that corps members receive long-term economic and noneconomic benefits from the program.

2. One exception is the experimental evaluation of employment and training programs for adults and youths funded under the Job Training Partnership Act that was undertaken by the U.S. Department of Labor in 1986. This net impact study involves approximately 15 program operators nationwide, will cost approximately \$20 million, and will be completed in 1991 or 1992.

3. Given that the model is usable and provides valid results on important postprogram outcomes, such as earnings and welfare dependency, then an important by-product—*consistency* in application across states—is likely to occur. This will maximize the usefulness of the information obtained from any single analysis by extending all states' knowledge of what is known about the effectiveness of employment and training programs among different state environments.

4. The primary exceptions include Dickinson, Johnson, and West (1987a) who examined CETA net impacts on the components of earnings, including whether employed, the hourly wage rate, hours worked per week, and weeks worked per year, and Bassi et al. (1984), who also examined the impact of CETA programs on welfare dependency.

5. Information on short-term earnings impacts occurring within less than a year can only be provided through primary data collection efforts, or through the use of UI Wage Records.

6. Because the net impact model is based exclusively on available administrative data, the conceptual framework is in large part data-determined. However, even though the conceptual framework is constrained by the features of available state data bases, virtually all previous national studies of the net impacts of employment and training programs summarized in the evaluation guide share several of the limitations of the net impact model described here.

7. Although these administrative data sources are very inexpensive, particularly relative to the costs of survey data, nontrivial data processing costs must be incurred to access the appropriate records from the system. Depending on the size of the files in the state, these costs could range from several hundred dollars to several thousand dollars (or more) per run. Moreover, prior to obtaining these data, it will be necessary to meet any state requirements concerning data confidentiality, and to take steps to maintain data confidentiality (e.g., remove all identifying information and create unique identifier for analysis purposes).

8. Some states that are not formally wage-reporting states have comparable earnings records available that are maintained by their Departments of Revenue and could be used in the analysis if the necessary interagency agreements are made.

9. For example, in some states, UI Wage Records do not include earnings for the following types of employees: certain corporate officers, church employees, individuals paid exclusively on commission, domestics who earn less than a certain amount per quarter (e.g., \$1,000), railroad employees, employees of small agricultural firms, casual laborers, and certain barbers or cosmetologists. As a result of these coverage gaps, approximately 80 percent of all state wages are generally included in the UI Wage Records data base.

10. For example, states that are very interested in developing net impact estimates for JTPA Title II-A youth programs might consider implementing an experimental design. Alternatively, states might consider conducting (relatively expensive) interviews of participants and comparison group

members to collect the detailed preprogram and postprogram employment and schooling data necessary for reliable analysis. In either case, the general research design and analysis plans described later in this chapter could be followed. States interested in such approaches, however, should first consult employment and training researchers who are knowledgeable in experimental design issues and questionnaire development to avoid the pitfalls that have plagued previous studies.

11. If, however, work experience or some other program activity is used extensively in a particular state so that the sample sizes are sufficient to support precise net impact estimates, it would also be possible to follow the other procedures outlined later in this chapter to estimate the net impacts for this activity.

12. On the other hand, an enrollee-based sample frame has some disadvantages. To avoid excluding long-term participants from the analysis, an enrollee sample frame causes a delay in analysis findings relative to a terminatee-based sample. In addition, because a given group of enrollees may terminate across different quarters, with such a design it is more difficult to estimate earnings impacts that correspond to specific time periods after program termination, such as the three-month period following the quarter after termination. However, alternative sample frames suffer from other problems that are more severe, which led us to the decision to use an enrollee-based sample.

13. The limited amount of missing data is, in part, a reflection of the procedures used by many agencies to assign "default" values when data are missing. Such procedures lead to measurement error, which can also introduce analytical complications as discussed later.

14. It is desirable to examine the missing data problems before making a final decision on whether to exclude such cases from the analysis sample. If the data items are missing for random reasons, then no harm is done by omitting such cases from the participant group. If, however, it is determined that the reason the data are missing is systematically related to the impact of the program (e.g., individuals with missing data on length of stay dropped out of the program and were less likely to be placed), this would reduce the internal validity of the overall analysis. Thus, some simple comparisons of the characteristics of participants with missing data and participants with complete data on program services will be performed as described in subsequent sections before a final decision is made to exclude cases with missing data from the analysis sample.

15. For example, in some SDAs women comprised as little as 25 percent of adult JTPA terminatees in PY 84, while in other SDAs women were over 80 percent of all adult terminatees in Title II-A programs during this period.

16. According to the JTPA legislation, to be eligible for Title II-A programs, adults must be 22 years of age or older and be economically disadvantaged. The act should be consulted for the exact definition of "economically disadvantaged." To the extent possible, the comparison group should only include individuals who meet the explicit eligibility criteria and who are similar to participants on characteristics emphasized in the legislation.

17. Based on a comparison of ES and JTPA data collected in selected states, the following individual characteristics are generally comparably measured: age, race/ethnicity (white, black, Hispanic, American Indian/Alaskan Native, Asian/Pacific Islander), education (whether received high school degree or equivalent), handicapped status (whether has physical or mental impairment that is a substantial handicap to employment), occupation (primary DOT code of previous job), veteran status (a veteran, a Vietnam-era veteran, recently separated, a disabled veteran), Food Stamps recipient, WIN registrant, and economically disadvantaged status. In addition, preprogram measures of UI Wage Records, AFDC grants, and whether a UI recipient will also be available and comparably measured for both participants and comparison group members. Although this list is not as complete as one would ideally like—measures of marital status, family size, dependent children, ex-offender status, limited English-speaking ability, and detailed data on preprogram employment experiences

are not available—it must be recognized that most of these characteristics were unavailable to previous national studies of the impact of employment and training programs. As such, this is not a limitation that is specific to the model described here.

18. Note that decreased emphasis on the economically disadvantaged measure might introduce additional measurement error into this variable. Not only did ES staff previously have no real incentive to accurately record the status of the applicant (i.e., since ES services do not depend on whether a person is economically disadvantaged), but they now have even less incentive to do so. As a result, it is likely that ES offices under-report serving such applicants. Thus, to the extent that only ES registrants who are recorded as economically disadvantaged are included in the sample frame for the comparison groups, their status should be measured reasonably accurately, which will minimize complications due to measurement error.

19. Several studies exclude from the comparison group individuals with very high preprogram earnings who were clearly ineligible to participate in employment and training programs (e.g., Dickinson, Johnson, and West (1986); Westat (1984)). By matching participants and comparison group members on economically disadvantaged status, however, such additional exclusions should no longer be necessary. Note that if the economically disadvantaged status variable is not available in some states for ES registrants, then procedures to exclude cases with high preprogram earnings must be implemented as described later in this section.

20. The statistical power of any hypothesis test relates to the likelihood of drawing a particular type of incorrect conclusion. The power of a test concerns what is called a Type II error, or incorrectly accepting the null hypothesis (e.g., that there are no significant differences in earnings between program participants and the comparison group) when the null hypothesis is false. Alternatively, the statistical power is the probability of detecting an effect (at the chosen significance level) when the effect of the specified size, in fact, exists (i.e., it is 1 minus the probability of making a Type II error). Because the probability of making a Type II error declines as sample size increases, larger samples are used to minimize Type II errors and maximize the power of the test.

21. With such a sample design we estimate that it will be possible to detect approximately a five (six) percentage point impact on earnings for adult men (adult women) with 90 percent power at a 0.10 significance level. That is, one would have 90 percent power at a 0.10 significance level of detecting an overall net increase in participants' earnings of as small as five or six percentage points.

22. In using enrollments from the prior program year to set the SSA number range in step 4 above, it may be necessary to adjust estimated enrollments to reflect changes in real program resources, that is, changes in federal allocations adjusted for inflation. Such adjustments can be made using information on the percentage change in program expenditures typically incurred for a given percentage change in the number of JTPA participants, which can be obtained from the cost analysis described later.

23. Note that because eligibility for ES services does not depend on economically disadvantaged status, it is likely that this indicator is measured with much more error for comparison group members than for JTPA participants. However, we expect that the error is more likely to be in not identifying some registrants as disadvantaged who in fact are. Thus, by only retaining in the comparison group those ES registrants who are indicated to be economically disadvantaged, the groups should be reasonably comparable on this dimension.

24. Although this would help to ensure similarity in terms of maximum earnings in the preprogram period, in the absence of data on economically disadvantaged status it is also desirable to match the samples more closely in terms of the time pattern and levels of preprogram earnings. For example, based on the preprogram pattern of participants' earnings, one could create specific cells that are

mutually exclusive and exhaustive of all possibilities, and then select comparison group members from these cells to match the distribution of participants.

25. Using quarterly values of the BLS *Consumer Price Index* for all Urban Wage and Salary Earners, one would deflate (divide) the values of the variables expressed in nominal dollar terms by the value of the price index in the same calendar quarter, and create measures of real earnings and real AFDC grants received in each preprogram quarter and real UI benefits received in the immediate preprogram quarter.

26. For data items obtained from ES or JTPA application forms, the range of acceptable values can be specified exactly. That is, if handicapped status is coded as 1 for yes and 2 for no, then any values other than 1 or 2 are clearly errors that likely occurred in entering the data into the MIS. Unless such errors can be readily corrected using other information on the file, they should be set to a common missing data code (e.g., -9) and dealt with as part of the procedures for handling missing data.

27. For example, one can use mean values, a hot-deck or cold-deck procedure, a regression equation, or other more complex methods to deal with missing data problems. In general, as long as the reason a variable is missing is not correlated with the variables representing program participation (e.g., classroom training, length of program participation), no bias is introduced in the estimate of net program impacts, although the standard error of the variable that has been imputed is reduced and the precision of the estimated impacts is overstated.

28. It would also be possible to estimate an OLS linear probability model of the likelihood of participating in JTPA to determine the major differences between the two groups. That is, one would estimate a regression equation with the dependent variable equal to 1 for JTPA participants and 0 for comparison group members, and the independent variables would be all measured characteristics included in the net impact model described later in this section. This approach has the advantage of estimating the independent influence of each measured characteristic, while controlling for the influence of all other characteristics, which eliminates the confounding effects of other variables that may be present when comparing mean characteristics. That is, a comparison of mean characteristics could indicate, for example, that JTPA participants are more likely to be minorities and less educated, whereas the regression approach would account for the differences in education by race and could reveal that, after adjusting for differences in race, there are no differences between participants and comparison group members in terms of education levels.

29. As indicated in the second and fourth guidelines for selecting independent variables discussed above, the only changes necessary in the independent variables in changing the dependent variable from the immediate preprogram quarter to the immediate preprogram year would be to ensure that quarterly earnings and AFDC grants in the second, third, and fourth preprogram quarters were excluded and that the unemployment rate was defined for the entire preprogram year rather than just for the immediate preprogram quarter.

30. Most standard regression programs allow one to perform an *F-test* of such an hypothesis, and also provide the calculated *F-statistic* for the test. Under the assumption that the error terms are normally distributed, the test statistic follows Snedecor's *F*-distribution with r degrees of freedom in the numerator and $N-K$ degrees of freedom in the denominator, where r is the number of restrictions being tested (i.e., the number of independent variables that have been multiplied by the JTPA dummy), and $N-K$ is the number of degrees of freedom when no restrictions are imposed (i.e., total sample size less the number of variables in the equation). If the test statistic exceeded the critical value for the specified level of significance, then the null hypothesis would be rejected and we would conclude that the preprogram earnings (or AFDC grants) equations for the two groups are not similar.

31. Note that individuals who are still in the program in a given quarter should not be included when analyzing the impact of JTPA on earnings or AFDC grants during that quarter because their earnings

may be unusually low (e.g., for classroom training or job search assistance participants) or unusually high (e.g., for OJT participants), which would bias the estimated overall net impacts.

32. Previous CETA studies indicate that program net impact estimates could be somewhat sensitive to the exclusion of long-term participants, primarily because the excluded individuals tended to be in Public Service Employment programs and were always working (Dickinson, Johnson, and West 1987b). Although this is generally not likely to be the case in JTPA, this suggests that, at a minimum, states should try to obtain longer follow-up data to use in additional analysis that includes all long-term participants to examine how sensitive the results are to this issue.

33. Because of the different sample sizes involved in analyzing impacts for different postprogram periods, the precision of the estimated 12-month net impacts will be less than the precision of the estimated impacts over a three-month period.

34. It may also be of interest to determine how the effectiveness of the JTPA differs among SDAs. This can be determined through including separate SDA-participant interaction terms using the general approach described below.

35. It should be noted that, in principle, similar analyses could be performed to determine whether net impacts vary across local labor market conditions. However, because the labor market variables would take on the same value for all persons in the same local area in a given time period, there is not likely to be sufficient variation to obtain precise estimates of how program impacts vary across local labor market conditions, except in large states, with many SDAs, and where there are considerable differences in labor market conditions across SDAs.

36. To formally test whether the program net impacts differ significantly across the groups of interest, an *F-test* is used. In this case, the test statistic follows an $F(r, N-K)$ distribution, where r is the number of restrictions imposed by the basic model (equal to the number of groups minus one), and $N-K$ is the number of degrees of freedom in the basic impact model. The hypothesis that the net impacts do not vary across the groups of interest (e.g., across racial groups) would be rejected for $r = 2$ and sufficiently large sample sizes at the 0.05 (0.01) significance level if the test statistic exceeded 2.99 (4.60). Most standard software analysis packages calculate this *F-statistic* as part of the analysis run.

37. If the characteristics defining the subgroups of interest are not measured equally well for the participants and comparison group members, however, the subgroup impacts will inappropriately reflect these differences. Because the presence of measurement error in an independent variable biases its estimated coefficient downward, if the amount of measurement error on a subgroup characteristic were greater in the JTPA sample, for example, than in the comparison group, the effect of that characteristic on the outcome variable would be smaller in the JTPA sample than in the ES registrant sample. The interaction term would inappropriately pick up such a difference and misleadingly indicate that JTPA impacts were smaller for individuals with that characteristic.

38. Reductions in transfer payments (e.g., AFDC grants) do not represent a benefit from the social perspective, since the increased benefit to taxpayers is offset by the loss of income to recipients, and there is therefore no change in the resources available to society as a whole. Thus, including estimated benefits from reductions in welfare dependency due to JTPA would involve a double counting of benefits.

39. Although very unlikely, in the extreme, the program could produce no net increase in output despite large increases in participants' postprogram earnings by simply reshuffling jobs from non-participants to participants.

BIBLIOGRAPHY

- Ashenfelter, Orley. "The Effect of Manpower Training on Earnings: Preliminary Results," in *Proceedings of the Twenty-Seventh Annual Winter Meeting of the Industrial Relations Research Association*, 1975, pp. 252-260.
- _____. "Estimating the Effects of Training Programs on Earnings," *Review of Economics and Statistics*, February 1978, pp. 47-57.
- Ashenfelter, Orley and David Card. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *Review of Economics and Statistics*, November 1985, pp. 648-660.
- Bassi, Laurie J. "The Effect of CETA on the Postprogram Earnings of Participants," *Journal of Human Resources*, Fall 1983, pp. 539-556.
- _____. "Estimating the Effect of Training Programs with Non-random Selection," *Review of Economics and Statistics*, February 1984, pp. 36-42.
- Bassi, Laurie J., Margaret C. Simms, Lynn C. Burbridge, and Charles L. Betsey. "Measuring the Effect of CETA on Youth and the Economically Disadvantaged" (Washington, D.C.: The Urban Institute, April 1984).
- Bloom, Howard S. and Maureen A. McLaughlin. "CETA Training Programs—Do They Work for Adults," Joint CBO-NCEP Report, July 1982.
- Cooley, Thomas F., Thomas W. McGuire, and Edward C. Prescott. "Earnings and Employment Dynamics of Manpower Trainees: An Exploratory Econometric Analysis," in Farrell E. Bloch (ed.), *Evaluating Manpower Training Programs* (Greenwich, CT: JAI Press, 1979), pp. 119-147.
- Dickinson, Katherine P., Terry R. Johnson, and Richard W. West. "An Analysis of the Impact of CETA Programs on the Components of Earnings," *Industrial and Labor Relations Review*, April 1987(a).
- _____. "An Analysis of the Sensitivity of Quasi-Experimental Evaluations of CETA Programs," *Evaluation Review*, August 1987(b), pp. 452-472.
- _____. "An Analysis of the Impact of CETA Programs on Participants' Earnings," *Journal of Human Resources*, Winter 1986, pp. 64-91.
- Fraker, Thomas and Rebecca Maynard. "An Assessment of Alternative Comparison Group Methodologies for Evaluating Employment and Training Programs," report prepared for DOL under Contract No. 20-11-82-15 (Princeton, NJ: Mathematica Policy Research Inc., June 1984).
- Heckman, James J. "Sample Selection Bias as a Specification Error," *Econometrica*, January 1979, pp. 153-161.
- Heckman, James J. and Richard Robb. "The Longitudinal Analysis of Earnings," Department of Economics, University of Chicago, unpublished manuscript, June 1982.
- _____. "Alternative Methods for Evaluating the Impact of Interven-

- tions: An Overview," Department of Economics, University of Chicago, unpublished manuscript, 1985.
- Johnson, Terry R., Katherine P. Dickinson, and Richard W. West. "An Evaluation of the Impact of ES Referrals on Applicant Earnings," *Journal of Human Resources*, Winter 1985, pp. 117-137.
- Johnson, Terry R. *A Guide for Net Impact Evaluations*, Volume 5, JTPA Evaluation Design Project, Department of Employment Security, Olympia, WA (March 1986).
- Kiefer, Nicholas M. "Federally Subsidized Occupational Training and the Employment and Earnings of Male Trainees," *Journal of Econometrics*, August 1978, pp. 111-125.
- _____. "The Economic Benefits from Four Government Training Programs," in Farrell E. Bloch (ed.), *Evaluating Manpower Training Programs* (Greenwich, CT: JAI Press, 1979), pp. 159-187.
- LaLonde, Robert and Rebecca Maynard. "How Precise Are Evaluations of Employment and Training Programs: Evidence from a Field Experiment," *Evaluation Review*, August 1987, pp. 428-451.
- Maddala, G. S. and Lung Fei Lee. "Recursive Models with Qualitative Endogenous Variables," *Annals of Economic and Social Measurement*, December 1976, pp. 525-545.
- Mallar, Charles et al. *Evaluation of the Economic Impact of the Job Corps Program: Third Follow-up Report* (Princeton, NJ: Mathematica Policy Research, Inc., September 1982).
- Masters, Stanley H. "The Effects of Supported Work on the AFDC Target Group," *Journal of Human Resources*, Fall 1981, pp. 600-636.
- Rubin, Donald B. "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Association*, 1979, pp. 318-328.
- Somers, Gerald G. and Ernst W. Stromsdorfer. *A Cost-Effectiveness Study of the In-School and Summer Neighborhood Youth Corps*, Industrial Relations Research Institute, University of Wisconsin, 1970.
- Westat, Inc. "Net Impact Report Number 1 (Supplement Number 1): The Impact of CETA on 1978 Earnings," report prepared for DOL under Contract No. 23-24-75-07, Westat, Inc., Rockville, MD (July 1982).
- _____. "Summary of Net Impact Results," report prepared for DOL under Contract No. 23-24-75-07, Westat, Inc., Rockville, MD (April 1984).
- Woodbury, Stephen A. "Estimating Preferences for Wage and Nonwage Benefits," paper presented at the NBER Conference on the Economics of Compensation, Cambridge, MA, November 21-22, 1980.