



Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment

James Heckman; Neil Hohmann; Jeffrey Smith; Michael Khoo

The Quarterly Journal of Economics, Vol. 115, No. 2. (May, 2000), pp. 651-694.

Stable URL:

<http://links.jstor.org/sici?sici=0033-5533%28200005%29115%3A2%3C651%3ASADBIS%3E2.0.CO%3B2-X>

The Quarterly Journal of Economics is currently published by The MIT Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/mitpress.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

SUBSTITUTION AND DROPOUT BIAS
IN SOCIAL EXPERIMENTS:
A STUDY OF AN INFLUENTIAL SOCIAL EXPERIMENT*

JAMES HECKMAN

NEIL HOHMANN

JEFFREY SMITH

WITH THE ASSISTANCE OF MICHAEL KHOO

This paper considers the interpretation of evidence from social experiments when persons randomized out of a program being evaluated have good substitutes for it, and when persons randomized into a program drop out to pursue better alternatives. Using data from an experimental evaluation of a classroom training program, we document the empirical importance of control group substitution and treatment group dropping out. Evidence that one program is ineffective relative to close substitutes is not evidence that the type of service provided by all of the programs is ineffective, although that is the way experimental evidence is often interpreted.

I. INTRODUCTION

In recent years a consensus has emerged that training programs for the disadvantaged have little effect on the earnings of participants. This consensus builds in part on experimental impact estimates from the recent National JTPA Study, which evaluated the training programs funded under the Job Training Partnership Act (JTPA), until recently the largest U. S. federal training program for the disadvantaged. That study reports that the training program had only small positive effects on the employment and earnings of adults and zero or even negative effects on youth. The impression of poor program performance produced by this evaluation contributed to a decision by Congress in 1995 to restructure JTPA and to cut funding for JTPA youth programs by over 80 percent. Moreover, this evidence of weak

* This research was supported by National Science Foundation grants SBR 91-11-455 and SBR 93-21-048, and by grants from the Russell Sage Foundation, the Lynde and Harry Bradley Foundation, and the Social Science and Humanities Research Council of Canada. A version of this paper was presented at the American Economic Association meetings in Anaheim, California, in January, 1993 [Heckman and Smith 1992]. We thank seminar participants at the Applications Workshop at the University of Chicago, the NBER Labor Studies Group, the W. E. Upjohn Institute for Employment Research, and the Universities of British Columbia, Québec à Montréal, Victoria, Western Ontario, and Windsor for helpful comments. We thank Edward Vytlačil and Andrew Wong for their expert programming assistance. Gary Becker, Lawrence Katz, Charles Mullin, Derek Neal, and two anonymous referees provided especially helpful comments.

© 2000 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

The Quarterly Journal of Economics, May 2000

effects for one program has been interpreted to mean that the services offered by all programs providing the same or very similar services are ineffective.

JTPA was evaluated by a social experiment in which the outcomes of a treatment group offered program services were compared with the outcomes of a control group randomly excluded from those services. This procedure is widely believed to yield less biased estimates of a program's effect than nonexperimental methods, which potentially suffer from selection bias (see, e.g., Burtless [1995]). In fact, experimental data require careful interpretation.

When experimental control group members choose to take alternative training, including the not infrequent case of taking the same training with alternative funding, then experimental data identify the *effect of the program*—an effect relative to the alternative programs available to control group members. When good substitutes are available to the control group, the difference in outcomes between treatment and control group members does not identify the effect of training relative to no training at all—what we call the *effect of training*. Knowledge of the training effect is important for evaluating the effectiveness of all programs. If there are very good substitutes for a program, the effect of the program estimated by experimental methods can be zero or negative even if the effect of training relative to no training at all is large and positive [Heckman 1992; Heckman and Smith 1993]. In the limit, if all programs are perfect substitutes, no program is effective in comparison with any other program although training may be effective compared with no training at all. Our analysis applies to the evaluation of any program for which at least one alternative program offering an equivalent service is available.

A large fraction of controls in the training experiment we analyze took training from other programs. The quality and duration of the substitute training are similar to those of treatment group training. Moreover, the effect of control group substitution is accentuated by the large fraction of treatment group members who drop out of the program prior to receiving training. Both dropping out by treatments and substitution by controls are so severe that the gap between treatment and control participation in training is not 100 percent, as theoretically desired, but instead falls to as low as 19 percent for some groups.

In the presence of substitution and dropping out, estimation of the training effect requires the use of nonexperimental meth-

ods, even when experimental data are available. Using such techniques to adjust for the effects of control group substitution and treatment group dropout in the experimental data, we obtain estimates of the effect of JTPA classroom training relative to no training. As is true of many educational investments, the effect of training on earnings is found to be negative during training and mostly positive thereafter. Using these estimates, we project the total net private returns to training relative to no training and find that net private returns are large. Based on this finding, we conclude that the evidence from the National JTPA Study on the effect of this type of training on earnings has been misinterpreted. Classroom training, from whatever source, is more effective than policy makers, journalists, and policy experts have been led to believe. At the same time, when the social costs of providing the training are correctly accounted for, it is not clear that the classroom training offered by JTPA passes a cost-benefit test. But this argument is different from the more conventional argument that "training doesn't work," which is based on low estimates of the gross impacts of these programs.

The National JTPA Study we examine is typical in its incidence of control group substitution and treatment group dropout. Table I shows the extent of substitution and dropout in several other major experimental evaluations of job training programs. The National Supported Work (NSW) demonstration is the one program in the table with low rates of substitution and dropout. It provided a very expensive service for which no close substitutes were available. In contrast, many of the other experiments, such as the one that evaluated Project Independence, an employment and training program for welfare recipients in Florida, had even higher rates of substitution and dropout than the JTPA experiment. Substitution and dropout are problems endemic to experimental evaluations of voluntary programs.¹

The paper develops in the following way. Section II considers the interpretation of experimental data in the presence of dropping out and substitution. Section III presents background information on the JTPA program and the National JTPA Study. Section IV describes the empirical evidence on control group substitution and treatment group dropout in the JTPA experiment. Section V presents experimental estimates of the effect of

1. Heckman [1992] presents several examples of this phenomenon in medical trials.

TABLE I
 FRACTION OF EXPERIMENTAL TREATMENT AND CONTROL GROUPS RECEIVING SERVICES
 IN EXPERIMENTAL EVALUATIONS OF EMPLOYMENT AND TRAINING PROGRAMS

Study	Authors/time period	Target group(s)	Fraction of treatments receiving services	Fraction of controls receiving services
1. NSW	Hollister et al. [1984] (9 months after RA)	Long-term AFDC women	0.95	0.11
		Ex-addicts	NA	0.03
		17-20 year old high school dropouts	NA	0.04
2. SWIM	Friedlander and Hamilton [1993] (Time period not reported)	AFDC women: applicants and recipients		
		a. Job search assistance	0.54	0.01
		b. Work experience	0.21	0.01
		c. Classroom training/OJT	0.39	0.21
		d. Any activity	0.69	0.30
		AFDC-U Unemployed fathers		
		a. Job search assistance	0.60	0.01
b. Work experience	0.21	0.01		
3. JOBSTART	Cave et al. [1993] (12 months after RA)	Youth high school dropouts		
		Classroom training/OJT	0.90	0.26
		a. Job search assistance	0.43	0.19
		b. Classroom training/OJT	0.42	0.31
4. Project independence	Kemple et al. [1995] (24 months after RA)	c. Any activity	0.64	0.40
		Teenage single mothers		
		Any education services	0.82	0.48
5. New chance	Quint et al. [1994] (18 months after RA)	Any training services	0.26	0.15
		Any education or training	0.87	0.55

Service receipt includes any employment and training services. RA denotes random assignment to treatment or control groups. In the NSW study, services received by controls are CETA and WIN jobs, for in the Long-term AFDC women group services received also include regular public sector employment. *Sources:* Masters and Maynard [1981], p. 148, Table A.15; Maynard [1980], p. 169, Table A14; Friedlander and Hamilton [1993], p. 22, Table 3.1; Cave et al. [1993], p. 95, Table 4.1; Kemple et al. [1995], p. 58, Table 3.5; Quint et al. [1994], p. 110, Table 4.9.

the program. Section VI presents a simple adjustment to the experimental estimates that takes account of both dropping out and substitution. Section VII uses more elaborate nonexperimental methods to estimate the effect of training on earnings. The evidence from these methods supports the inference from the

simple method of Section VI. Section VIII shows that evidence from standard bounding methods is consistent with the point estimates obtained from conventional econometric estimators. A concluding section summarizes the paper.

II. SOCIAL EXPERIMENTS, DROPOUTS, AND SUBSTITUTION

Unlike researchers conducting experiments in chemistry or biology, researchers conducting a social experiment have only partial control over the level of treatment actually received by treatment and control group members. A social experiment compares the outcomes of persons whose options include participation in the evaluated program to the outcomes of persons who lack this option. Legal and ethical considerations prevent researchers from compelling participation among designated treatment group members or excluding controls from alternative treatments in ways that would improve the interpretability of social experiments. As a result, the experimental impact estimates—the differences in mean outcomes between the treatment and control groups—estimate the effect of program availability, rather than the effect of program participation.

It is helpful to distinguish two important policy questions:

- Q1: What is the effect on mean earnings attributable to the availability of JTPA training given the other training options in place?
- Q2: What is the mean difference in earnings attributable to receiving JTPA training compared with no training at all?

The answer to Q1 is the difference in mean earnings compared with what would have been received by participants in the JTPA program if they had exercised their best non-JTPA training option—the *effect of the program*. The answer to Q2 is the total effect of JTPA classroom training, i.e., the difference in earnings relative to what participants would have earned if they had received no training at all—the *effect of training*. When good substitutes for an evaluated program are available, the effect of the program will be small even if the effect of training is large.

Estimates of the effect of a program provide information about whether that single program should be scaled back or discontinued assuming that other programs remain in place. Estimates of the effect of training should guide decisions about discontinuing all training programs. A program-by-program evalu-

ation of all programs could conclude that each should be eliminated even though any one program is effective compared with no program at all.

Consider a simple model of the choice of a training program in an environment with several training options. Suppose that in some period, say $s = 0$, a person may choose among J available training options, and may also choose to take no training ($j = 0$).² If a person selects choice j , then in period $s (>0)$, he or she receives earnings $Y_{j,s}$ and incurs direct private costs $c_{j,s}$, where $c_{j,s}$ is zero in periods in which no training takes place. Assuming that the person has a constant discount rate δ and lives for S periods after period zero, the present value in period zero of future discounted earnings from choosing j is

$$V_j = \begin{cases} \sum_{s=0}^S \delta^s (Y_{j,s} - c_{j,s}) & j \in \{1, 2, \dots, J\}; \\ \sum_{s=0}^S \delta^s Y_{0,s} & j = 0. \end{cases}$$

If persons have unbiased expectations and seek to maximize their expected present value of discounted earnings, they pick T so that at period $s = 0$,

$$T = \arg \max_{j \in J} E[V_j],$$

where, for simplicity of notation, we suppress the conditioning variables. Let V_T denote the value of the option selected. Let $j = 1$ denote the training program being evaluated (in our empirical analysis, JTPA classroom training). Applicants accepted into JTPA training are denoted by $D = 1$. The social experiment we study randomly assigns accepted applicants into one of two groups: a treatment group allowed to receive program services and a control group excluded from program services. Thus, the set of training options available to a control group member does not include $j = 1$. Let T_{-1} denote the best choice given that "1" is omitted from the choice set. For persons with $D = 1$, let $RN = 1$ indicate random assignment to the treatment group and $RN = 0$ indicate assignment to the control group. The experimental estimator of net returns to JTPA classroom training compares

2. We assume that $J \geq 2$ so that substitution into alternative programs is a concern.

mean treatment group earnings with mean control group earnings and identifies

$$(1) \quad R_0 = E(V_T | D = 1, RN = 1) - E(V_{T_{-1}} | D = 1, RN = 0).$$

This is the mean *effect of the program* as defined in the introduction to this paper. Under standard conditions, replacing population means with sample means produces a consistent estimator of R_0 . Note that R_0 is not the same as R_1 , the mean effect of participation in the program being evaluated compared with participation in no program at all for participants in the program:

$$(2) \quad R_1 = E(V_1 - V_0 | D = 1, T = 1).$$

This is the population answer to Q2 for those who actually take training. This is the evaluation parameter that much of the literature claims to estimate. It is the *effect of training* as defined in the introduction to this paper.³

The two parameters are the same ($R_0 = R_1$) if the following conditions both hold:

(AS-1) There are no dropouts in the treatment group so that $RN = 1 \Rightarrow T = 1$

and

(AS-2) There is no substitution into alternative training programs in the control group so that $RN = 0 \Rightarrow T_{-1} = 0$.⁴

Suppose that an experiment estimates R_0 to be \$0. If both (AS-1) and (AS-2) hold, then \$0 is also a consistent estimate of R_1 . However, if (AS-1) and (AS-2) fail to hold, an estimate of \$0 for R_0 is consistent with a wide range of possible values of R_1 . The difference in R_0 and R_1 solely attributable to violations of (AS-2) is *substitution bias*. The difference attributable to failure of (AS-1) is *dropout bias*. In Section IV we present empirical evidence on violations of both (AS-1) and (AS-2) in the National JTPA Study.⁵

3. Conditioning on $D = 1$ defines the parameter “the mean effect of participating in $T = 1$ compared with no program at all for those who sought to go into the program and were accepted, whether or not they actually participated in the program defined by $T = 1$.” See Heckman [1992, 1997], Heckman and Smith [1998], and Heckman, LaLonde, and Smith [1999] for discussions of a variety of parameters of interest in evaluating social programs and what economic questions they answer.

4. These conditions are jointly sufficient but not necessary.

5. Heckman, Smith, and Taber [1998] analyze the dropout problem but not the substitution problem.

First, we provide some background information on the program we study and on the experiment that generated our data.

III. THE JTPA PROGRAM AND THE NATIONAL JTPA STUDY

Until recently, the Job Training Partnership Act (JTPA) funded the primary federal training program for disadvantaged youth and adults in the United States. This program provided classroom training in occupational skills (CT-OS), basic education (often GED preparation), wage subsidies for on-the-job training at private firms, and job search assistance to persons receiving means-tested government transfers or who had low family incomes in the preceding six months. Although funded at the federal level, the program was primarily administered by the states and by local training centers with independent authority. Classroom training in occupational skills, which forms the primary focus of this study, typically consists of short courses (usually less than six months) provided by community colleges, proprietary schools, or nonprofit organizations. These courses aim to prepare trainees for occupations such as word processing, electronics repair, and home health care.⁶

The National JTPA Study (NJS) was an experimental evaluation of JTPA conducted at a nonrandom subset of 16 of the more than 600 JTPA training centers. Doolittle and Traeger [1990] show that these centers roughly resemble the population of centers in terms of their observable characteristics. Random assignment took place from 1987 to 1989, with the exact dates varying across training centers.⁷ In the NJS, applicants accepted into the program were first recommended to receive particular training services and then randomly assigned to either a treatment group given access to JTPA services or a control group excluded from receiving JTPA services for eighteen months. Follow-up surveys collected information on the earnings and employment outcomes of persons in the experiment. We use these self-reported data to construct the outcome and

6. See National Commission for Employment Policy [1987] or Heckman [1999a] for more detailed descriptions of the JTPA program.

7. Doolittle and Traeger [1990] describe the training center selection process and the implementation of the study in greater detail.

training measures used in this study.⁸ All dollar amounts are in nominal dollars.⁹

In order to analyze the effects of departures from (AS-1) and (AS-2) most clearly, we confine our empirical analysis throughout the paper to persons recommended for classroom training in occupational skills, a group that comprises about one-third of the experimental sample and a similar proportion of the overall JTPA trainee population [U. S. Department of Labor 1992].¹⁰ We omit from our analysis persons recommended to receive subsidized on-the-job training at private firms, the one JTPA service for which few alternative providers exist.¹¹

IV. EVIDENCE OF SUBSTITUTION AND DROPPING OUT IN THE NATIONAL JTPA STUDY

The severity of control group substitution depends on the availability of alternative training options. Unlike the National

8. We use the self-reported earnings data as our outcome measure in preference to the administrative data from state Unemployment Insurance (UI) systems that were also collected for the experimental sample because the latter are available only for persons at twelve of the sixteen training centers in the experiment.

9. The range of years in the main study is 1989–1991, a period of a quiescent price level. In all models we include year dummies to control for year effects.

10. For comparability, our measure of classroom training for both the treatment and control groups is self-reported. It includes high school instruction, GED training, attendance in a two- or four-year college, graduate or professional school, vocational school, and adult education. The JTPA program generally does not provide funding for four-year college degrees or for graduate or professional school. The small number of treatment and control group members receiving these services must therefore have obtained them from another source.

11. By focusing on the receipt of classroom training, we designate as dropouts those treatment group members not receiving such training. About 18.9 percent of treatment group members recommended to receive classroom training received other training services without also receiving classroom training. However, over half of these alternative services consisted of job search assistance or other low-intensity services found to have small impacts in other studies (see, e.g., Gueron and Pauly [1991]). The remainder consists of on-the-job training. The data do not distinguish on-the-job training subsidized by JTPA from that provided in the course of regular employment. Incorporating receipt of other types of training into our empirical framework is hampered by the fact that job search assistance and other training types are measured very poorly in the self-report data [Smith 1999]. The available data suggest that persons not receiving classroom training in the treatment group were more likely to receive these other services than such persons in the control group. In that case, the experimental estimate of the effect of classroom training is upward biased, since it includes in some part the effect of other training. At the same time, the nonexperimental estimates presented in Section VII which use the treatment group dropouts as a comparison group are biased down due to the receipt of other services by the dropouts. That we find large differences between the experimental estimates and our nonexperimental estimates of the training effect, R_1 , in the presence of these biases serves only to strengthen our main argument, that the experimental impact estimate substantially understates R_1 .

Supported Work Demonstration employment subsidy program which offered a unique and expensive treatment with no close substitutes, and in which there was little substitution or dropping out (see Table I), JTPA offered standard services that were provided by many institutions and public agencies. In many cases, JTPA contracted with third parties to provide these services, which were also available through other programs or private purchase, often at subsidized prices. The possibility of substitution bias was accentuated in fourteen of the sixteen training centers in the JTPA experiment where control group members received lists of alternative service providers in their community, thereby making them aware of programs that they might not otherwise have known about.¹²

Table II presents the training experiences of treatment ($RN = 1$) and control ($RN = 0$) group members among accepted applicants ($D = 1$) during the first nineteen months after random assignment. Separate results are reported for adult men and adult women (ages 22 and over), and male and female youth (ages 16 to 21), the same groups analyzed in the official experimental impact reports.

Among controls, the incidence of substitute classroom training ranges from 27 percent for adult males to almost 40 percent for female youth. At the same time, many treatment group members drop out of the program, so that receipt of classroom training varies from a low of 49 percent among adult males to a high of 59 percent among female youth.¹³ Nonetheless, in all four demographic groups a larger fraction of the treatment group received training than the control group, with the difference being statistically significant in all cases. Thus, while assumptions (AS-1) and (AS-2) are clearly violated in the JTPA data, random assignment reduces the incidence of classroom training among controls.¹⁴

12. See Heckman, Smith, and Wittekind [1997] for a study of the determinants of program awareness.

13. Of the treatment group members self-reporting classroom training, 16.8 percent are not recorded as receiving any training in JTPA administrative records. This implies that one or the other, or both, of the reports is in error or that treatment group members took training from non-JTPA sources. Analyses of the self-reported and administrative records reveal no obvious way to resolve the discrepancy. To the extent that treatment group members received classroom training from non-JTPA providers, our estimates of the effect of JTPA training reported in this paper partly reflect the effect of classroom training obtained from other sources.

14. The high incidence of classroom training receipt among the controls is not the result of cross-over, i.e., of controls foiling the experimental protocol by receiving JTPA services. As noted in Bloom et al. [1993], only about 3 percent of controls crossed over and received JTPA services.

TABLE II
CHARACTERISTICS OF CLASSROOM TRAINING IN THE 19 MONTHS FOLLOWING RANDOM ASSIGNMENT

	Adult men		Adult women		Male youth		Female youth				
	Treatment	Control	Prob ($t > T $)	Treatment	Control	Prob ($t > T $)	Treatment	Control	Prob ($t > T $)		
Sample size	744	325		1,697	816		377	174	734	352	
Number receiving CT	363	89		952	272		210	60	430	141	
Percent receiving CT	48.8%	27.4%	0.00	56.1%	33.3%	0.00	55.7%	34.5%	58.6%	40.1%	0.00
Characteristics of persons with one or more training spells											
Average total months of training	6.7	7.6	0.19	7.2	8.0	0.04	7.0	6.4	6.7	7.0	0.56
Average total hours of training	680.7	699.0	0.83	705.9	779.3	0.17	745.7	661.4	765.3	585.2	0.00
Average hours per month of training	110.3	93.5	0.04	100.5	91.4	0.01	110.2	109.5	108.8	87.3	0.00
Fraction of training months employed	50.2%	46.9%	0.77	34.7%	38.1%	0.51	47.8%	59.2%	32.8%	48.0%	0.02
Percent paying for training	16.8%	41.6%	0.00	11.6%	39.0%	0.00	16.7%	48.3%	13.0%	36.2%	0.00
Average monthly payment	\$209	\$358	0.44	\$25	\$101	0.00	\$39	\$103	\$43	\$226	0.21

The sample is rectangular and includes all persons from the sixteen experimental sites with valid data on training receipt. *T*-tests are of the null hypothesis that the means of the treatment and control samples are equal within demographic groups. Classroom training payments are the amount paid by the entire household. Average payments are for each trainee and include trainees who reported zero expenditures.

Figure I presents the fraction of the control and treatment groups in classroom training in the months after random assignment for each demographic group. Consistent with the findings from Table II, within our 33-month window¹⁵ of observation the cumulative level of classroom training receipt for the treatment group exceeds that for the control group. The controls take training later than the treatment group members on average, probably due to the necessity of locating an alternative provider or funding source.

Since most training spells last longer than one month, the patterns of monthly training incidence in Figure I combine the effects of the initiation of new classroom training spells with the continuation of existing ones. An examination of differences in the rates of new spell starts between treatments and controls reveals a large “dose” effect for treatments in the first two or three months after random assignment for all four demographic groups, after which the difference in spell start rates between treatments and controls essentially disappears.

Incidence does not measure the intensity or quality of the alternative services received. If the classroom training received by controls is less intensive or of lower quality than that received by the treatment group, and if more intensive or higher quality training has a larger impact on earnings, then the effect of substitution on the experimental estimates may be less severe than the high level of incidence among controls might suggest. We have good data on the intensity of the training received by controls and some weaker evidence on other aspects of the quality of the substitute training. Table II presents data on the intensity and cost of training for the subsample of each group that reported at least one classroom training spell.

Among those treatments and controls who received classroom training, there is little systematic difference in the average duration of training. Differences in average total months and average total hours in training are statistically significant only in the case of adult women, where controls spent more months in classroom training, and in the case of female youth, where controls spent fewer total hours in training.

Although total hours of training are similar for controls and treatments who take it, mean hours per month in training are

15. The sample includes the month in which random assignment occurs and the following 32 months. These are referred to as the 33 months following random assignment.

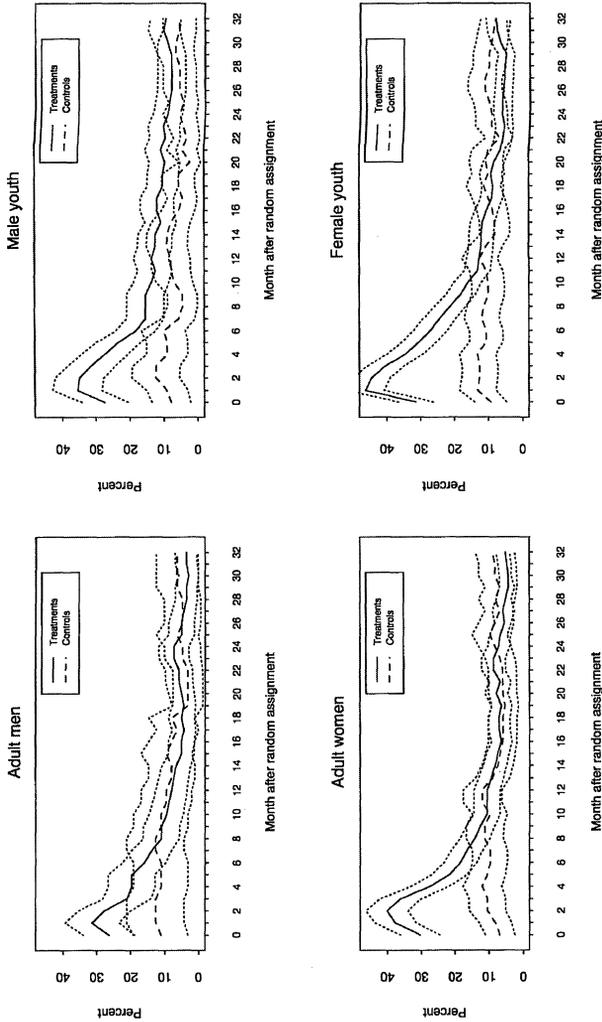


FIGURE I

Percentage Receiving Classroom Training

The percentages are the proportion of persons among the sample who report the receipt of classroom training in each month following random assignment. The sample includes only those persons who responded for the entire 32 months of the survey. Month 0 is the month of random assignment. Standard error bars indicate ± 2 standard errors about the mean.

lower for controls in all four demographic groups—usually statistically significantly so. On average, controls spread their training out over a longer time period. In three of the four groups, controls more often combine work with training, which may help account for their lower average intensity of training. The differences in percent paying and amount paid are large and statistically significant in all cases.¹⁶ However, note that a majority of the controls who took training did not pay for their own instruction. They were able to finance their classroom training through the myriad of federal education programs, such as Pell Grants and Carl Perkins Grants, available to the disadvantaged.

Our evidence on other aspects of the quality of the substitute training received is sketchier but supports the view that substitute classroom training is of high quality. An analysis of the data from one location, Corpus Christi, Texas, reveals that approximately 50 percent of both the treatment and control group members who received classroom training received it at the same community college.

The experimental data from the National JTPA Study clearly indicate that assignment to the control group raises the cost of obtaining classroom training. Equally evident, however, is that many controls seek and find training despite their exclusion from JTPA services. When they obtain training, controls participate in programs of duration and intensity comparable to that of JTPA. Given the small gap in the receipt of training between the treatment and control groups, it would take an enormous amount of perverse selection bias to overturn the intuition that reported experimental impact estimates greatly understate the effect of classroom training relative to no training at all for those who took the training (parameter R_1). In the following sections we present estimates of R_1 that account for both control group substitution and treatment group dropping out. First, however, we report the *experimental estimates of the effect of the program on earnings which form the point of departure for our analysis.*

16. For those respondents reporting a training spell in the period after random assignment, the survey first asks "(Do/Did) you or your family pay anything for this school or training?" For respondents giving a positive response, the survey then asks "What was the total amount your family paid (in addition to funds from grants or scholarships)?" The figures reported in the tables are the mean of the responses to this question, with zeros included for persons responding negatively to the first question.

V. EXPERIMENTAL ESTIMATES OF PROGRAM RETURNS

This section presents experimental estimates of the effect of the program, R_0 . We take as the outcome measure Y_s , the monthly self-reported earnings for persons recommended to receive classroom training (denoted the classroom training “treatment stream” in the experimental analysis) for each of the 33 months after random assignment.¹⁷

We define Δ_s and μ_s to be the mean difference between treatment and control group earnings and direct costs, respectively, in month s after random assignment:

$$\Delta_s = E(Y_s|D = 1, RN = 1) - E(Y_s|D = 1, RN = 0),$$

and

$$\mu_s = E(c_s|D = 1, RN = 1) - E(c_s|D = 1, RN = 0).$$

Assuming a common monthly discount rate δ , the mean effect of the program on discounted earnings is:

$$\begin{aligned} R_0 &= E(V_T|D = 1, RN = 1) - E(V_{T-1}|D = 1, RN = 0) \\ &= \sum_{s=0}^S \delta^s (\Delta_s - \mu_s). \end{aligned}$$

Figure II presents estimates of Δ_s by month after random assignment, conditional on background variables.¹⁸ Estimates for all groups in the first few months after random assignment are negative, consistent with positive opportunity costs for training and with the fact that treatment group members are more likely to take training in the months immediately following random assignment. Although the estimates rise in later months, the mean of Δ_s over the 33-month sample is small or negative, ranging from \$21 for adult females to -\$10 for female youth.

We estimate the net returns to the program, R_0 , by discount-

17. In contrast, Bloom et al. [1993] and Orr et al. [1995] use accumulated earnings in the 18 (or 30) months after random assignment as the measured outcome. Our estimates differ from theirs for a number of other reasons as well. We do not restrict ourselves to the 18-month impact sample as in Bloom et al., nor do we combine earnings information from self-reports and state unemployment insurance records as in Orr et al. We omit monthly observations for which earnings or classroom training receipt data are unavailable. We also trim off the top 1 percent of the earnings observations in each month in both the experimental treatment and control groups rather than attempting to identify outliers on a case-by-case basis.

18. In results not reported here, we find as expected that the experimental impact estimates are not sensitive to the conditioning.

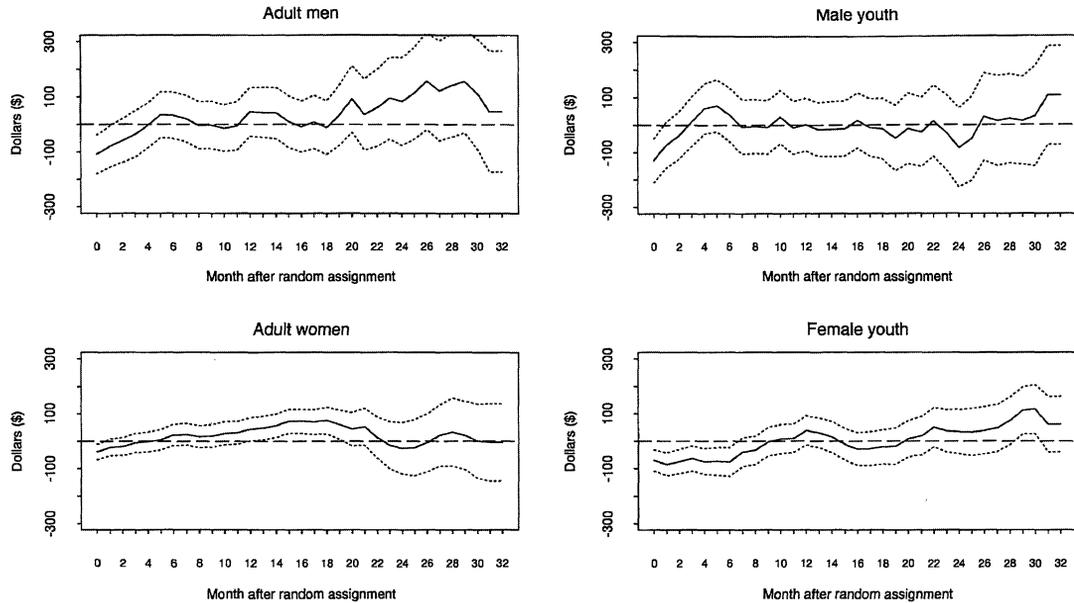


FIGURE II

Experimental Estimates of the Monthly Effect of the JTPA Program

The dependent variable in an OLS regression is self-reported monthly earnings. The sample consists of all person-months in the 32 months after random assignment (RA) with valid values for all variables. Regressors include indicators for treatment status, calendar month, month after RA, treatment status*month after RA, race, marital status, education, training center of random assignment, age, and English language preference. The top 1 percent of earnings values are dropped in each month in both the treatment and control groups. Standard error bars indicate ± 2 Eicher-White robust standard errors about the mean.

ing the stream of monthly estimates $\Delta_s - \mu_s$ for each group using a δ that corresponds to an annual discount rate of 3 percent. Mean self-reported payments from Table II are used as the measure of monthly costs of training, c_s . We estimate R_0 under three assumptions: returns persist only in the 33 months after random assignment for which we have data, or returns persist for a total of 5 or 10 years. To estimate the latter, we must extrapolate Δ_s for periods after the experiment. A U. S. General Accounting Office [1996] report finds that experimentally estimated program effects for JTPA persist at a roughly constant level for at least five years.¹⁹ We assume that Δ_s persists at the mean level of the last 12 months of the 33-month sample. Figure I reveals that, in the last year of the sample, differences in the incidence of training between the treatment and control groups are negligible. Thus, the monthly effect in these and succeeding months primarily reflects differences in the quantity and quality of training induced by randomization.

Table III presents estimates of R_0 , the effect of the program.²⁰ The estimates depend almost entirely on the two earnings streams because the discounted differences in direct training costs over the 33 months following random assignment are fairly small for all four groups. For an annual discount rate of $r = 0.03$, we obtain estimates of R_0 of \$1248 and \$755 for adult men and women, respectively, and \$190 and \$222 for male and female youth, respectively. The returns differ substantially among demographic groups, with estimated net returns for adult males an order of magnitude larger than those for male youth. Table III includes the

19. Couch [1992] finds a similar persistence in the impact estimates from the experimental evaluation of the National Supported Work Demonstration.

20. For purposes of comparison, the experimental estimates of the impact of classroom training in the 18 months after random assignment for the classroom training treatment stream from Bloom et al. [1993] are \$418 for adult males, \$398 for adult females, -\$259 for male youth, and -\$542 for female youth (see Exhibits S.6 and S.12). For the 30 months after random assignment, the impact estimates from Orr et al. [1995] are \$630 for adult women, \$1287 for adult men, -\$438 for male youth, and -\$33 for female youth (see Exhibits 5.7 and 5.17 in their report). The 30-month estimates are adjusted for dropouts in the treatment group and for the small number of controls who receive JTPA services (but not for substitution into alternative sources of training) using the method described in Section VI of this paper. The male youth 30-month estimates refer only to the subsample of male youth not arrested between their sixteenth birthday and random assignment. None of the impact estimates in either report is statistically significant. Heckman, LaLonde, and Smith [1999] discuss differences in the experimental estimates presented in the two official reports. Under the assumption that the program effect lasts 33 months, the pattern of estimates among demographic groups in the official impact reports matches the pattern in Table III. This is not the case for our estimates which extrapolate beyond the available data.

TABLE III
MEAN DISCOUNTED EARNINGS AND ESTIMATES OF THE DISCOUNTED RETURNS TO THE JTPA PROGRAM R_0

	Adult males			Adult females			Youth males			Youth females		
	33 months	5 years	10 years	33 months	5 years	10 years	33 months	5 years	10 years	33 months	5 years	10 years
r	Discounted earnings											
0	25625	49303	101921	12737	25437	53659	20829	41765	88288	11441	22128	45877
0.03	24513	45596	87652	12163	23472	46029	19883	38525	75709	10941	20457	39439
0.10	22149	38254	63369	10945	19584	33054	17875	32115	54321	9878	17147	28483
	Estimates of the discounted returns											
0	1337 (102)	3949 (478)	9755 (1357)	787 (57)	1002 (300)	1481 (860)	209 (102)	491 (409)	1117 (1146)	269 (58)	1704 (231)	4892 (642)
0.03	1248 (98)	3574 (430)	8214 (1132)	755 (54)	947 (270)	1330 (717)	190 (97)	441 (369)	942 (957)	222 (55)	1500 (209)	4048 (536)
0.10	1061 (87)	2838 (338)	5609 (756)	687 (48)	833 (211)	1062 (478)	152 (88)	343 (291)	642 (640)	125 (49)	1101 (165)	2623 (360)
	Estimates of the internal rates of returns											
	1.17	1.27	1.28	2.63	2.63	2.63	0.66	0.79	0.80	0.21	0.53	0.58

Discounted earnings are the present discounted value of mean monthly control group earnings discounted at $1/(1+r)$, where r ranges over 0, 0.03, and 0.10. Estimates of R_0 are the present discounted value of the effect of the program based on program effects with the indicated durations. Monthly earnings beyond the 33-month sample are set at the mean level of months 22 to 33 after random assignment. Estimates are of private returns and include estimated average monthly tuition payments. The internal rate of return is the annual rate of return r such that the net present value of the earnings/cost stream, discounted at $1/(1+r)$, is equal to zero. Rates of return are reported as fractions, *not as percentages*. Internal rate of return estimates are also private estimates and so include the estimated monthly tuition payments. Estimated standard errors appear in parentheses.

mean earnings of the controls for each demographic group in the 33 months after random assignment. This figure provides a baseline against which to compare the estimated program effects.

Net returns increase as the assumed persistence in impacts increases. The salient feature of the estimates is their relatively small size. These low returns reveal the small relative effectiveness of the JTPA program compared with the other programs in place. Raising the annual discount rate to $r = 0.10$ reduces these present values by approximately 20 percent. Reducing the discount rate to $r = 0$ raises these present values by approximately 10 percent. Within a plausible range, the choice of discount rate does not affect the estimates greatly.

Table III also presents private internal rates of return for the program. This is the annual rate of return which equates the present values of earnings for persons with and without the option of JTPA training.²¹ Under different assumptions about the persistence of benefits (33 months, 5 and 10 years), the estimated rates of return are quite large, ranging from 21 percent to 263 percent annually. It is not surprising that the rates of return are so large because the difference in forgone earnings between treatments and controls is small and limited to the first few months following random assignment (see Figure II). In addition, the mean direct costs of training for treatment group members are lower than those of control group members (see Table II).

VI. ESTIMATING TRAINING EFFECTS IN THE PRESENCE OF SUBSTITUTION AND DROPOUT USING AN INTUITIVE ESTIMATOR

Our evidence on widespread substitution and dropping out reported in Section IV suggests that R_0 does not identify the effect of training compared with no training. In this section we develop a simple method for using experimental data to estimate R_1 in the presence of substitution and dropping out. Using the JTPA data, we show that the method produces estimates of R_1 that differ substantially from estimates of R_0 .

If there is dropping out but no substitution, so that assumption (AS-1) is violated but assumption (AS-2) is not, and T is either zero or one, and dropouts experience no effect from partial receipt of training, then the effect of training, R_1 , can be identified from

21. Formally, it is the annualized value of the monthly interest rate r_i that sets $\sum_{s=0}^S (1/(1+r_i)^s)(\Delta_s - \mu_s) = 0$ using the estimated values of Δ_s and μ_s .

the experimental estimator provided that the following assumption holds:

$$(AS-3) \quad E(Y_0|D = 1, RN = 1, Q = 1), \\ = E(Y_0|D = 1, RN = 0, Q = 1),$$

where $Q = 1$ is the event "quit the program after being enrolled in it." For persons denied access to the program by randomization, this is a counterfactual event. Of importance to this paper, assumption (AS-3) states that treatment group dropouts have the same mean earnings as their counterparts in the control group who would have been dropouts if they had been in the treatment group. This is a strong assumption, especially for the NJS, where dropouts may have received partial treatment [Doolittle and Traeger 1990]. Under (AS-3) the effect of training is identified from Δ_s/p_s , where p_s is the proportion of the treatment group that receives training by month s after random assignment. If dropouts receive partial treatment which increases earnings, the true training effect is understated.²²

When both dropping out and substitution characterize experimental data, the problem of estimating the effect of training, R_1 , is essentially the same as that facing an analyst using nonexperimental data [Heckman 1992]. For simplicity, we leave the conditioning on $D = 1$ implicit in all of the expressions that follow. Let q_s be the proportion of persons who receive training by month s after random assignment in any program in the experimental control group. Assume that for persons with $T = 0$ there is no effect of training on the outcome in either experimental group. Then if both (AS-1) and (AS-2) fail to hold, the experimentally determined gross effect of training in postrandom-assignment period s is

$$\Delta_s = p_s E(Y_{1,s} - Y_{0,s} | T = 1, RN = 1) \\ - q_s E(Y_{T-1,s} - Y_{0,s} | T_{-1} > 1, RN = 0).$$

In words, Δ_s is the difference between the mean effect of classroom training on those taking it in the treatment group and in the control group, weighted by the proportions receiving training in each group.²³ Alternatively, this is the weighted difference be-

22. See the discussion and references in Heckman, Smith, and Taber [1998], who show that this estimator is an instrumental variables estimator. This estimator is a version of the Horvitz-Thompson [1952] estimator used in biostatistics. This formula emerges as a special case of the general formula presented in the next footnote, and is a special case of equation (3) discussed in the text below.

23. This formula implicitly assumes the validity of randomization, so that the no-training earnings are on average the same for treatments and controls, and we can subtract off a common mean ($E(Y_0|RN = 1) = E(Y_0|RN = 0)$). It also assumes one of two additional conditions: (a) that persons offered $T = 1$ who take training,

tween the effect of JTPA-provided training and training provided through other sources on those taking it. We cannot identify the effect of JTPA classroom training using the experimental data if we do not know the effect of classroom training from other sources. Put differently, even with experimental data, nonexperimental assumptions are required to identify the training effect in the presence of dropping out and substitution. One simple assumption that guarantees identification of the effect of training is an exchangeability condition:

$$(AS-4) \quad E(Y_{1,s} - Y_{0,s}|T = 1, RN = 1) \\ = E(Y_{T-1,s} - Y_{0,s}|T_{-1} > 1, RN = 0) \forall s.$$

In words, this says that the mean effect of training on those taking it among treatment group members equals the mean effect on those taking it among control group members. This assumption allows for heterogeneous responses to treatment both within and across programs but assumes, like the standard instrumental variables treatment effect estimator (see, e.g., Heckman and Robb [1985]), that participation in training is not based on the idiosyncratic unobserved components of the impacts (see Heckman [1997] and Heckman, LaLonde, and Smith [1999]). Ex ante, all training programs are identical in the eyes of all agents in terms of expected returns, but costs of participation may vary, provided that the costs are independent of the returns. Ex post, there may be considerable heterogeneity in returns.

Under (AS-4) a consistent estimate of the monthly effect of training can be obtained by dividing the monthly experimental estimate through by $p_s - q_s$. The mean net return to training is

take $T = 1$ in preference to training $T > 1$ or (b) that the training $T > 1$ is of comparable quality to $T = 1$, i.e., $(E(Y_{1,s} - Y_{0,s}|T = 1, RN = 1) = E(Y_{T,s} - Y_{0,s}|T > 1, RN = 1)$, where $Y_{T,s}$ is the outcome for training option T . The formula in the text is a special case of a more general formula derived as follows. Let $D_T = 1 (T > 1)$, where 1 is the indicator function ($= 1$, if the inequality is satisfied; $= 0$ otherwise). Define $Y_s = Y_{0,s}(1 - T - D_T) + Y_{1,s}T + Y_{T,s}D_T$. Then $E(Y_s - Y_{0,s}|RN = 1) = E(Y_{1,s} - Y_{0,s}|T = 1, RN = 1) \Pr(T = 1|RN = 1) + E(Y_{T,s} - Y_{0,s}|T > 1, RN = 1) \Pr(T > 1|RN = 1)$. Under assumption (a), $p_s = \Pr(T = 1|RN = 1)$ since $\Pr(T > 1|RN = 1) = 0$. Under assumption (b) $p_s = \Pr(T > 1|RN = 1) + \Pr(T = 1|RN = 1)$ and $E(Y_{1,s} - Y_{0,s}|T = 1, RN = 1) = E(Y_{T,s} - Y_{0,s}|T > 1, RN = 1)$. Using similar reasoning, define $D_{T-1} = 1(T_{-1} > 1)$, and let $\tilde{Y}_s = Y_{0,s}(1 - D_{T-1}) + Y_{T-1,s}D_{T-1}$ so $E(\tilde{Y}_s - Y_{0,s}|RN = 0) = E(Y_{T-1,s} - Y_{0,s}|T_{-1} > 0, RN = 0) \cdot \Pr(T_{-1} > 0|RN = 0)$. Thus, $E(Y_s - Y_{0,s}|RN = 1) - E(\tilde{Y}_s - Y_{0,s}|RN = 0) = E(Y_{1,s} - Y_{0,s}|T = 1, RN = 1) \Pr(T = 1|RN = 1) + E(Y_{T,s} - Y_{0,s}|T > 1, RN = 1) \Pr(T > 1|RN = 1) - E(Y_{T-1,s} - Y_{0,s}|T_{-1} > 1, RN = 0) \Pr(T_{-1} > 1|RN = 0)$ which collapses to the formula in the text under either of the two conditions (a) or (b). As discussed in footnote (11), the discrepancy between the self-reported and administrative records suggests that assumption (a) may be false.

also identified by subtracting out the net differential in costs attributable to the two training sequences. Thus,

$$(3) \quad R_1 = \sum_{s=0}^S \frac{\delta^s}{p_s - q_s} (\Delta_s - \mu_s),$$

which may be consistently estimated by replacing each term on the right-hand side by its sample analogue. This estimator is widely used in the literature (including in Bloom et al. [1993] and Orr et al. [1995], the official reports from the National JTPA Study, where it is employed to adjust for treatment group dropout, and for the small fraction of controls who "cross over" and receive JTPA services, but not close substitutes for those services).²⁴ In the case where earnings impacts per hour enrolled or dollar spent are desired, the $p_s - q_s$ in the denominator is replaced by the difference in hours or dollars spent, respectively. Assumption (AS-4) may be applied to hours, days, incidence, or dollar cost.

(AS-4) is a very strong assumption, but it is supported by the evidence in Section VII of similar impacts of training compared with no training in the treatment and control groups. Even if the assumption is strictly false, its relaxation is unlikely to reverse the main conclusion that the effect of training, R_1 , is substantially greater than the effect of the program, R_0 . In our data there is substantial evidence of dropping out and substitution. As a result, the value of $1/(p_s - q_s)$ is around five. Even if small deviations from (AS-4) occur, the effect of training should still be several times larger than the effect of the program. It is unlikely that most reasonable adjustments will reverse this conclusion, and the evidence from more elaborate estimation and bounding procedures reported in Sections VII and VIII supports this conjecture.

Table IV presents alternative estimates of R_1 based on adjustment of the monthly estimates by $1/(p_s - q_s)$, where the denominator is the difference in either the incidence or the hours of training received between the treatment and control groups.²⁵ For estimates based on the incidence of training, the patterns of

24. Versions of this estimator have a long history. To the best of our knowledge, it was first used in Mallar, Kerachsky, and Thornton [1980] for the case $q_s = 0$. It is identical to estimators defined to deal with contamination bias in nonexperimental evaluations as presented in Heckman and Robb [1985]. See the references in Heckman, Smith, and Taber [1998] for the history of this estimator. Its origins in statistics are older. The Horvitz-Thompson estimator is based on this idea and goes back at least to the 1950s. (See Horvitz and Thompson [1952].)

25. These estimates were first reported in Heckman and Smith [1992]. Kane [1994] applies the estimator (3) to the JTPA impact estimates and confirms our initial findings.

TABLE IV
ESTIMATES OF THE DISCOUNTED RETURNS TO JTPA TRAINING R_1

	Adult males			Adult females			Youth males			Youth females		
	33 months	5 years	10 years	33 months	5 years	10 years	33 months	5 years	10 years	33 months	5 years	10 years
r	Estimates of discounted returns adjusted by training incidence											
0	5268 (530)	17429 (2737)	44454 (7873)	3687 (251)	4622 (1221)	6700 (3474)	847 (526)	2203 (2149)	5218 (5980)	2929 (296)	12186 (1511)	32756 (4318)
0.03	4862 (504)	15690 (2460)	37290 (6562)	3538 (238)	4371 (1098)	6032 (2898)	754 (500)	1962 (1937)	4371 (4995)	2608 (281)	10850 (1359)	27292 (3601)
0.10	4012 (450)	12284 (1925)	25183 (4369)	3215 (211)	3851 (861)	4843 (1933)	564 (446)	1487 (1526)	2926 (3346)	1947 (248)	8243 (1064)	18061 (2401)
	Estimates of discounted returns adjusted by training hours											
0	3739 (427)	12367 (1676)	31542 (4654)	3875 (274)	5114 (1501)	7868 (4325)	758 (467)	2071 (2013)	4989 (5641)	1281 (200)	6998 (959)	19702 (2729)
0.03	3446 (408)	11129 (1512)	26454 (3889)	3714 (259)	4818 (1348)	7019 (3604)	670 (444)	1839 (1813)	4172 (4710)	1088 (190)	6178 (863)	16332 (2276)
0.10	2831 (368)	8700 (1195)	17852 (2608)	3366 (228)	4209 (1053)	5523 (2398)	491 (395)	1385 (1426)	2777 (3150)	690 (169)	4579 (677)	10642 (1519)
	Estimates of the internal rates of return adjusted by training incidence											
	0.79	0.96	0.97	2.45	2.45	2.45	0.45	0.62	0.65	0.43	0.70	0.73
	Estimates of the internal rates of return adjusted by training hours											
	0.73	0.90	0.91	2.44	2.44	2.44	0.41	0.59	0.62	0.26	0.56	0.61

Estimates based on incidence of classroom training receipt are constructed using estimated monthly program effects (Δ_s) adjusted upward by $1/(p_s - q_s)$, where p_s is the proportion of treatments who have received some level of classroom training by month s , and q_s is defined similarly for controls. Estimates based on hours of classroom training receipt are constructed using estimated monthly program effects (Δ_s) adjusted upward by $1/(p_s - q_s)$, where p_s is the average cumulative hours of classroom training received by treatments by month s , and q_s is defined similarly for controls. These constant hourly impact figures are then multiplied by the average cumulative hours of classroom training received by treatments who report at least one CT training spell by month s (i.e., treatments who actually receive training) to yield the monthly effect of training based on constant hourly effects. Monthly earnings beyond the 33-month sample are set at the mean level of months 22 to 33 after random assignment. Estimates are of private returns and include estimated average monthly tuition payments. The internal rate of return is the annual rate of return r such that the net present value of the earnings/cost stream, discounted at $1/(1 + r)$, is equal to zero. Estimated standard errors appear in parentheses.

the returns to training, R_1 , for different demographic groups are very similar to the patterns of the returns to the program, R_0 . The estimated returns to classroom training are four to five times larger than estimated returns to the JTPA program. For estimates based on the intensity of training, as indicated by differences in the number of hours of training completed, the patterns are similar, but the differences between demographic groups are smaller. Again, using discount rates of 10 and zero percent lowers and raises these present values by 20 and 10 percent, respectively. Table IV also presents the associated private internal rates of return to training, which are substantial.²⁶

VII. EVIDENCE FROM ALTERNATIVE ESTIMATION METHODS

The dramatic difference between the estimates of R_0 and R_1 presented in Sections V and VI suggests that the unadjusted estimates from the JTPA experiment seriously understate the effect of classroom training. However, the adjusted estimates presented in Section VI depend on the strong assumption that, on average, classroom training received by persons in the treatment and control groups is equally effective. Although the evidence in Table II provides some support for this assumption, the training effect could still differ substantially between the treatment and control groups.

In this section we present two sets of alternative nonexperimental estimates of the effect of JTPA training. All of these estimates rely on assumptions about the process of selection into training and the earnings process of trainees. The first set uses either the pretraining earnings of the trainees or the earnings of the dropouts to proxy for what the treatment group trainees would have received if they had not taken training. The second set consists of estimates from traditional econometric evaluation estimators described in Heckman and Robb [1985]. Both sets of estimates reinforce the conclusion of Section VI that the unadjusted experimental estimates seriously understate the training effect. In addition, when we apply the methods in this section to

26. Note that the discounted returns rise in going from Table III to Table IV as one relatively large number (the difference in earnings) and one small number (the difference in costs) are both being multiplied by the same monthly adjustment factor of $1/(p_s - q_s)$ before returns are calculated. In contrast, the internal rate of return falls because the temporal pattern of the adjustment factors, which rise and then fall as the treatment group "dose" effect plays out, accentuates the costs relative to the benefits of training.

estimate the effect of control group training, we obtain estimates similar to those obtained for JTPA training using the treatment group. These results provide additional support for the assumption that controls who take training receive the same quality of training as those in the treatment group.

A. Simple Differences-in-differences Cross-Section Estimates

To define the nonexperimental estimators used in this section, it is necessary to introduce some additional notation. In a given month, a treatment group member can be in one of four states. Three states are for persons who at some point receive classroom training ($T = 1$); they can be observed prior to, during, or after that training. The fourth state is for persons who do not receive any training ($T = 0$) over our sample period. We denote potential monthly earnings in each state as follows: Y_p for earnings in months following random assignment, but prior to training (prior "p"), Y_d for earnings in months during which training is received (during "d"), Y_a for earnings in months after the end of training (after "a") and Y_0 for earnings when $T = 0$.

In this framework, earnings and training histories can be used to identify separate earnings effects from training for months during and after training. The mean monthly effect of current training on persons for whom $T = 1$, Δ_d , is the mean difference between monthly earnings during training and monthly earnings in the absence of any training:

$$(4) \quad \Delta_d = E(Y_d|T = 1) - E(Y_0|T = 1).$$

Similarly, the mean monthly effect of completing training, Δ_a , is the mean difference between monthly posttraining earnings and monthly earnings in the absence of any training:

$$(5) \quad \Delta_a = E(Y_a|T = 1) - E(Y_0|T = 1).$$

The second term on the right-hand side of both expressions, monthly earnings in the absence of training for persons for whom $T = 1$, is an unobserved counterfactual.

In order to estimate Δ_d and Δ_a , we must find appropriate proxies to substitute for $E(Y_0|T = 1)$. We use two proxies: the monthly earnings of nontrainees, $E(Y_0|T = 0)$, and the monthly earnings of trainees prior to receipt of training, $E(Y_p|T = 1)$. We construct cross-section estimates by taking mean differences with respect to $E(Y_0|T = 0)$ and before-after estimates by taking mean differences with respect to $E(Y_p|T = 1)$.

Using the earnings of nontrainees or of trainees prior to training in place of the unobserved counterfactual raises the possibility of selection bias. If the monthly earnings of trainees in the absence of training correspond to the unobserved counterfactual conditional on characteristics X , or

$$(AS-5) \quad E(Y_0|T = 1, X) = E(Y_0|T = 0, X),$$

then the cross-section estimates consistently estimate R_1 . This assumption is termed "selection on observables" by Heckman and Robb [1985] and is used, for example, by Barnow, Cain, and Goldberger [1980]. It is a weaker assumption than traditionally used in matching (see Heckman, LaLonde, and Smith [1999] or Heckman, Ichimura, and Todd [1997]). Similarly, if, conditional on X , the monthly earnings of trainees prior to training identify the unobserved counterfactual, or

$$(AS-6) \quad E(Y_0|T = 1, X) = E(Y_p|T = 1, X),$$

then our before-after estimates of Δ_d and Δ_a will be consistent.

Table V presents differences-in-differences and cross-section estimates of Δ_d and Δ_a . For all of the person-months in the experimental treatment group with usable data, we estimate the regression,

$$Y_s = \alpha_0 + \alpha_1 D_p + \alpha_2 D_d + \alpha_3 D_a + X' \beta_X + D_s \beta_s + u,$$

where D_p , D_d , and D_a are indicators for person-months in the three states of prior to training, during training, and after training, respectively, where the baseline group is persons with no training, and where X is a vector of observed characteristics including calendar month of random assignment, and D_s is a vector of indicator variables for the month after random assignment in which the earnings observation occurs. By conditioning on the month s , we remove any common trend in earnings in the months after random assignment operating on all groups. The coefficient on D_a estimates the (conditional) mean difference between the posttraining earnings of trainees and the earnings of nontrainees. This coefficient corresponds to the cross-section estimate of Δ_a . The difference between the coefficients on D_a and D_p estimates the mean difference between posttraining earnings and pretraining earnings for trainees. This difference is the conditional difference-in-differences estimate of Δ_a . Cross-section and conditional difference-in-differences estimates of Δ_d are similarly defined. As we use

TABLE V
NONEXPERIMENTAL ESTIMATES OF THE MONTHLY EFFECTS OF TRAINING Δ_d AND Δ_a ON TREATMENT GROUP MEMBERS

	Adult males		Adult females		Youth males		Youth females	
	Difference- in differences	Cross- section						
Sample size	19438		42943		11328		22052	
Mean earnings gain	777		386		632		347	
Estimates of Δ_d for persons with 1-4 months of training								
	-57 (75)	-231 (48)	-51 (32)	-127 (21)	-17 (93)	-189 (59)	-57 (48)	-67 (32)
Estimates of Δ_d for persons with >4 months of training, having completed 1-4 months								
	-68 (59)	-273 (46)	-75 (25)	-189 (17)	-129 (64)	-180 (46)	-66 (29)	-111 (22)
Estimates of Δ_d for persons with >4 months of training, having completed >4 months								
	-96 (86)	-302 (66)	-117 (32)	-231 (24)	-207 (77)	-259 (62)	-44 (38)	-89 (29)
Estimates of Δ_a for persons with 1-4 months of training								
	280 (92)	105 (55)	111 (43)	36 (29)	169 (107)	-3 (61)	77 (57)	66 (38)
Estimates of Δ_a for persons with >4 months of training								
	230 (90)	25 (60)	215 (39)	101 (31)	70 (90)	18 (55)	169 (42)	124 (33)

The dependent variable in the OLS regression is pooled self-reported monthly earnings. The regression sample consists of all person-months for treatments at the sixteen experimental sites in the 33 months after random assignment (RA) with valid values for all variables. Regressors include indicators for training status, calendar month, month after RA, race, marital status, education, training center of RA, age, and English language preference. The excluded training status is never receiving training. Separate sets of training status indicators are used to estimate effects for training spells of one to four months or more than five months in duration. The top 1 percent of earnings values are dropped in each month. Mean earnings is the mean level of monthly earnings for members of the control group who reported earnings over the full 33-month period. Estimated standard errors appear in parentheses.

multiple monthly observations from the same individuals, we report robust standard errors to allow for serial correlation in u .

We estimate Δ_d separately for three groups: persons currently in a one–four month spell of classroom training, persons currently in the first four months of a spell of five or more months, and persons currently in the fifth or later month of a spell of five or more months. Both the conditional difference-in-differences and cross-section estimates of Δ_d are negative for all groups, but the conditional difference-in-differences estimates are smaller in absolute value in all cases. In addition, longer durations of training are associated with a more negative effect on monthly earnings. These estimates are consistent with the negative experimental estimates in the first few months after random assignment in Figure II.

We estimate Δ_a separately for persons who received one–four months and five or more months of classroom training. The estimates in both cases are generally large, positive, and statistically significant. The large size of these estimates relative to the estimates of Δ_s suggests that the monthly experimental estimates substantially understate the effect of training completion on monthly earnings.²⁷

Table VI presents estimates of the net returns to training, R_1 , constructed using the difference-in-differences and cross-section estimates of the mean effects of training, Δ_d and Δ_a , from Table V. The estimated net return to training is the discounted present value of the future stream of training effects on net earnings assuming training begins in the first month after random assignment:

$$R_1 = \sum_{s=0}^m \delta^s (\Delta_d - E[c_s | D = 1, T = 1]) + \sum_{s=m+1}^M \delta^s \Delta_a.$$

27. The difference-in-differences estimates of Δ_a are uniformly higher than the cross-section estimates. This suggests that (AS-5) is not valid and that, conditional on X , nontrainee earnings are higher than trainee earnings would have been if they had not taken training. The differences between the two sets of estimates are consistent with a model with selection into the program on the basis of a time-invariant fixed effect, where persons with low values of the fixed effect are more likely to take training. The difference-in-differences estimator differences out the bias, as does the symmetric differences estimator in Heckman and Robb [1985], but the cross-section estimator does not. It is likely that the difference-in-differences estimator is biased upward due to the dip in earnings prior to training [Heckman and Smith 1999]. Heckman, Ichimura, Smith, and Todd [1998] and Heckman, Ichimura, and Todd [1997, 1998] discuss alternative methods for accounting for selection bias in nonexperimental evaluations. Heckman, LaLonde, and Smith [1999] summarize the evidence in this literature and discuss this specific model of bias.

TABLE VI
NONEXPERIMENTAL ESTIMATES OF THE DISCOUNTED RETURNS TO TRAINING, R_1 FOR TREATMENT GROUP MEMBERS

Training length	Adult males			Adult females			Youth males			Youth females		
	33 months	5 years	10 years	33 months	5 years	10 years	33 months	5 years	10 years	33 months	5 years	10 years
Private returns based on before-after estimates												
3 months	7362 (2529)	13349 (4490)	24086 (8013)	2801 (1194)	5175 (2122)	9433 (3787)	4421 (2962)	8044 (5257)	14541 (9380)	1789 (1572)	3428 (2791)	6367 (4980)
6 months	4964 (2233)	9889 (4158)	18720 (7622)	4601 (972)	9202 (1811)	17451 (3321)	533 (2216)	2032 (4126)	4719 (7562)	3572 (1046)	7193 (1949)	13685 (3572)
Social returns based on before-after estimates												
3 months	2730 (2529)	8717 (4490)	19453 (8013)	-1884 (1194)	490 (2122)	4749 (3787)	-184 (2962)	3440 (5257)	9936 (9380)	-2842 (1572)	-1203 (2791)	1736 (4980)
6 months	405 (2233)	5330 (4158)	14161 (7622)	-33 (972)	4568 (1811)	12817 (3321)	-3974 (2216)	-2476 (4126)	211 (7562)	-977 (1046)	2644 (1949)	9136 (3572)
Private returns based on cross-section estimates												
3 months	2066 (1503)	4326 (2668)	8377 (4761)	510 (802)	1272 (1425)	2639 (2544)	-804 (1693)	-860 (3004)	-959 (5358)	1464 (1048)	2874 (1861)	5403 (3321)
6 months	-1277 (1496)	-746 (2781)	206 (5094)	1136 (766)	3298 (1429)	7174 (2620)	-1037 (1361)	-644 (2526)	62 (4625)	2203 (808)	4860 (1506)	9624 (2761)
Social returns based on cross-section estimates												
3 months	-2566 (1503)	-307 (2668)	3744 (4761)	-4175 (802)	-3413 (1425)	-2046 (2544)	-5409 (1693)	-5464 (3004)	-5563 (5358)	-3167 (1048)	-1757 (1861)	772 (3321)
6 months	-5837 (1496)	-5306 (2781)	-4353 (5094)	-3498 (766)	-1336 (1429)	2540 (2620)	-5544 (1361)	-5151 (2526)	-4446 (4625)	-2346 (808)	311 (1506)	5075 (2761)

Returns are the present discounted value of the training effect, discounted at $1/(1+r)$, where $r = 0.03$, and based on estimated monthly effects of training Δ_t and Δ_s . The duration of training is set at 3 or 6 months and the effects of training persist for either 33 months, 5 years, or 10 years. Private returns include the estimated monthly tuition payments. Social returns include the estimated marginal costs incurred by the training provider, adjusted upward by 1.5 to reflect the deadweight cost of taxation.

A discount rate of $r = 0.03$ is used. This expression is appropriate for observations for which the length of the training spell is m months, the effects of training are assumed to persist for M months and no further costs are incurred after m months. We examine net returns when training spells are of three months or six months duration and the effects persist for 33 months, 5 years, or 10 years. Net returns are examined using two measures of the costs of training c_s : private costs only and both private and social costs of training. Mean private costs are based on reported monthly tuition among those receiving training. Mean social costs are based on the marginal cost of providing JTPA classroom training to an additional person as reported by Heinrich [1996],²⁸ scaled up by a factor of 1.50 to account for the deadweight cost of taxation (see, e.g., Browning [1987]).²⁹

The estimates reported in Table VI are very sensitive to assumptions about the persistence of the training effects. Including social costs also has a substantial effect. Among the estimates of social returns that rely solely on the 33 months for which we have data, 11 of 16 are negative, and only one, for female youth, is positive and statistically significant at conventional levels. Cases with positive private returns often have negative net social returns. Generally, the estimated private returns to JTPA classroom training are in the neighborhood of a few thousand dollars. These estimates are consistent with the estimated bounds on the effect of training reported in Section VIII below. They are also close to the adjusted experimental estimates of R_1 reported in Table IV, although the two sets of estimates are based on different assumptions. Table VI shows JTPA classroom training in a strikingly different light than the unadjusted experimental estimates. At the same time, subsidization of classroom training may not be supported under a strict cost-benefit criterion that looks only at net social returns, inclusive of the welfare cost of taxation.

In estimating the effect of JTPA classroom training, we have

28. We use marginal costs because they are easier to obtain, given that average cost estimates must somehow account for the overhead costs which are shared among all the types of training provided by JTPA.

29. This measure of social costs will overstate true social costs if classroom training reduces the probability of receiving unemployment insurance or AFDC payments. The resulting decline in deadweight loss associated with revenues for these programs may offset some of the costs of training. Experimental estimates of the effect of access to JTPA on AFDC receipt indicate no significant effect among adult women and youth, and a *positive* effect for adult men (see Orr et al. [1995]). There are no estimates of the effect of access to JTPA on receipt of unemployment insurance benefits from the JTPA experiment.

restricted our sample to treatment group members. We can also use the control group to estimate the effect of non-JTPA classroom training. Table VII is constructed for the control group in the same manner as Table V for the treatment group. The effect of training among controls appears to be broadly similar to the effect among treatments, providing support for assumption (AS-4) that justifies the intuitive estimator of Section VI. Estimates of Δ_α are somewhat higher for controls, which is consistent with the higher average out-of-pocket costs shown in Table II.³⁰

B. Traditional Econometric Estimates of the Training Effect

In this section we complement the nonexperimental estimates of the training effect just discussed by briefly presenting estimates obtained using a variety of other commonly used nonexperimental estimators. Heckman and Robb [1985], Heckman and Smith [1996], and Heckman, LaLonde, and Smith [1999] offer comprehensive discussions of alternative methods for eliminating selection bias. In this section we apply some of these methods using monthly earnings in each of the nineteen months after random assignment as the dependent variable in separate regressions.³¹

Each method is based on different assumptions about the heterogeneity of the training effect, the information sets of potential participants, and the decision rule that governs selection into training conditional on reaching random assignment.³² Let X denote a set of observable variables affecting earnings, and let Z denote a set of observable variables affecting participation in classroom training. It is assumed that some Z are not in X . Methods I–IV assume that conditioning on a set of observed variables controls for selection bias. Different methods use this information in different ways. (Heckman, LaLonde, and Smith [1999] provide an extensive discussion of this point.) Method I conditions on X , as in Barnow, Cain, and Goldberger [1980], Method II conditions on X and Z , as in Heckman, Ichimura, Smith and Todd [1998]. Method III, based on the analysis of Rosenbaum

30. The analogue to Table VI constructed using control group data is available from the authors upon request.

31. Earnings in months after month 19 are assumed to equal the mean earnings in the final six months of the available data. The sample is restricted to those persons who either receive no training or have completed training within one year after random assignment in order to isolate the effect of training completion for extrapolation to months beyond 19.

32. Details on the construction of the various estimators are discussed in the notes to Table VIII and in an appendix available from the authors upon request.

TABLE VII
NONEXPERIMENTAL ESTIMATES OF THE MONTHLY EFFECTS OF TRAINING Δ_d AND Δ_a for Control Group Members

Sample size	Adult males		Adult females		Youth males		Youth females	
	8696		20815		5447		10449	
	Difference-in-differences	Cross-section	Difference-in-differences	Cross-section	Difference-in-differences	Cross-section	Difference-in-differences	Cross-section
	Estimates of Δ_d for persons with 1-4 months of training							
	-283 (111)	-215 (116)	-107 (50)	-149 (41)	-93 (75)	-66 (98)	-138 (49)	-227 (44)
	Estimates of Δ_d for persons with >4 months of training, having completed 1-4 months							
	-146 (95)	-372 (64)	-86 (44)	-98 (36)	-133 (101)	-256 (83)	-9 (34)	-172 (38)
	Estimates of Δ_d for persons with >4 months of training, having completed >4 months							
	-198 (120)	-425 (76)	-115 (58)	-127 (45)	10 (107)	-113 (93)	-91 (43)	-254 (42)
	Estimates of Δ_a for persons with 1-4 months of training							
	85 (141)	153 (117)	95 (72)	53 (55)	-3 (99)	25 (70)	47 (65)	-42 (56)
	Estimates of Δ_a for persons with >4 months of training							
	379 (139)	152 (87)	116 (72)	104 (56)	211 (115)	88 (105)	238 (62)	74 (58)

The dependent variable in the OLS regression is pooled self-reported monthly earnings. The regression sample consists of all person-months for controls at the sixteen experimental sites in the 33 months after random assignment with valid values for all variables. Regressors include indicators for training status, calendar month, month after RA, race, marital status, education, training center of random assignment, age, and English language preference. The excluded training status is never receiving training. Separate sets of training status indicators are used to estimate effects for training spells of one to four months or more than five months in duration. The top 1 percent of earnings values are dropped in each month. Estimated standard errors are in parentheses.

and Rubin [1983], conditions on the estimated probability of participation in the program, or propensity score. Method IV is the commonly used instrumental variables approach. As noted by Heckman [1997], Heckman and Smith [1998], and Heckman, LaLonde, and Smith [1999], the IV method is closely related to the method of matching. Method V is the sample selection correction method of Heckman [1979], which assumes that the unobservables in the outcome and participation equations are jointly normally distributed.³³ It is the only method among these five that allows for selection on unobserved (by the economist) components of gain.

Table VIII presents the results of estimating R_1 using each method. Note that these estimates, unlike those in the preceding subsection, do not distinguish between the effect of being in training and the effect of having completed training. Rather, like the adjusted estimates in Section VI, they combine the two effects into a net effect of training relative to no training at all.

The estimates in Table VIII are generally of the same sign, and of similar magnitude, as those in Table VII, although it is important to keep in mind the large standard errors obtained from many of the estimators. The estimates from Method V, which are motivated by selection on unobservable components of gain not fully captured by the available X and Z , are generally larger than Methods I–III, which assume selection only on observables. The unstable performance of the IV estimator has been found in many studies. See the discussion and Monte Carlo estimates in Heckman, LaLonde, and Smith [1999]. The larger estimates from Method V suggest that unobserved factors that positively affect the receipt of training are negatively correlated with the unobservables in the earnings equation. This ordering of the estimates is reasonable if persons tend to drop out of the program after random assignment as a result of receiving attractive alternative arrangements for training or employment outside the program. However, the evidence in Heckman, Ichimura, Smith, and Todd [1998] indicates that the parametric normality assumption used to construct this estimator is often at odds with data of the type analyzed here.

33. Heckman, Ichimura, Smith, and Todd [1998] extend this estimator to a semiparametric setting. Heckman and Robb [1986] establish the relationship between matching and selection models based on the propensity score.

TABLE VIII
NONEXPERIMENTAL ESTIMATES OF THE DISCOUNTED RETURNS TO TRAINING, R_1 , CORRECTED FOR SELECTION

Adult males			Adult females			Youth males			Youth females		
33 months	5 years	10 years	33 months	5 years	10 years	33 months	5 years	10 years	33 months	5 years	10 years
I. Regression-adjusted on X											
1912	4226	8840	1191	3037	6719	647	1537	3313	1843	4163	8789
(387)	(890)	(1977)	(195)	(456)	(1016)	(426)	(991)	(2205)	(241)	(564)	(1257)
II. Regression-adjusted on X and Z											
2255	4964	10367	2250	5085	10740	1936	3897	7809	1186	3120	6977
(447)	(1026)	(2278)	(234)	(548)	(1220)	(573)	(1342)	(2991)	(341)	(800)	(1784)
III. Regression-adjusted on X and Z through propensity score											
1672	3901	8348	1527	3635	7840	1472	3081	6291	1917	4454	9514
(516)	(1111)	(2436)	(304)	(713)	(1589)	(926)	(1856)	(3999)	(286)	(661)	(1471)
IV. Instrumental-variables estimates											
8201	14836	28071	3816	9556	21005	-2407	-5214	-10812	10520	20358	39983
(2408)	(5489)	(12171)	(1425)	(3349)	(7469)	(2071)	(4883)	(10895)	(1944)	(4655)	(10414)
V. Heckman [1979] method											
17025	28568	51595	9343	20310	42185	1074	1331	1845	4466	9346	19079
(2709)	(6233)	(13846)	(1582)	(3714)	(8279)	(2118)	(4985)	(11119)	(1637)	(3828)	(8528)

Returns are the present discounted value of the estimated monthly effects of training, discounted at $1/(1+r)$, where $r = 0.03$. The selection-corrected monthly effect of training for treatment group members is estimated through separate regressions for months 0 through 18 after random-assignment (RA). For succeeding months, the effect of training is taken to be the mean of the training effects for months 13–18. In order to isolate the effect of training completion, the sample is restricted to either those who receive no training or those who complete their training within 12 months of RA. The top 1 percent of earnings values in each month are excluded. The dependent variable Y in each regression is a person's self-reported earnings in that month. Exogenous regressors X in the earnings equation include indicators for race, marital status, education, site of random assignment, and age. T is a treatment indicator. $T = 1$ signifies that an individual has participated in training by the month of the regression. Participation-related regressors Z include month of random assignment, household size, indicators for progressively higher levels of total family income, and indicators for receipt of adult basic education, vocational training, and job search assistance at or before random assignment. The estimated training effect for Method I is the coefficient on T in an OLS regression of Y on X . For Method II, it is the coefficient on T in an OLS regression of Y on X and Z (see Heckman, Ichimura, Smith, and Todd [1998]). For Method III, it is the coefficient on T in an OLS regression of Y on X , P , and T , where P is the predicted probability of participation from a probit of T on Z . (See Heckman and Robb [1986] for discussion of this method.) For Method IV, the 2SLS model, it is the coefficient on T in an OLS regression of Y on X and P , where P is the predicted value from an OLS regression of T on X and Z . For Method V, the Heckman two-step model, it is the coefficient on T in an OLS regression of Y on X , M , and T , where M is the estimated inverse Mills ratio from a probit of T on Z . The estimated standard error of Method V does not incorporate the additional variance component resulting from the first-stage estimation. Estimated standard errors are in parentheses.

VIII. BOUNDING THE EFFECT OF TRAINING

The recent literature on program evaluation draws on ideas from the robust estimation and sensitivity analysis literatures in statistics to develop methods for bounding the impacts of treatments without imposing exact identifying assumptions or imposing specific functional forms as we have done in Sections VI and VII. In this section we estimate bounds on the training effect using the experimental data and the bounding strategies originally developed in Robins [1989] and Manski [1990] and refined in later work by Horowitz and Manski [1995] and applied by Hotz, Mullin, and Sanders [1997].³⁴ This approach represents a complement to the intuitive estimator of Section VI and the more traditional econometric estimators presented in Section VII. Evidence from application of these bounds supports our conclusion that the effect of the program understates the effect of training.

We find bounds on

$$\Delta_s = E(Y_{1,s}|T_s = 1) - E(Y_{0,s}|T_s = 1),$$

where Δ_s denotes the (gross) training effect for month s after random assignment, Y_s denotes earnings in month s and where $T_s = 1$ for treatment group members who initiate training by month s and zero otherwise. The second term on the right-hand side is not observed; we bound Δ_s by estimating bounds on this term.³⁵

The estimated bounds presented in Table IX result from successively imposing the following assumptions:

(AS-7) The support of the earnings distribution in each month equals (Y^l, Y^u) , where Y^u equals the maximum value of earnings observed in the data and $Y^l = 0$.

(AS-8) Anyone who takes alternative training by month s when in the control group would take JTPA training by month s if in the treatment group; that is, $T_{-1,s} > 1 \Rightarrow T_s = 1$.

(AS-9) The training effect is nonnegative following completion of training and the earnings process is stationary.

34. The essential ideas in these methods appear in the robust estimation literature. Clear antecedents for their application to the selection problem are the papers by Glynn, Laird, and Rubin [1986] and Holland [1986]. Heckman and Vytlačil [2000] and Heckman [2001] present the intellectual history of these methods.

35. See Heckman, Hohmann, Smith, and Khoo [1998] for a formal derivation of the bounds employed here.

TABLE IX
ESTIMATED BOUNDS ON THE DISCOUNTED RETURNS TO JTPA TRAINING R_1

Adult males			Adult females			Youth males			Youth females		
33 months	5 years	10 years	33 months	5 years	10 years	33 months	5 years	10 years	33 months	5 years	10 years
Lower bound with (AS-7) imposed											
-61254 (363)	-113194 (688)	-216799 (1788)	-48755 (179)	-94469 (359)	-185655 (949)	-52544 (314)	-104590 (565)	-208407 (1440)	-40457 (167)	-79469 (300)	-157288 (764)
Lower bound with (AS-7) and (AS-8) imposed											
-44294 (698)	-81262 (1271)	-155004 (3256)	-37048 (296)	-73848 (572)	-147254 (1497)	-41059 (578)	-84791 (1002)	-172024 (2519)	-33611 (281)	-68192 (470)	-137171 (1162)
Lower bound with (AS-7), (AS-8), and (AS-9) imposed											
-2214 (800)	-942 (1441)	1595 (3678)	1133 (373)	3002 (707)	6729 (1837)	-1772 (670)	-2496 (1156)	-3940 (2901)	-1129 (341)	96 (581)	2540 (1449)
Upper bound with (AS-7) imposed											
21824 (321)	40864 (628)	78842 (1648)	13007 (144)	25155 (286)	49388 (753)	17160 (274)	33047 (505)	64736 (1299)	11208 (141)	22186 (261)	44084 (674)
Upper bound with (AS-7), (AS-8), and, optionally, (AS-9) imposed											
13015 (698)	29303 (1271)	61795 (3256)	8255 (296)	15924 (572)	31222 (1497)	9609 (578)	19532 (1002)	39325 (2519)	4902 (281)	11916 (470)	25909 (1162)

Estimates of bounds on R_1 are the present discounted value of the estimated monthly bounds on the effect of training less the average monthly cost of training, discounted at $1/(1+r)$, where $r = 0.03$. Standard errors of R_1 are derived from the discounted sum of individual monthly variance estimates on the training effect bounds. Estimated standard errors are in parentheses.

We obtain upper and lower bounds on the discounted net returns to training by constructing bounds for each month.

Consistent with the evidence from similar bounding strategies in the literature (see, e.g., Heckman and Smith [1993] and Heckman, Smith, and Clements [1997]), the estimated bounds reported in Table IX are very wide. For example, for adult women the bounds from imposing only (AS-7) under the assumption that the training effect persists 33 months are $-\$48,775$ and $\$13,000$.

Only when we impose (AS-7), (AS-8), and (AS-9) together, do the bounds have some bite. For adult women the lower bounds on the training effect exceed the estimated program effects in Table III for all three assumptions regarding the duration of the effect. For example, the lower bound in Table IX for the 33 month duration for adult females is $\$1133$, which exceeds the estimated program effect in Table III of $\$755$ (for $r = 0.03$). Despite their width, the evidence from these bounds provides additional support for the view that the experimentally estimated program effects represent downward-biased estimates of the effect of training relative to no training.

Overall, we draw two lessons from the estimates using the intuitive IV estimator in Section VI, the estimates from traditional cross-section econometric estimators presented in Section VII, and the bounds presented in this section. First, as is well-known, alternative assumptions about the selection process can yield very different estimates of the impact of training on earnings. Second, and more important for our purposes, the conclusions drawn from the methods employed in Sections VI and VII are robust to a wide variety of alternative methods for dealing with postrandom-assignment selection into JTPA. In virtually every case that we examine, the nonexperimental estimates of the effect of training relative to no training are positive and are well in excess of the experimental estimates of the effect of the JTPA program relative to the available alternatives. The nonparametric bounds, while wide, support this conclusion.

IX. CONCLUSION

While Heckman [1992], Heckman and Smith [1993, 1995], and others raise the issue of substitution bias in theory, and the problem of dropout bias is widely discussed, this paper provides the first systematic empirical examination of the importance of

control group substitution, combined with treatment group drop-out, in an actual social experiment.³⁶ Using data on persons in the National JTPA Study recommended to receive classroom training, we demonstrate that experimental control group members receive substantial training. This training is similar in duration and intensity to the training received by JTPA participants. At the same time, many treatment group members drop out of JTPA without receiving any training. These two factors reduce the difference in the fraction receiving training in the treatment and control groups from the commonly assumed value of 1.0 to about 0.2.

In the presence of substitution and dropping out, the simple experimental estimator no longer corresponds to the effect of training on those who receive it. Applying nonexperimental estimation methods to earnings data for the experimental treatment group, we find a sizable negative effect of participation in training on the monthly earnings of persons while they are taking training and a large positive effect on monthly earnings of completing training. Individual net returns to training are estimated under a variety of conditions. Returns are generally several times larger than those calculated using the unadjusted experimental estimates as estimates of the training effect. Lower bound estimates of the training effect constructed from simple bounding strategies rule out the unadjusted experimental estimates for some groups. Nonexperimental estimates constructed using the data from the control group suggest that the effect of alternative classroom training programs is roughly similar to the effect of JTPA classroom training.

Access to JTPA only marginally enhances the training opportunities of prospective participants, but classroom training from whatever source appears to have a positive effect on its recipients. Experimental evaluations of each of the substitutes for JTPA would also find small program effects. Such experimental evidence does not prove that classroom training is ineffective. The weight of the evidence suggests that it is effective, at least when measured by private returns. The question of the sign of net social returns remains an open one and depends crucially on assumptions about the duration of training effects, the deadweight cost of

36. Puma et al. [1990] note substantial substitution in the program they analyze but do not adjust their estimates for this source of bias.

taxation and the discount rate. For other components of the JTPA program such as publicly subsidized on-the-job training, there are many fewer substitutes, and reported estimates are less vulnerable to the substitution bias discussed in this paper. The reported experimental estimates of program effectiveness for these components are more likely to be reliable guides to public policy.³⁷

Our evidence suggests that experimental evaluations cannot be treated as if they automatically produce easily interpreted and valid answers to questions about the effectiveness of social programs. Reporting experimental estimates by themselves without placing them in the context in which treatments and controls operate invites misinterpretation.

DATA APPENDIX

This appendix provides information on the data we use from the NJS and on the construction and characteristics of our analysis sample. Between November, 1987, and September, 1989, 20,601 applicants accepted into JTPA ($D = 1$) at one of the sixteen training centers in the NJS were randomly assigned to either the treatment group ($RN = 1$) or the control group ($RN = 0$). The random assignment ratio was (almost always) two treatments to every control.

About eighteen months after random assignment the experimental samples were contacted for the first of two follow-up interviews. A random subset of the experimental samples was contacted about 32 months after random assignment for a second follow-up interview. For cost reasons, the interviews were staggered around the 18 and 32 month dates. Some 17,689 persons reported data during the follow-up period (12,069 treatments and 5620 controls). Data from the follow-up surveys on earnings, employment, and other outcomes were converted into monthly data aligned relative to random assignment by Abt Associates. These monthly data, in combination with information on demographic characteristics from the Background Information Form (BIF) administered at the time of random assignment, form the basis of our analysis. A total of 4315 respondents have at least 33 months of follow-up data. In total, we have access to 440,623

37. Recall that Orr et al. [1995] use the intuitive IV estimator presented in Section VI to adjust for dropouts. See also Heckman, Smith, and Taber [1998].

person-month observations (an average of 24 months per person), although these observations may have missing values for particular variables.

For reasons cited in the main text, we focus exclusively on persons assigned to the classroom training treatment stream. This restriction reduces our sample size from 17,689 persons with some follow-up data to 6188 persons (or 154,118 person-months). Within the classroom training treatment stream we have, for adult males, 889 treatments and 392 controls, for adult females, 2042 treatments and 977 controls, for male youth, 437 treatments and 207 controls, and for female youth 852 treatments and 392 controls.

The follow-up surveys collected detailed information on all spells of education and training. Our measure of classroom training combines the following types of training from the survey data: high school, GED preparation, two-year college, four-year college, graduate or professional school, vocational education, and adult education classes. The survey data do not distinguish between training provided by JTPA and that received elsewhere.

Appendix 1 presents selected demographic and other characteristics of our sample, broken down by demographic group and by random assignment status. These variables are drawn from the BIF data. The third column for each subgroup presents the *p*-value from a test of equal means for the indicated characteristic between treatment and control group members. The sample sizes for this table are slightly lower than those listed above as the table includes only persons with nonmissing values for all of the demographic characteristics displayed.

UNIVERSITY OF CHICAGO, AMERICAN BAR FOUNDATION, AND NBER
UNIVERSITY OF CHICAGO
UNIVERSITY OF WESTERN ONTARIO AND NBER
MINNESOTA PUBLIC RADIO

APPENDIX 1: DEMOGRAPHIC CHARACTERISTICS OF TRAINEES AND NONTRAINEES

	Adult male trainees			Adult male nontrainees			Adult female trainees			Adult female nontrainees		
	Treat-ments	Con-trols	Prob ($t > T $)	Treat-ments	Con-trols	Prob ($t > T $)	Treat-ments	Con-trols	Prob ($t > T $)	Treat-ments	Con-trols	Prob ($t > T $)
Sample size	363	89		381	236		952	272		745	544	
Percent black	33.1%	37.1%	0.48	42.8%	36.4%	0.12	27.0%	26.5%	0.86	40.4%	37.7%	0.32
Percent Hispanic	11.0%	10.1%	0.80	5.0%	11.4%	0.01	15.9%	17.3%	0.58	10.1%	11.0%	0.58
Percent with 12 years schooling	68.6%	77.5%	0.08	69.8%	67.4%	0.53	60.5%	63.2%	0.41	62.6%	59.0%	0.20
Percent with >12 years schooling	26.4%	25.8%	0.91	23.1%	22.5%	0.85	16.6%	18.8%	0.42	15.4%	14.3%	0.58
Percent employed at random assignment	20.7%	29.2%	0.11	16.3%	18.2%	0.54	18.1%	20.6%	0.36	16.0%	17.8%	0.38
Percent received AFDC at RA	14.3%	20.2%	0.21	16.8%	17.4%	0.85	63.4%	59.6%	0.25	57.3%	62.7%	0.05
	Male youth trainees			Male youth nontrainees			Female youth trainees			Female youth nontrainees		
	Treat-ments	Con-trols	Prob ($t > T $)	Treat-ments	Con-trols	Prob ($t > T $)	Treat-ments	Con-trols	Prob ($t > T $)	Treat-ments	Con-trols	Prob ($t > T $)
Sample size	210	60		167	114		430	141		304	211	
Percent black	22.4%	26.7%	0.51	27.5%	27.2%	0.95	24.2%	24.8%	0.88	28.3%	25.6%	0.50
Percent Hispanic	30.0%	21.7%	0.18	21.6%	24.6%	0.56	28.1%	25.5%	0.54	17.1%	24.6%	0.04
Percent with 12 years schooling	41.4%	41.7%	0.97	30.5%	40.4%	0.09	49.3%	56.7%	0.13	47.4%	45.0%	0.60
Percent with >12 years schooling	6.2%	11.7%	0.23	0.6%	1.8%	0.40	4.7%	9.2%	0.09	4.3%	2.4%	0.22
Percent employed at random assignment	18.6%	28.3%	0.13	18.6%	17.5%	0.83	20.7%	25.5%	0.25	16.4%	21.8%	0.13
Percent received AFDC at RA	16.2%	16.7%	0.93	13.8%	8.8%	0.19	45.1%	42.6%	0.60	38.2%	39.3%	0.79

The sample is rectangular and includes all persons from the sixteen experimental sites, recommended to receive classroom training, T -tests are of the null hypothesis that means of the treatment, and control samples are equal within demographic and training groups.

REFERENCES

- Barnow, Burt, Glen Cain, and Arthur Goldberger, "Issues in the Analysis of Selectivity Bias," in E. Stromsdorfer and G. Farkas, eds., *Evaluation Studies Review Annual, Volume 5* (San Francisco: Sage, 1980), pp. 290–317.
- Bloom, Howard, Larry Orr, George Cave, Stephen Bell, and Fred Doolittle, *The National JTPA Study: Title II-A Impacts on Earnings and Employment at 18 Months* (Bethesda, MD: Abt Associates, 1993).
- Browning, Edgar, "On the Marginal Welfare Cost of Taxation," *American Economic Review*, LXXVII (1987), 11–23.
- Burtless, Gary, "The Case for Randomized Field Trials in Economic and Policy Research," *Journal of Economic Perspectives*, IX (1995), 63–84.
- Cave, George, Hans Bos, Fred Doolittle, and Cyril Toussaint, *JOBSTART: Final Report on a Program for School Dropouts* (New York: Manpower Demonstration Research Corporation, 1993).
- Couch, Kenneth, "New Evidence on the Long-Term Effects of Employment and Training Programs," *Journal of Labor Economics*, X (1992), 380–388.
- Doolittle, Fred, and Linda Traeger, *Implementing the National JTPA Study* (New York: Manpower Demonstration Research Corporation, 1990).
- Friedlander, Daniel, and Gayle Hamilton, *The Saturation Work Initiative Model in San Diego: A Five-Year Follow-up Study* (New York: Manpower Demonstration Research Corporation, 1993).
- Glynn, Robert, Nan Laird, and Donald Rubin, "Mixture Modeling vs. Selection Modeling," in H. Wainer, ed., *Drawing Inferences From Self-Selected Samples* (New York: Springer Verlag, 1986), pp. 115–142.
- Gueron, Judith, and Edward Pauly, *From Welfare to Work* (New York: Russell Sage Foundation, 1991).
- Heckman, James, "Sample Selection Bias as a Specification Error," *Econometrica*, XLVII (1979), 153–161.
- , "Randomization and Social Program Evaluation," in C. Manski and I. Garfinkel, eds., *Evaluating Welfare and Training Programs* (Boston: Harvard University Press, 1992), pp. 201–230.
- , "Instrumental Variables: A Study of Implicit Behavioral Assumptions in One Widely Used Estimator," *Journal of Human Resources*, XXXII (1997), 441–462.
- , *Performance Standards in a Government Bureaucracy: Analytical Essays on the JTPA Performance Standards System*, unpublished monograph prepared for W. E. Upjohn Institute for Employment Research, 1999.
- , "Causal Parameters and Policy Analysis: A Twentieth Century Retrospective," *Quarterly Journal of Economics*, CXV (2000), 45–98.
- , "The Econometric Approach to Program Evaluation," in J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Volume V* (Amsterdam: North-Holland, 2001).
- Heckman, James, Neil Hohmann, and Jeffrey Smith, with Michael Khoo, "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment," University of Western Ontario Research Report No. 9819, 1998.
- Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd, "Characterizing Selection Bias Using Experimental Data," *Econometrica*, LXVI (1998), 1017–1098.
- Heckman, James, Hidehiko Ichimura, and Petra Todd, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, LXIV (1997), 605–654.
- Heckman, James, Hidehiko Ichimura, and Petra Todd, "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, LXV (1998), 261–294.
- Heckman, James, Robert LaLonde, and Jeffrey Smith, "The Economics and Econometrics of Active Labor Market Policies," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics: Volume III* (Amsterdam: North-Holland, 1999).
- Heckman, James, and Richard Robb, "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data* (Cambridge: Cambridge University Press, 1985), pp. 156–243.

- Heckman, James, and Richard Robb, "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in H. Wainer, ed., *Drawing Inferences from Self-Selected Samples* (New York: Springer Verlag, 1986), pp. 63–107.
- Heckman, James, and Jeffrey Smith, "Experimental and Nonexperimental Evaluation of Training Programs," unpublished manuscript, University of Chicago, 1992, presented at the American Economic Association meetings, Anaheim, California, January 1993.
- Heckman, James, and Jeffrey Smith, "Assessing the Case for Randomized Evaluation of Social Programs," in K. Jensen and P. K. Madsen, eds., *Measuring Labour Market Measures: Evaluating the Effects of Active Labour Market Policies* (Copenhagen: Ministry of Labour, 1993), pp. 35–96.
- Heckman, James, and Jeffrey Smith, "Assessing the Case for Social Experiments," *Journal of Economic Perspectives*, IX (1995), 85–110.
- Heckman, James, and Jeffrey Smith, "Experimental and Non-experimental Evaluation," in G. Schmid, J. O'Reilly, and K. Schömann, eds., *International Handbook of Labour Market Policy and Evaluation* (London: Edward Elgar, 1996), pp. 37–88.
- Heckman, James, and Jeffrey Smith, "Evaluating the Welfare State," in S. Strom, ed., *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial* (Cambridge: Cambridge University Press, 1998), pp. 241–318.
- Heckman, James, and Jeffrey Smith, "The Pre-Program Earnings Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies," *Economic Journal*, CIX (1999), 1–37.
- Heckman, James, Jeffrey Smith, and Nancy Clements, "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, LXIV (1997), 487–536.
- Heckman, James, Jeffrey Smith, and Christopher Taber, "Accounting for Dropouts in Evaluations of Social Programs," *Review of Economics and Statistics*, LXXIX (1998), 1–14.
- Heckman, James, Jeffrey Smith, and Marybeth Wittekind, "Awareness and Participation in Social Welfare Programs," University of Chicago, unpublished manuscript, 1997.
- Heckman, James, and Edward Vytlacil, "The Econometric Evaluation of Social Programs," in J. Heckman and E. Leamer, eds. *Handbook of Econometrics, Volume V* (Amsterdam: North-Holland, 2001).
- Heinrich, Carolyn, "JTPA Unit Training Costs," University of Chicago, unpublished manuscript, 1996.
- Holland, Paul, "A Comment on Remarks by Hartigan and Rubin," in H. Wainer, ed., *Drawing Inferences From Self-Selected Samples* (New York: Springer Verlag, 1986), pp. 149–151.
- Hollister, Robinson, Peter Kemper, and Rebecca Maynard, *The National Supported Work Demonstration* (Madison: University of Wisconsin Press, 1984).
- Horowitz, Joel, and Charles Manski, "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, LXIII (1995), 281–302.
- Horvitz, Daniel, and David Thompson, "A Generalization of Sampling with Replacement from a Finite Population," *Journal of the American Statistical Association*, XLVII (1952), 663–685.
- Hotz, V., Joseph, Charles Mullin, and Seth Sanders, "Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analyzing the Effects of Teenage Childbearing," *Review of Economic Studies*, LXIV (1997), 575–605.
- Kane, Thomas, "Reconciling Experimental and Non-Experimental Evidence on the Returns to Postsecondary Training," Harvard University, unpublished manuscript, 1994.
- Kemple, James, Daniel Friedlander, and Veronica Fellerath, *Florida's Project Independence: Benefits, Costs and Two-Year Impacts of Florida's JOBS Program* (New York: Manpower Demonstration Research Corporation, 1995).
- Mallar, Charles, Stuart Kerachsky, and Craig Thorton, "The Short-Term Economic Impact of the Job Corps Program," in E. Stromsdorfer and G. Farkas, eds., *Evaluation Studies Review Annual, Volume 5* (Beverly Hills: Sage Publications, 1980).

- Manski, Charles, "Nonparametric Bounds on Treatment Effects," *American Economic Review*, LXXX (1990), 319-323.
- Masters, Stanley, and Rebecca Maynard, *The Impact of Supported Work on Long-Term Recipients of AFDC Benefits* (New York: Manpower Demonstration Research Corporation, 1981).
- Maynard, Rebecca, *The Impact of Supported Work on Young School Dropouts* (New York: Manpower Demonstration Research Corporation, 1980).
- National Commission for Employment Policy, *The Job Training Partnership Act* (Washington, DC: U. S. Government Printing Office, 1987).
- Orr, Larry, Stephen Bell, Winston Lin, George Cave, and Fred Doolittle, *The National JTPA Study: Impacts, Benefits and Costs of Title IIA* (Bethesda, MD: Abt Associates, 1995).
- Puma, Michael, Nancy Burstein, Katie Merrell, and Gary Silverstein, *Evaluation of the Food Stamp Employment and Training Program: Final Report* (Bethesda, MD: Abt Associates, 1990).
- Quint, Janet, Barbara Fink, and Sharon Rowser, *New Chance: Interim Findings on a Comprehensive Program for Disadvantaged Mothers and their Children* (New York: Manpower Demonstration Research Corporation, 1994).
- Robins, James M., "The Analysis of Randomized and Nonrandomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in L. Sechrest, H. Freeman, and A. Uley, eds., *Health Service Research Methodology: A Focus on AIDS, NCHSR* (Washington, DC: U. S. Public Health Service, 1989), pp. 113-159.
- Rosenbaum, Paul, and Donald Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, LXX (1983), 41-55.
- Smith, Jeffrey, "How Well Do We Measure Public Job Training?" University of Western Ontario, unpublished manuscript, 1999.
- United States Department of Labor, *JTPA Title IIA and III Enrollments and Terminations During Program Year 1990 (July 1990-June 1991)* (Washington, DC: Employment and Training Administration, Office of Strategic Planning and Development, 1992).
- United States General Accounting Office, "Job Training Partnership Act: Long-Term Earnings and Employment Outcomes," GAO Report to Congressional Requesters HEHS-96-40, March 1996.