1998

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

James Heckman

Neil Hohmann

Jeffrey A. Smith

ISSN:0318-725X
ISBN:0-7714-2141-9

RESEARCH REPORT 9819

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

by

James Heckman
Neil Hohmann
Jeffrey Smith
with the assistance of Michael Khoo

November 1998

Department of Economics
Social Science Centre
University of Western Ontario
London, Ontario, Canada
N6A 5C2
econref@julian.uwo.ca

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

James Heckman[1]
Neil Hohmann
Jeffrey Smith
with the assistance of Michael Khoo

First Draft: January 1993
Current Draft: November 17, 1998

[1]Heckman is Henry Schultz Distinguished Service Professor of Economics and Director of the Center for Social Program Evaluation, University of Chicago, a Senior Research Fellow of the American Bar Foundation and a Faculty Research Fellow of the NBER. Hohmann and Khoo are Research Associates at the Center for Social Program Evaluation at the University of Chicago. Khoo is also a graduate student at Columbia. Smith is Assistant Professor of Economics at the University of Western Ontario and a Faculty Research Fellow of the NBER. This research was supported by NSF SBR 91-11-455, NSF 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 AEA meetings in Anaheim, California, January, 1993. (Heckman and Smith, 1992). We thank seminar participants at the Applications Workshop at the University of Chicago, the NBER Labor Studies Group and the Universities of British Columbia, Victoria and Western Ontario for helpful comments. We thank Edward Vytlacil and Andrew Wong for their expert programming assistance. Gary Becker, Larry Katz, Charles Mullin, Derek Neal and two anonymous referees provided especially helpful comments.

Abstract

This paper considers the interpretation of results from social experiments when persons randomized out of the program being evaluated have good substitutes for it, and when persons randomized into the program drop out to pursue better alternatives. Using data on persons randomized into (or out of) a classroom training program, we document the empirical importance of substitution in the control group and dropping out in the treatment group. Evidence that one program is ineffective compared to close substitutes for it is not evidence that the type of service provided by all of the programs is ineffective, although that is the way the experimental evidence is often interpreted.
JEL Numbers: H43, C93, C14

1 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 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 to 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 to 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 methods, 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 evidence, we conclude that the evidence from the National JTPA Study for the effectiveness of this type of training on earnings has been misinterpreted. Classroom training, from whatever source, is more effective than policymakers, 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 1-1 shows the extent of substitution and dropout in several other major experimental evaluations of job training programs. The National Supported Work 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 2 considers the interpretation of experimental data in the presence of dropping out and substitution. Section 3 presents background information on the JTPA program and the National JTPA Study. Section 4 describes the empirical evidence on control group substitution and treatment group dropout in the JTPA experiment. Section 5 presents experimental estimates of the effect of the program. Section 6 presents a simple adjustment to the experimental estimates that takes account of both dropping out and substitution. Section 7 shows that evidence from standard bounding methods is consistent with the estimates reported in Section 6. Section 8 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 6. A concluding section summarizes the paper.

2 Social Experiments, Dropouts and Substitution

Unlike researchers conducting experiments in chemistry or plant 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

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

the other training options in place?

Q2: What is the mean difference in earnings attributable to receiving JTPA training compared to no training at all?

The answer to Q1 is the difference in mean earnings compared to what would have been received by participants in the JTPA program had they 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 had they 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 evaluation of all programs could conclude that each should be eliminated even though any one program is effective compared to 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 one member j of a set J of available training options, where J includes the option of taking no training ($j = 0$) and may include any of J potential training programs ($j \in \{1, 2, \ldots, J\}$).[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

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

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

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

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

where 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 $D = 1$. A social experiment 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 J 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 mean treatment group earnings to mean control group earnings and identifies

$$R_0 = E[V_T \mid D = 1, RN = 1] - E[V_{T_{-1}} \mid D = 1, RN = 0]. \tag{1}$$

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 to participation in no program at all for participants in the program:

$$R_1 = E[V_1 - V_0 \mid D = 1, T = 1]. \tag{2}$$

This is the population answer to Q2 for those who actually take training. This is the evaluation parameter claimed to be estimated in much of the literature. 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:

[3] Conditioning on $D = 1$ alone in equation (2) defines the parameter "the mean effect of participating in $T = 1$ compared to 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 discussion of a variety of parameters of interest in evaluating social programs and what economic questions they answer.

(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 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 4, we present empirical evidence on violations of both **(AS-1)** and **(AS-2)** in the National JTPA Study.[5] First, we provide some background information on the program we study and on the experiment that generated our data.

3 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. Though 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 non-profit organizations. These courses aim to prepare trainees for occupations such as word processing, electronics repair and home health care.[6]

[4]These conditions are jointly sufficient but not necessary.

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

[6]See National Commission for Employment Policy (1987) or Heckman (1998) for more detailed descriptions of the JTPA program.

The National JTPA Study (NJS) was an experimental evaluation of JTPA conducted at a non-random 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 18 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 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]

---

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

[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 12 of the 16 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]Our measure of classroom training is self-reported, and includes high school instruction, GED training, attendance in a 2- or 4-year college, graduate/professional school, vocational school, and/or adult education. The JTPA program generally does not provide funding for 4-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, and Woodbury and Spiegelman, 1987). 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,

4 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 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 1-1), 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 are 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 were provided with lists of alternative service providers in their community, thereby making persons aware of programs which they might not otherwise have known about.[12]

Table 4-1 presents the training experiences of treatment ($RN = 1$) and control ($RN = 0$) group members among accepted applicants ($D = 1$) during the first 19 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, adult males are the group with the lowest incidence of training. Approximately 27 percent found substitute classroom training. In contrast, almost 40 percent of the female youth controls found substitute classroom training. 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] While the rates of control substitution

1998). 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 8 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.

[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

and treatment dropping out are quite high, 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]

Figure 4-1 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 4-1, within our 33-month window[15] of observation the cumulative level of classroom training receipt for the treatment group remains higher than that for the control group. Training is delayed but not prevented for many controls.

Since most training spells last longer than one month, the patterns of monthly training incidence in Figure 4-1 combine the effects of the initiation of new classroom training spells with the continuation of existing ones. To separate out these two effects, Figure 4-2 plots the differences in the rates of new spell starts between treatments and controls by demographic group. For all four demographic groups, we find a large "dose" effect for treatments in the first two or three months after random assignment, 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 much 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 4-1 presents data on the intensity and cost of training for the subsample

records reveal no obvious way to resolve the discrepancy. To the extent that treatment group members may have received classroom training from non-JTPA providers, estimates of the effect of JTPA training reported in this paper partly reflect the effect of classroom training obtained from other sources, although access to these other sources could be attributable to the initial participation in the JTPA program.

[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.

[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.

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, on average, in the length of training taken. 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 both controls and treatments who take it, mean hours per month in training are lower for controls in all four demographic groups — usually statistically significantly so. On average, controls spread their training out over a longer time period. The differences in percent paying and amount paid are large and statistically significant in all cases.[16,17] However, note that a majority of the controls who took training did not pay for their own instruction. Persons randomized out of JTPA were able to finance their classroom training through the myriad of federal government education programs available to low income and disadvantaged persons.[18,19]

Our evidence on other aspects of the quality of the substitute training received is more sketchy but supports the view that substitute classroom training is of high quality. An analysis of the

[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.

[17]Table 4-1 reveals that adult treatment group members work during training about as often as control group members, despite being less likely to pay for their training. In contrast, among youth, controls were substantially more likely to work during training, indicating that they may have had to combine work and training in order to finance unsubsidized instruction.

[18]These include Pell Grants and Carl Perkins Grants, among other programs.

[19]It is of interest to examine whether differences in tuition costs between the treatment and control groups explain the higher incidence of training in the treatment group. Table 4-1 presents the mean amounts paid for training by treatments and controls, expressed in terms of both monthly and annual tuition for comparison purposes. Cameron and Heckman (1992, revised 1998) report that a $1000 rise in tuition payments results in a 12 percentage point decrease in community college participation rates. The elasticity of enrollment with respect to tuition required to rationalize Table 4-1 varies by demographic group. For adult males it is about the same as that reported in Cameron and Heckman — 12.8 percentage points. For female youth it is a bit lower and for adult women and male youth it is two to three times higher. The Cameron-Heckman estimate falls in the range of estimates reported in the literature, including the study of Kane (1994).

data from one location, Corpus Christi, Texas, reveals that approximately 60 percent of the control group members who received classroom training obtained it at the same community college where treatment group trainees received training.[20]

The experimental data from the National JTPA Study clearly indicate that assignment to the control sample 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. The unadjusted estimates from the JTPA experiment compare the experiences of a treatment group, 55 percent of whom receive training, to a control group, 33 percent of whom receive training. The absence of differences in the duration and intensity of the services received in the two groups, and the weaker evidence that many controls took training at the same locations as treatment group members, suggests that small experimental estimates are consistent with sizeable effects of training on earnings. 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.

5 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 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.[21]

[20]Both treatment and control group members were asked where they got it. But many did not answer the question and among those who did the data were coded with so many spelling errors that analysis of them becomes extremely difficult.

[21]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

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:

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

and

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

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

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

Figure 5-1 presents estimates of Δ_s by month after random assignment, conditional on background variables.[22] 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 discounting the stream of monthly estimates $\Delta_s - \mu_s$ for each group using a δ that corresponds to an annual discount rate of three percent. Mean self-reported payments from Table 4-1 are used as the measure of monthly costs of training c_s. We estimate R_0 under two 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

as well. We do not restrict ourselves to the 18 month impact sample as in Bloom, *et al.* (1993), nor do we combine earnings information from self-reports and state unemployment insurance records as in Orr, *et al.* (1995). We omit monthly observations for which earnings or classroom training receipt data are unavailable. We also trim off the top one 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.

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

must extrapolate Δ_s for periods after the experiment. A U.S. General Accounting Office (1996) report finds that experimentally-estimated training effects in the JTPA program persist at a roughly constant level for at least five years.[23] We assume that Δ_s persists at the mean level of the last twelve months of the 33 month sample. Figure 4-1 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 5-1 presents estimates of R_0, the effect of the program.[24] 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 an 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 5-1 includes the 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 to the other programs in place. Raising the annual discount rate to

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

[24]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, \$1,287 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 6 of this paper. The male youth 30-month estimates refer only to the subsample of male youth not arrested between their 16th 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 5-1. This is not the case for our estimates which extrapolate beyond the available data.

$r = 0.10$ reduces these present values by approximately 20 percent. Reducing the discount rate to $r = 0$ raises these present values by approximately 10 percent. Within plausible range, the choice of discount rate does not affect the estimates greatly.

Table 5-1 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.[25] 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 foregone earnings between treatments and controls is small and limited to the first few months following random assignment (see Figure 5-1). In addition the mean direct costs of training for treatment group members are lower than those of control group members (see Table 4-1).

6 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 4 suggests that R_0 does not identify the effect of training compared to 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 the experimental estimator provided that the following assumption holds:

(AS-3) $E[Y_0 \mid D = 1, RN = 1, T = 0] = E[Y_0 \mid D = 1, RN = 0, T = 0]$.

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 had

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

they 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.[26]

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 post-random-assignment period s is

$$\Delta_s = p_s E\left[Y_{1,s} - Y_{0,s} \,|T = 1\,, RN = 1\right] - q_s E\left[Y_{T_{-1},s} - Y_{0,s} \,|T_{-1} > 1\,, RN = 0\right].$$

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.[27] Alternatively, this is the weighted difference between the effect of JTPA-provided training

[26]See the discussion and references in Heckman, Smith and Taber (1998), who show that this estimator is an instrumental variables estimator. 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.

[27]This formula implicitly assumes the validity of randomization so that the no training earnings are on average the same for treatments and controls so that we can subtract off a common mean ($E(Y_0 \mid RN = 1) = E(Y_0 \mid RN = 0)$), and one of two additional conditions: (a) that persons offered $T = 1$ who take training, take $T = 1$ in preference to training $T > 1$ and/or (b) that the training $T > 1$ is of comparable quality to $T = 1$, *i.e.* ($E(Y_{1,s} - Y_{0,s} \mid T = 1, RN = 1) = E(Y_{T,s} - Y_{0,s} \mid T > 1, RN = 1)$ 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 $E(Y_s - Y_{0,s} \mid RN = 1) = E(Y_{1,s} - Y_{0,s} \mid T = 1, RN = 1)\Pr(T = 1 \mid RN = 1) + E(Y_{T,s} - Y_{0,s} \mid T > 1, RN = 1)\Pr(T > 1|RN = 1)$ .Under assumption (a), $p_s = \Pr(T = 1 \mid RN = 1)$ since $\Pr(T > 1 \mid RN = 1) = 0$. Under assumption (b) $p_s = \Pr(T > 1 \mid RN = 1) + \Pr(T = 1 \mid RN = 1)$ and $E(Y_{1,s} - Y_{0,s} \mid T = 1, RN = 1) = E(Y_{T,s} - Y_{0,s} \mid T > 1, RN = 1)$.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 $E(\tilde{Y}_s - Y_{0,s} \mid RN = 0) = E(Y_{T_{-1},s} - Y_{0,s} \mid T_{-1} > 0, RN = 0) \cdot \Pr(T_{-1} > 0 \mid RN = 0)$. Thus $E(Y_s - Y_{0,s} \mid RN = 1) - E(\tilde{Y}_s - Y_{0,s} \mid RN = 0) = E(Y_{1,s} - Y_{0,s} \mid T = 1, RN = 1)\Pr(T = 1 \mid RN = 1) + E(Y_{T,s} - Y_{0,s}|T > 1, RN =$

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 is an exchangeability condition:

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

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

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

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.[28] 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

1) $\Pr(T>|RN=1)-E(Y_{T_{-1},s}-Y_{0,s}\mid T_{-1}>1,RN=0)\Pr(T_{-1}>1\mid RN=0)$ 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 suggest that assumption (a) may be false.

[28]Versions of this estimator have a long history. To the best of our knowledge, it was first used in Mallar, Kerachsky and Thorton (1980) for the case $q_s = 0$. It is employed in Bloom, *et al.* (1993) and Orr, *et al.* (1995) to adjust

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 8 of similar impacts of training compared to 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 5. 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 procedures reported in Section 8 supports this conjecture.

Table 6-1 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.[29] For estimates based on the incidence of training, the patterns of 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 6-1 also presents the associated private internal rates of return to training, which are substantial.

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). 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.

[29]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.

7 Bounding the Effect of Training

The recent literature proposes methods for bounding the impacts of treatments in various empirical contexts without imposing exact identifying assumptions or imposing specific functional forms. In this section, we estimate bounds on the training effect using the experimental data and the bounding strategies originally developed in Robins (1989) and refined in later work by Horowitz and Manski (1995) and Hotz, Mullin and Sanders (1997). 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).[30]

In this approach various moments of the outcome distribution for controls are used in combination with assumptions about the support of the outcome distribution and the training participation process, to obtain upper and lower bounds on the training effect using more general assumptions than those required to obtain point estimates. Thus, this approach represents an alternative to either the intuitive estimator of Section 6 or the more traditional econometric estimators presented below in Section 8. Consistent with much of the empirical literature that estimates nonparametric bounds on treatment effects (*e.g.*, Heckman and Smith, 1993, and Heckman and Smith with Clements, 1997), we find that the bounds are often quite wide. However, in certain cases, they provide information that supports our other evidence that the estimated program effects represent strongly downward-biased estimates of the effect of training.

We seek to obtain bounds on

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

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. Using the law of iterated expectations,

$$E(Y_{0,s} \mid T_s = 1) = E(Y_{0,s} \mid T_s = 1, T_{-1,s} > 1)\Pr(T_s = 1 \mid T_{-1,s} > 1)$$

[30]Heckman (1999) presents a history of these ideas.

$$+ E(Y_{0,s} \mid T_s = 1, T_{-1,s} = 0) \Pr(T_s = 1 \mid T_{-1,s} = 0)$$

where $T_{-1,s} > 1$ for treatment group members who would have initiated alternative training by month s, had they been in the control group, and is zero otherwise.

None of the terms in the decomposition is identified from the experimental data on treatment and control outcomes. However, the conditional probabilities can be bounded by applying the Fréchet (1951) bounds to the observed marginal probabilities $\Pr(T_s = 1)$ and $\Pr(T_{-1,s} > 1)$ in the treatment and control groups, respectively (see Heckman and Smith, 1993; 1997). The conditional mean $E(Y_{0,s} \mid T_s = 1, T_{-1,s} = 0)$ is bounded by assuming that among the controls who did not take training, either those with the highest earnings or those with the lowest earnings would have taken training if in the treatment group and using their earnings to construct the bounds. Finally, the conditional mean $E(Y_{0,s} \mid T_s = 1, T_{-1,s} > 1)$ can be bounded by invoking the following assumption:

> **(AS-5)** 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 $Y^l = 0$.

We obtain upper and lower bounds on the discounted net returns to training by constructing bounds for each month after random assignment using the methods just described, calculating the discounted sum of the upper and lower bounds, respectively, and then subtracting off the discounted difference in direct training costs from each bound. The assumed monthly interest rate for discounting is that appropriate for an annual rate of $r = 0.03$. In order to consider cases where the training effect persists beyond the available data, we assume that the monthly bounds on the training effect persist after month 33 at the mean of the last twelve observable months.

The first rows of each panel of Table 7-1 present estimated upper and lower bounds, respectively, calculated using assumption **(AS-5)**. Standard errors appear in parentheses. Given the weak assumptions required to obtain them, it is not surprising that these bounds are very wide and thus not at all informative. For example, for adult women, the bounds under the assumption that the training effect persists 33 months are -\$48,775 and \$13,000. These bounds easily bracket all of the estimated program and training effects in Tables 5-1 and 6-1.[31]

[31] Heckman and Smith (1993) and Heckman and Smith with Clements (1997) report similar results in their application of the Fréchet bounds.

The bounds on the training effect can be tightened by making additional assumptions. For example, consider the following assumption:

(AS-6) 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, $T_{-1,s} > 1 \Rightarrow T_s = 1$.

This assumption is plausible since persons denied training by randomization are slowed in their quest for training from alternative sources. Under this assumption, if receipt of training is more common in the treatment group, so that $\Pr(T_s = 1) > \Pr(T_{-1,s} > 1)$, then $\Pr(T_s = 1 \mid T_{-1,s} > 1) = 1$ and

$$\Pr(T_s = 1 \mid T_{-1,s} = 0) = \frac{\Pr(T_s = 1) - \Pr(T_{-1,s} > 1)}{\Pr(T_{-1,s} = 0)}.$$

In addition, **(AS-6)** allows the identification of $E(Y_{0,s} \mid T_s = 1, T_{-1,s} = 0)$ because by another application of the law of iterated expectations

$$\begin{aligned} E(Y_{0,s} \mid T_{-1,s} = 0) = {} & E(Y_{0,s} \mid T_s = 1, T_{-1,s} = 0) \Pr(T_s = 1 \mid T_{-1,s} = 0) \\ & + E(Y_{0,s} \mid T_s = 0, T_{-1,s} = 0) \Pr(T_s = 0 \mid T_{-1,s} = 0). \end{aligned}$$

In this equation, the left-hand side is identified from the mean earnings of controls who do not take training and the two probabilities are identified by **(AS-6)**. In addition, because of random assignment, we can identify $E(Y_{0,s} \mid T_s = 0, T_{-1,s} = 0)$ using the mean earnings of the treatment group dropouts. Replacing these terms with their sample analogues allows us to solve for $E(Y_{0,s} \mid T_s = 1, T_{-1,s} = 0)$. For the other counterfactual earnings term, we continue to rely on the bounds implied by **(AS-5)**.

The second row of each panel of Table 7-1 presents estimates obtained using **(AS-5)** and **(AS-6)**. In every case, these bounds are narrower than those that rely only on **(AS-5)**. At the same time, they remain too wide to provide any useful information about the training effect.

A final assumption narrows just the lower bound estimates. If we assume that

(AS-7) The training effect is non-negative following completion of training, then, if the earnings process is stationary, we can use the earnings of the treatment group trainees in the months after training as an upper bound on their earnings in the absence of training. In formal terms, **(AS-7)** plus stationarity of the earnings process together imply

$$E(Y_{0,s} \mid T_s = 1, T_{-1,s} = 1) \leq E(Y_{1,r} \mid T_r = 1), \tag{4}$$

where $T_r = 1$ for persons who complete training prior to month r after random assignment. Note that both **(AS-6)** and **(AS-7)** are consistent with simple models in which individuals have unbiased expectations and participate when the expected benefits of training outweigh the direct and indirect costs.

The final row of each panel of Table 7-1 presents estimates based on assumptions **(AS-6)** and **(AS-7)**. Only the lower bounds differ from those based on **(AS-5)** and **(AS-6)**. In this case, the bounds actually have some bite. For adult women, the lower bounds on the training effect exceed the estimated program effects in Table 5-1 for all three assumptions regarding the duration of the effect. For example, the lower bound in Table 7-1 for the 33 month duration for adult females is \$1133, which exceeds the estimated program effect in Table 5-1 of \$755 (for $r = 0.03$). Thus, 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.

8 Evidence From Alternative Estimation Methods

The dramatic difference between the estimates of R_0 and R_1 presented in Sections 5 and 6 suggests that the unadjusted estimates from the JTPA experiment seriously understate the effect of classroom training. However, the adjusted estimates presented in Section 6 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 4-1 on the characteristics of training in the two groups 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 pre-training earnings of the trainees or the earnings of the dropouts to proxy for what the treatment group trainees would have received had they 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 6 that the unadjusted experimental estimates seriously understate the training effect. In addition, when we apply the methods in this section to estimate the effect of control group training, we obtain estimates of the effect of training 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 which underlies the adjusted experimental estimates reported in Section 6.

8.1 Simple Before-After and 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 have received 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:

$$\Delta_d = E\left[Y_d \mid T = 1\right] - E\left[Y_0 \mid T = 1\right]. \tag{5}$$

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

$$\Delta_a = E\left[Y_a \mid T = 1\right] - E\left[Y_0 \mid T = 1\right]. \tag{6}$$

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 non-trainees, $E[Y_0 | T = 0]$, 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 non-trainees or of trainees prior to training in place of the unobserved counterfactual raises the possibility of selection bias in the estimates of Δ_d and Δ_a. If the monthly earnings of trainees in the absence of training correspond to the monthly earnings of non-trainees conditional on observable characteristics X, or

(AS-8) $E[Y_0 | T = 1, X] = E[Y_0 | T = 0, X]$,

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). This assumption underlies one version of the method of matching (Heckman, 1986, and Heckman, LaLonde and Smith, 1999). Similarly, if, conditional on X, the monthly earnings of trainees prior to training identify what trainees would have earned in the absence of training, or

(AS-9) $E[Y_0 | T = 1, X] = E[Y_p | T = 1, X]$,

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

Table 8-1 presents before-after 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 post-training earnings of trainees and the earnings of non-trainees. 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 post-training earnings and pre-training earnings for trainees. This difference is the before-after estimate of Δ_a. Cross-section and before-after estimates of Δ_d are similarly defined. As we use multiple monthly observations from the same individuals, we report robust standard errors to allow for serial correlation in u.[32]

We estimate Δ_d separately for three groups: persons currently in a 1-4 month spell of classroom training, persons currently in the first 4 months of a spell of 5 or more months, and persons currently in the fifth or later month of a spell of 5 or more months. Both the before-after and cross-section estimates of Δ_d are negative for all groups, but the before-after 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 5-1.

We estimate Δ_a separately for persons who received 1-4 months and 5 or more months of classroom training. The estimates in both cases are generally large, positive and statistically significant.[33] 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.[34]

[32]We also estimated cross-section and before-after estimates in separate regressions in which all person-months were not pooled. Instead, the before-after estimates were estimated using the person-months of trainees only, while the cross-section estimates were estimated using the person-months of all non-trainees and the relevant observations for trainees (person-months in training for Δ_d and person-months after training for Δ_a.) The cross-section estimates were not sensitive to the change in specification. Although the before-after estimates were more sensitive to the change in specification (estimates tended to be higher for adult controls and for male youth in both experimental groups when estimated separately), these differences do not affect our substantive conclusions. The details of these alternative specifications are available on request from the authors.

[33]For males, the effect of training is larger for persons with 5 or more months of training. The reverse is true for females, but the differences are not statistically significant.

[34]The before-after estimates of Δ_a are uniformly higher than the simple cross-section estimates. This suggests that (AS-8) is not valid and that, conditional on X, nontrainee earnings are higher than trainee earnings would

Table 8-2 presents estimates of the net returns to training, R_1, constructed using the before-after and cross-section estimates of the mean effects of training Δ_d and Δ_a from Table 8-1. 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:

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

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)[35], scaled up by a factor of 1.50 to account for the deadweight cost of taxation (see, *e.g.*, Browning, 1987).[36]

The estimates reported in Table 8-2 are very sensitive to assumptions about the persistence of the training effects. Including social costs also has a substantial effect. Cases with positive private returns often have negative net social returns. Generally, the estimated private returns to JTPA

have been had they 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 before-after estimator differences out the bias as in the symmetric differences estimator in Heckman and Robb (1985) but the cross-section estimator does not. It could also be that the before-after estimator is biased upward due to the dip in earnings prior to training. Heckman, Ichimura, Smith and Todd (1998) and Heckman, Ichimura and Todd (1997,1998) discuss alternative methods for accounting for selection bias in non-experimental evaluations. Heckman, Smith and LaLonde (1999) summarize the evidence in this literature.

[35] 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.

[36] 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 deadwight 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.

classroom training are in the neighborhood of a few thousand dollars. They are consistent with the estimated bounds on the effect of training in Table 7-1. They are also close to the adjusted experimental estimates of R_1 reported in Table 6-1, although the two sets of estimates are based on different assumptions. Internal rates of return corresponding to the estimates in Table 8-2 are generally large and positive for private returns calculated using the before-after estimates, small and positive for private returns calculated using the cross-section estimates, and frequently negative for social returns. Table 8-2 shows JTPA classroom training in a strikingly different light than the unadjusted experimental estimates. At the same time, even when private returns are large and positive, 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 restricted our sample to treatment group members. We can also use the members of the control group to estimate the effect of non-JTPA classroom training. Table 8-3 is constructed for the control group in the same manner as Tables 8-1 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)** in Section 6 which justifies a simple adjustment of the experimental estimator to identify the effect of training. Estimates of Δ_a are somewhat higher for controls, which is consistent with the higher average out-of-pocket costs shown in Table 4-1.[37]

8.2 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 19 months after random assignment as the dependent variable in separate regressions.[38]

[37] The analogue to Table 8-2 constructed using control group data is available from the authors upon request.

[38] 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

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. 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 but they use this information in different ways. (Heckman, LaLonde and Smith, 1999, provide an extensive discussion of this point). Method I estimates the mean effect of training conditional on X. Method II estimates the mean effect of training conditional on X and Z. This method relies on the assumption that, conditional on X and Z, any residual variation in outcomes is statistically independent of T. Method III, based on the analysis of Rosenbaum and Rubin (1983), conditions on the estimated probability of participation in the program, or propensity score, to eliminate selection bias in estimates of R_1. 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.[39] 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.[40] It is the only method among those five that allows for selection on unobserved (by the economist) components of gain.

Table 8-4 presents the results of estimating R_1 using each method. Note that these estimates, unlike those in Section 8.1, do not distinguish between the effect of being in training and the effect of having completed training. Rather, like the adjusted estimates in Section 6, they combine the two effects into a net effect of training relative to no training at all.

The estimates in Table 8-4 are generally of the same sign, and of similar magnitude, as those in Table 8-2, although it is important to keep in mind the large standard errors obtained from many

one year after random assignment in order to isolate the effect of training completion for extrapolation to months beyond 19.

[39]Conditioning on X and Z equates selection bias across treatment and comparison groups so it can be removed by differencing across the two groups.

[40]Details on the construction of the various estimators are discussed in the notes to Table 8-4 and in the Appendix. Heckman, Ichimura, Smith and Todd (1998) extend this estimator to a semiparametric setting.

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).[41] 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 so we should not make too much of these results.

Overall, we draw two lessons from the estimates based on traditional cross-section econometric estimators presented in Table 8-4. 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 Section 6 and in Section 8.1 are robust to a wide variety of alternative methods for dealing with post-random-assignment selection into JTPA. In practically 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.

9 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

[41] A number of different sets of instruments were tested for Method IV. Estimates of the training effect using different instrument sets showed the same pattern of large effects relative to the experimental estimates and large standard errors. These estimates are available on request from the authors.

treatment group dropout, in an actual social experiment.[42] 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 sizeable negative effect on monthly earnings of being in 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 taxation and the discount rate. For other components of the JTPA program like publicly-subsidized 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

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

to be reliable guides to public policy.[43]

Our evidence suggests that experimental evaluations cannot be treated as if they automatically produce easily interpreted and valid answers 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.

[43]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).

References

[1] 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, CA: Sage, 1980), 290-317.

[2] 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).

[3] Burtless, Gary, "The Case for Randomized Field Trials in Economic and Policy Research," *Journal of Economic Perspectives*, IX, 1995, 63-84.

[4] Cave, George., Hans. Bos, Fred Doolittle and Cyril Toussaint, *JOBSTART: Final Report on a Program for School Dropouts* (New York, NY: Manpower Demonstration Research Corporation, 1993).

[5] Couch, Kenneth, "New Evidence on the Long-Term Effects of Employment and Training Programs," *Journal of Labor Economics*, X, 1992, 380-388.

[6] Doolittle, Fred and Linda Traeger, *Implementing the National JTPA Study* (New York, NY: Manpower Demonstration Research Corporation, 1990).

[7] Friedlander, Daniel and Gayle Hamilton, *The Saturation Work Initiative Model in San Diego: A Five-Year Follow-up Study* (New York, NY: Manpower Demonstration Research Corporation, 1993).

[8] Glynn, Robert, Nan Laird, Donald Rubin, in "Mixture Modeling vs. Selection Modeling," in Howard Wainer, ed, *Drawing Inferences From Self-Selected Samples*, New York: Springer Verlag, 1986), 115-142.

[9] Gueron, Judith and Edward Pauly, *From Welfare to Work* (New York, NY: Russell Sage Foundation, 1991).

[10] Heckman, James, "Randomization and Social Program Evaluation," in C. Manski and I. Garfinkel, eds., *Evaluating Welfare and Training Programs* (Boston, MA: Harvard University Press, 1992), 201-230.

[11] Heckman, James, "Instrumental Variables: A Study of Implicit Behavioral Assumptions in One Widely Used Estimator," *Journal of Human Resources*, XXXII (1997), 441-462.

[12] Heckman, James, ed., *Performance Standards in a Government Bureaucracy: Analytical Essays on the JTPA Performance Standards System* (Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1998).

[13] Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. "Characterizing Selection Bias Using Experimental Data," *Econometrica,* LXVI (1998), 1017-1098.

[14] 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.

[15] Heckman, James, Hidehiko Ichimura, and Petra Todd, "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, LXV (1998), 261-294.

[16] 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* (Amsterdam: North-Holland, 1999).

[17] 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), 156-243.

[18] Heckman, James and Jeffrey Smith, "Experimental and Nonexperimental Evaluation of Training Programs," unpublished manuscript, University of Chicago, 1992, presented at the ASSA meetings, Anaheim, California, January, 1993.

[19] 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), 35-96.

[20] 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), 37-88.

[21] 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.

[22] Heckman, James, Jeffrey Smith, and Christopher Taber. "Accounting for Dropouts in Evaluations of Social Programs," *Review of Economics and Statistics*, LXXIX, (1998), 1-14.

[23] Heckman, James, Jeffrey Smith and Marybeth Wittekind, "Awareness and Participation in Social Welfare Programs," University of Chicago, unpublished manuscript, 1997.

[24] 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), 149-151.

[25] 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.

[26] Kane, Thomas, "Reconciling Experimental and Non-Experimental Evidence on the Returns to Postsecondary Training," Harvard University, unpublished manuscript, 1994.

[27] Kane, Thomas, "College Entry by Blacks since 1970: The Role of College Costs, Family Background, and the Returns to Education," *Journal of Political Economy*, CII (1994), 878-911.

[28] Kemple, James, Daniel Friedlander and Veronica Fellerath, *Florida's Project Independence: Benefits, Costs and Two-Year Impacts of Florida's JOBS Program* (New York, NY: Manpower Demonstration Research Corporation, 1995).

[29] Mallar, Charles, Stuart Kerachsky and Craig Thorton, "The Short-Term Economic Impact of The Job Corps Program," in Ernest Stromsdorfer and George Farkas, eds, *Evaluation Studies Review Annual*, Vol. 5. (Beverly Hills, CA: Sage Publications, 1980).

[30] Masters, Stanley and Rebecca Maynard, *The Impact of Supported Work on Long-Term Recipients of AFDC Benefits* (New York, NY: Manpower Demonstration Research Corporation, 1981).

[31] Maynard, Rebecca, *The Impact of Supported Work on Young School Dropouts* (New York, NY: Manpower Demonstration Research Corporation, 1980).

[32] National Commission for Employment Policy, *The Job Training Partnership Act* (Washington, DC, U.S. Government Printing Office, 1987).

[33] 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).

[34] 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).

[35] Quint, Janet, Barbara Fink and Sharon Rowser, *New Chance: Interim Findings on a Comprehensive Program for Disadvantaged Mothers and their Children* (New York, NY: Manpower Demonstration Research Corporation, 1994).

[36] 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, D.C.: U.S. Public Health Service, 1989).

[37] Rosenbaum, Paul and Donald Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, LXX (1983), 41-55.

[38] 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.

[39] Woodbury, Stephen and Robert Spiegelman, "Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois," *American Economic Review*, LXXVII (1987), 513-530.

TABLE 1-1

TREATMENT GROUP DROPOUT AND CONTROL GROUP SUBSTITUTION IN EXPERIMENTAL EVALUATIONS OF ACTIVE LABOR MARKET POLICIES

(Fraction of Experimental Treatment and Control Groups Receiving Services)

Study	Authors/Time Period	Target Group(s)	Fraction of Treatments Receiving Services[1]	Fraction of Controls Receiving Services
1. NSW[2]	Hollister, et al. (1984)	Long Term AFDC Women[3]	0.95	0.11
	(9 months after RA[4])	Ex-addicts	NA	0.03
		17 - 20 year old high school dropouts	NA	0.04
2. SWIM	Friedlander and Hamilton (1993)	AFDC Women: Applicants and Recipients		
	(Time period not reported)	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
		c. Classroom Training/OJT	0.34	0.22
		d. Any activity	0.70	0.23
3. JOBSTART	Cave, et al. (1993) (12 months after RA)	Youth High School Dropouts		
		Classroom Training/OJT	0.90	0.26
4. Project Independence	Kemple, et al. (1995) (24 months after RA)	AFDC Women: Applicants and Recipients		
		a. Job Search Assistance	0.43	0.19
		b. Classroom Training/OJT	0.42	0.31
		c. Any activity	0.64	0.40
5. New Chance	Quint, et al. (1994) (18 months after RA)	Teenage Single Mothers		
		Any education services	0.82	0.48
		Any training services	0.26	0.15
		Any education or training	0.87	0.55

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.

[1] Service receipt includes any employment and training services.
[2] The services received by the controls in the NSW study are CETA and WIN jobs.
[3] For the Long Term AFDC Women, this service receipt also includes regular public sector employment during the period.
[4] RA denotes random assignment to treatment or control groups.

TABLE 4-1

CHARACTERISTICS OF CLASSROOM TRAINING (CT)
IN THE 19 MONTHS FOLLOWING RANDOM ASSIGNMENT
Classroom Training Treatment Stream
19 Month Rectangular Sample
16 Experimental Sites

	Adult Men			Adult Women			Male Youth			Female Youth		
	Treatment	Control	Prob(t>|T|)[1]	Treatment	Control	Prob(t>|T|)	Treatment	Control	Prob(t>|T|)	Treatment	Control	Prob(t>|T|)
Total Sample Size	744	325	NA	1,697	816	NA	377	174	NA	734	352	NA
Number Receiving CT	363	89	NA	952	272	NA	210	60	NA	430	141	NA
Percent Receiving CT	48.8%	27.4%	0.0	56.1%	33.3%	0.0	55.7%	34.5%	0.0	58.6%	40.1%	0.0
Characteristics of Training Received[2]												
Average Total Months of Training	6.7	7.6	0.2	7.2	8.0	0.0	7.0	6.4	0.5	6.7	7.0	0.6
Average Total Hours of Training	680.7	699.0	0.8	705.9	779.3	0.2	745.7	661.4	0.4	765.3	585.2	0.0
Average Hours per Month of Training	110.3	93.5	0.0	100.5	91.4	0.0	110.2	109.5	1.0	108.8	87.3	0.0
Fraction of Training Months Employed	50.2%	46.9%	0.8	34.7%	38.1%	0.5	47.8%	59.2%	0.2	32.8%	48.0%	0.0
Percent Paying for Training[3]	16.8%	41.6%	0.0	11.6%	39.0%	0.0	16.7%	48.3%	0.0	13.0%	36.2%	0.0
Average Monthly Payment per Trainee[4]	$209	$358	0.4	$25	$101	0.0	$39	$103	0.0	$43	$226	0.2
Imputed Average Annual Payment per Trainee[5]	$2,507	$4,301	0.4	$302	$1,211	0.0	$468	$1,238	0.0	$522	$2,715	0.2

[1] T-tests are of the null hypothesis that means of the treatment and control samples are equal within demographic groups.
[2] Sample includes those respondents who reported at least one training spell.
[3] Payments for classroom training are the amount the household paid for school or training.
[4] The average over all trainees includes trainees who reported zero private expenditures.
[5] The imputed annual payment is the observed monthly cost of each trainee multiplied by 12.

TABLE 5-1

DISCOUNTED MEAN EARNINGS AND ESTIMATES OF THE DISCOUNTED RETURNS TO THE JTPA PROGRAM R_0[1]

(Based on Estimates of Δ_s)

Classroom Training Treatment Stream

	Adult Males			Adult Females			Youth Males			Youth Females		
r	Duration of Effect			Duration of Effect			Duration of Effect			Duration of Effect		
	33 Months	5 Years	10 Years	33 Months	5 Years	10 Years	33 Months	5 Years	10 Years	33 Months	5 Years	10 Years
	Discounted Earnings[2]			Discounted Earnings			Discounted Earnings			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
	Discounted Returns			Discounted Returns			Discounted Returns			Discounted Returns		
0	1337	3949	9755	787	1002	1481	209	491	1117	269	1704	4892
	(102)	(478)	(1357)	(57)	(300)	(860)	(102)	(409)	(1146)	(58)	(231)	(642)
0.03	1248	3574	8214	755	947	1330	190	441	942	222	1500	4048
	(98)	(430)	(1132)	(54)	(270)	(717)	(97)	(369)	(957)	(55)	(209)	(536)
0.10	1061	2838	5609	687	833	1062	152	343	642	125	1101	2623
	(87)	(338)	(756)	(48)	(211)	(478)	(88)	(291)	(640)	(49)	(165)	(360)
	ESTIMATES OF THE INTERNAL RATES OF RETURN TO JTPA PROGRAM[3] (Based on Estimates of Δ_s)											
	1.17	1.27	1.28	2.63	2.63	2.63	0.66	0.79	0.80	0.21	0.53	0.58

[1] Estimates of R_0 are the present discounted value of the effect of the program, discounted at $1/(1+r)$ where r ranges over 0, 0.03, and 0.10, and based on program effects. Estimates are of private returns and include estimated monthly tuition payments.

[2] Discounted earnings are the present discounted value of mean monthly control group earnings. Monthly earnings beyond the 33 month sample are set at the mean level of months 22 to 33 after random assignment.

[3] 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 private estimates and include the estimated monthly tuition payments.

TABLE 6-1
ESTIMATES OF THE DISCOUNTED RETURNS TO JTPA TRAINING R_l
(Based on Adjustment of Δ_s)
Classroom Training Treatment Stream

r	Adult Males			Adult Females			Youth Males			Youth Females		
	Duration of Effect			Duration of Effect			Duration of Effect			Duration of Effect		
	33 Months	5 Years	10 Years	33 Months	5 Years	10 Years	33 Months	5 Years	10 Years	33 Months	5 Years	10 Years
Adjusted by Incidence of Classroom Training Receipt[1]												
0	5268	17429	44454	3687	4622	6700	847	2203	5218	2929	12186	32756
	(530)	(2737)	(7873)	(251)	(1221)	(3474)	(526)	(2149)	(5980)	(296)	(1511)	(4318)
0.03	4862	15690	37290	3538	4371	6032	754	1962	4371	2608	10850	27292
	(504)	(2460)	(6562)	(238)	(1098)	(2898)	(500)	(1937)	(4995)	(281)	(1359)	(3601)
0.10	4012	12284	25183	3215	3851	4843	564	1487	2926	1947	8243	18061
	(450)	(1925)	(4369)	(211)	(861)	(1933)	(446)	(1526)	(3346)	(248)	(1064)	(2401)
Adjusted by Hours of Classroom Training Receipt[2]												
0	3739	12367	31542	3875	5114	7868	758	2071	4989	1281	6998	19702
	(427)	(1676)	(4654)	(274)	(1501)	(4325)	(467)	(2013)	(5641)	(200)	(959)	(2729)
0.03	3446	11129	26454	3714	4818	7019	670	1839	4172	1088	6178	16332
	(408)	(1512)	(3889)	(259)	(1348)	(3604)	(444)	(1813)	(4710)	(190)	(863)	(2276)
0.10	2831	8700	17852	3366	4209	5523	491	1385	2777	690	4579	10642
	(368)	(1195)	(2608)	(228)	(1053)	(2398)	(395)	(1426)	(3150)	(169)	(677)	(1519)
ESTIMATES OF THE INTERNAL RATES OF RETURN TO JTPA TRAINING (Based on Adjustment of Δ_s)												
Adjusted by Incidence of Classroom Training Receipt												
	0.79	0.96	0.97	2.45	2.45	2.45	0.45	0.62	0.65	0.43	0.70	0.73
Adjusted by Hours of Classroom Training Receipt												
	0.73	0.90	0.91	2.44	2.44	2.44	0.41	0.59	0.62	0.26	0.56	0.61

[1] Estimates based on incidence of classroom training receipt are constructed using estimated monthly program effects (Δ_s) adjusted upwards 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.

[2] Estimates based on hours of classroom training receipt are constructed using estimated monthly program effects (Δ_s) adjusted upwards 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.

TABLE 7-1
ESTIMATED BOUNDS ON THE DISCOUNTED RETURNS TO JTPA TRAINING R_l[1]
(Based on Estimates of Monthly Bounds with $r = 0.03$)
Classroom Training Treatment Stream

Adult Males			Adult Females			Youth Males			Youth Females		
Duration of Effect			Duration of Effect			Duration of Effect			Duration of Effect		
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 on Discounted Private Returns											
(AS-5) Imposed											
-61254	-113194	-216799	-48755	-94469	-185655	-52544	-104590	-208407	-40457	-79469	-157288
(363)[2]	(688)	(1788)	(179)	(359)	(949)	(314)	(565)	(1440)	(167)	(300)	(764)
(AS-5) and (AS-6) Imposed											
-44294	-81262	-155004	-37048	-73848	-147254	-41059	-84791	-172024	-33611	-68192	-137171
(698)	(1271)	(3256)	(296)	(572)	(1497)	(578)	(1002)	(2519)	(281)	(470)	(1162)
(AS-5), (AS-6) and (AS-7) Imposed											
-2214	-942	1595	1133	3002	6729	-1772	-2496	-3940	-1129	96	2540
(800)	(1441)	(3678)	(373)	(707)	(1837)	(670)	(1156)	(2901)	(341)	(581)	(1449)
Upper Bound on Discounted Private Returns											
(AS-5) Imposed											
21824	40864	78842	13007	25155	49388	17160	33047	64736	11208	22186	44084
(321)	(628)	(1648)	(144)	(286)	(753)	(274)	(505)	(1299)	(141)	(261)	(674)
(AS-5) and (AS-6) Imposed											
13015	29303	61795	8255	15924	31222	9609	19532	39325	4902	11916	25909
(698)	(1271)	(3256)	(296)	(572)	(1497)	(578)	(1002)	(2519)	(281)	(470)	(1162)
(AS-5), (AS-6) and (AS-7) Imposed											
13015	29303	61795	8255	15924	31222	9609	19532	39325	4902	11916	25909
(698)	(1271)	(3256)	(296)	(572)	(1497)	(578)	(1002)	(2519)	(281)	(470)	(1162)

[1]Estimates of bounds on R_l are the present put 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 $r = 0.03$.
[2]Standard errors of R_l are derived from the discounted sum of variance estimates on the monthly bounds on the training effect.

TABLE 8-1
NONEXPERIMENTAL ESTIMATES OF THE MONTHLY EFFECTS OF TRAINING Δ_d AND Δ_a
Before-After and Cross-Section Estimates
Dependent Variable: Pooled Monthly Earnings[1]
Classroom Training Treatment Stream
Treatments at the 16 Experimental Sites

Sample Size / Mean Monthly Earnings[2]	Adult Males 19438 777		Adult Females 42943 386		Youth Males 11328 632		Youth Females 22052 347	
	B-A	C-S	B-A	C-S	B-A	C-S	B-A	C-S
Estimates of Δ_d[3]								
Persons With 1-4 Months of Training								
Estimates	-57	-231	-51	-127	-17	-189	-57	-67
	(75)	(48)	(32)	(21)	(93)	(59)	(48)	(32)
Persons With > 4 Months of Training, Currently Completed 1-4 Months								
Estimates	-68	-273	-75	-189	-129	-180	-66	-111
	(59)	(46)	(25)	(17)	(64)	(46)	(29)	(22)
Persons With > 4 Months of Training, Currently Completed > 4 Months								
Estimates	-96	-302	-117	-231	-207	-259	-44	-89
	(86)	(66)	(32)	(24)	(77)	(62)	(38)	(29)
Estimates of Δ_a[4]								
Persons With 1-4 Months of Training								
Estimates	280	105	111	36	169	-3	77	66
	(92)	(55)	(43)	(29)	(107)	(61)	(57)	(38)
Persons With > 4 Months of Training								
Estimates	230	25	215	101	70	18	169	124
	(90)	(60)	(39)	(31)	(90)	(55)	(42)	(33)

[1]The dependent variable in an OLS regression is self-reported monthly earnings. The sample consists of all person-months 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. The top one percent of earnings values are dropped in each month in both the treatment and control groups.

[2]Mean Earnings is the mean level of monthly earnings for the control group members who reported earnings over the full 33-month survey.

[3]Before-After estimates of Δ_d (the effect of training while currently receiving training) are the difference in coefficients on the indicator for currently training and the indicator for months prior to training. Cross-Section estimates are the coefficient for currently training (relative to never receiving training).

[4]Before-After estimates of Δ_a (the effect of training after completion of training) are the difference in coefficients on the indicator for completion of training and the indicator for months prior to training. Cross-Section estimates are the coefficient for completion of training (relative to never receiving training).

TABLE 8-2

NONEXPERIMENTAL ESTIMATES OF THE DISCOUNTED RETURNS TO TRAINING R_t[1]

(Based on Estimates of Δ_d and Δ_a with $r = 0.03$)

Treatment Group

Classroom Training Treatment Stream

	Adult Males			Adult Females			Youth Males			Youth Females		
Training Duration	Duration of Effect			Duration of Effect			Duration of Effect			Duration of Effect		
	33 Months	5 Years	10 Years	33 Months	5 Years	10 Years	33 Months	5 Years	10 Years	33 Months	5 Years	10 Years
Returns Based on Before-After Estimates												
Private Returns[2]												
3 months	7362	13349	24086	2801	5175	9433	4421	8044	14541	1789	3428	6367
	(2529)	(4490)	(8013)	(1194)	(2122)	(3787)	(2962)	(5257)	(9380)	(1572)	(2791)	(4980)
6 months	4964	9889	18720	4601	9202	17451	533	2032	4719	3572	7193	13685
	(2233)	(4158)	(7622)	(972)	(1811)	(3321)	(2216)	(4126)	(7562)	(1046)	(1949)	(3572)
Social Returns[3]												
3 months	2730	8717	19453	-1884	490	4749	-184	3440	9936	-2842	-1203	1736
	(2529)	(4490)	(8013)	(1194)	(2122)	(3787)	(2962)	(5257)	(9380)	(1572)	(2791)	(4980)
6 months	405	5330	14161	-33	4568	12817	-3974	-2476	211	-977	2644	9136
	(2233)	(4158)	(7622)	(972)	(1811)	(3321)	(2216)	(4126)	(7562)	(1046)	(1949)	(3572)
Returns Based on Cross-Section Estimates												
Private Returns												
3 months	2066	4326	8377	510	1272	2639	-804	-860	-959	1464	2874	5403
	(1503)	(2668)	(4761)	(802)	(1425)	(2544)	(1693)	(3004)	(5358)	(1048)	(1861)	(3321)
6 months	-1277	-746	206	1136	3298	7174	-1037	-644	62	2203	4860	9624
	(1496)	(2781)	(5094)	(766)	(1429)	(2620)	(1361)	(2526)	(4625)	(808)	(1506)	(2761)
Social Returns												
3 months	-2566	-307	3744	-4175	-3413	-2046	-5409	-5464	-5563	-3167	-1757	772
	(1503)	(2668)	(4761)	(802)	(1425)	(2544)	(1693)	(3004)	(5358)	(1048)	(1861)	(3321)
6 months	-5837	-5306	-4353	-3498	-1336	2540	-5544	-5151	-4446	-2346	311	5075
	(1496)	(2781)	(5094)	(766)	(1429)	(2620)	(1361)	(2526)	(4625)	(808)	(1506)	(2761)

[1]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. Monthly effects are estimated separately for persons receiving 1-4 and 5 or more months of training.

[2]Private returns include the estimated monthly tuition payments.

[3]Social returns include the estimated marginal costs incurred by the training provider, adjusted upwards by 1.5 to reflect the deadweight cost of taxation.

TABLE 8-3
NONEXPERIMENTAL ESTIMATES OF THE MONTHLY EFFECTS OF TRAINING Δ_d AND Δ_a
Before-After and Cross-Section Estimates
Dependent Variable: Pooled Monthly Earnings[1]
Classroom Training Treatment Stream
Controls at the 16 Experimental Sites

Sample Size	Adult Males 8696		Adult Females 20815		Youth Males 5447		Youth Females 10449	
	B-A	C-S	B-A	C-S	B-A	C-S	B-A	C-S
Estimates of Δ_d[2]								
Persons With 1-4 Months of Training								
Estimates	-283	-215	-107	-149	-93	-66	-138	-227
	(111)	(116)	(50)	(41)	(75)	(98)	(49)	(44)
Persons With > 4 Months of Training, Currently Completed 1-4 Months								
Estimates	-146	-372	-86	-98	-133	-256	-9	-172
	(95)	(64)	(44)	(36)	(101)	(83)	(34)	(38)
Persons With > 4 Months of Training, Currently Completed > 4 Months								
Estimates	-198	-425	-115	-127	10	-113	-91	-254
	(120)	(76)	(58)	(45)	(107)	(93)	(43)	(42)
Estimates of Δ_a[3]								
Persons With 1-4 Months of Training								
Estimates	85	153	95	53	-3	25	47	-42
	(141)	(117)	(72)	(55)	(99)	(70)	(65)	(56)
Persons With > 4 Months of Training								
Estimates	379	152	116	104	211	88	238	74
	(139)	(87)	(72)	(56)	(115)	(105)	(62)	(58)

[1]The dependent variable in an OLS regression is self-reported monthly earnings. The sample consists of all person-months 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. The top one percent of earnings values are dropped in each month in both the treatment and control groups.

[2]Before-After estimates of Δ_d (the effect of training while currently receiving training) are the difference in coefficients on the indicator for currently training and the indicator for months prior to training. Cross-Section estimates are the coefficient for currently training (relative to never receiving training).

[3]Before-After estimates of Δ_a (the effect of training after completion of training) are the difference in coefficients on the indicator for completion of training and the indicator for months prior to training. Cross-Section estimates are the coefficient for completion of training (relative to never receiving training).

TABLE 8-4

NONEXPERIMENTAL ESTIMATES OF THE DISCOUNTED RETURNS TO TRAINING R_j CORRECTED FOR SELECTION[1]

Classroom Training Treatment Stream

Experimental Treatment Group

Adult Males			Adult Females			Youth Males			Youth Females		
Duration of Effect			Duration of Effect			Duration of Effect			Duration of Effect		
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[2,3,4]											
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[5,6]											
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[7]											
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[8]											
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[9]											
17025	28568	51595	9343	20310	42185	1074	1331	1845	4466	9346	19079
(2709)[10]	(6233)	(13846)	(1582)	(3714)	(8279)	(2118)	(4985)	(11119)	(1637)	(3828)	(8528)

[1] 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. The effect of training in each month is estimated through a separate regression. The dependent variable Y in each regression is a person's earnings in that month.

[2] Exogenous regressors X in the earnings equation include indicators for race, marital status, education, site of random assignment, and age. The top one percent of earnings values are excluded.

[3] T is a treatment indicator. T=1 signifies that an individual has participated in training by the month of the regression.

[4] Estimator I is the coefficient on T in an OLS regression of Y on X .

[5] 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.

[6] Estimator II is the coefficient on T in an OLS regression of Y on X and Z.

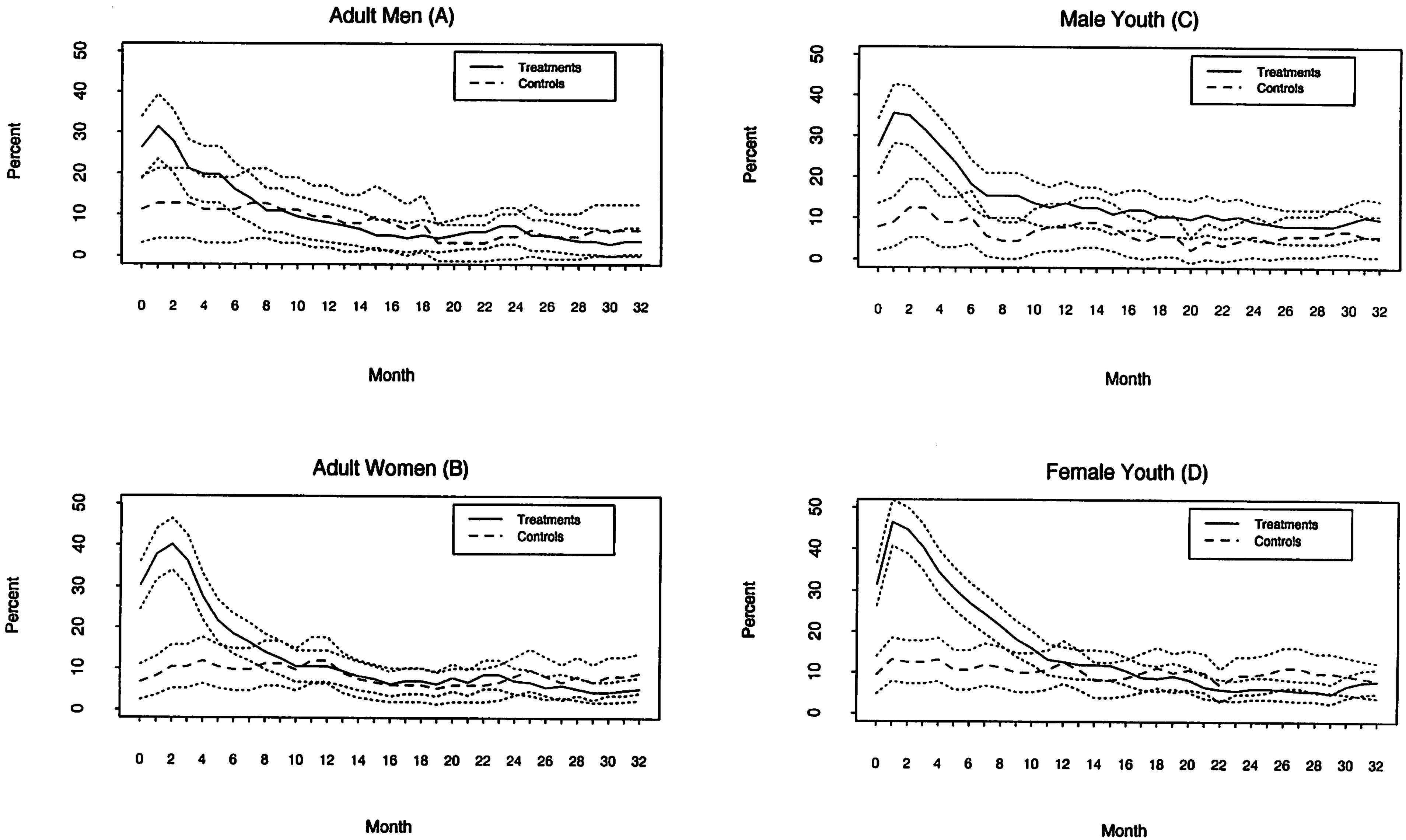
[7] Estimator III is the coefficient on T in an OLS regression of Y on X, P, and T, where P is the predicted probability of participation from a probit of T on Z.

[8] Estimator IV, the 2SLS estimator, 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.

[9] Estimator V, the Heckman two-step estimator, 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.

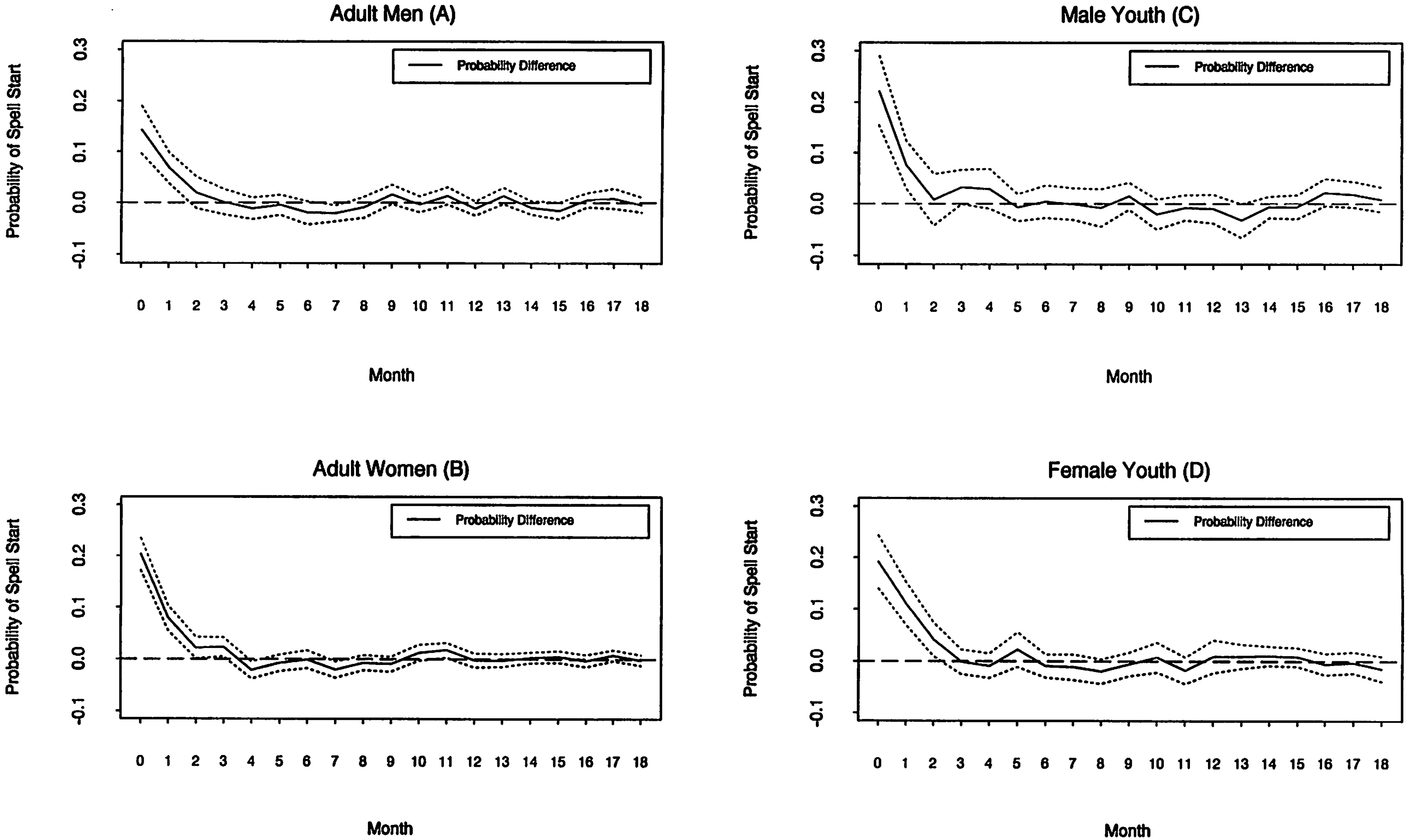
[10] The standard errors are unadjusted for variation introduced by uncertainty in first-stage parameter estimates.A correction was implemented, but led to very large standard errors that were judged to be uninformative.

Figure 4-1
PERCENTAGE RECEIVING CLASSROOM TRAINING
Treatments and Controls at the 16 Experimental Sites
Classroom Training Treatment Stream
33 Month Rectangular Sample



1. Month t=0 is the month of random assignment.
2. The monthly information graphed here is derived from self-reported data on receipt of training.
3. Respondents must not have missing values for receipt of classroom training at any point in the 33 months following random assignment.
4. Standard error bars indicate +/- 2 standard errors about the mean.

Figure 4-2
DIFFERENCES IN PROBABILITY OF SPELL STARTS
Treatments and Controls at the 16 Experimental Sites
Classroom Training Treatment Stream
19 Month Rectangular Sample



1. Differences in probability of spell start are drawn from a series of month-by-month regressions from month 0 to month 18, where t=0 is the month of random assignment. The difference in start rates is indicated by the coefficent on treatment status.
2. The regression model is a linear probability model with an indicator for the start of a new classroom training spell as the dependent variable. Regressors include an indicator for treatment group membership, lagged employment and training status, demographic and training center characteristics, and pre-random assignment training experiences.
3. Standard error bars indicate +/- 2 Eicher-White robust standard errors about the mean.

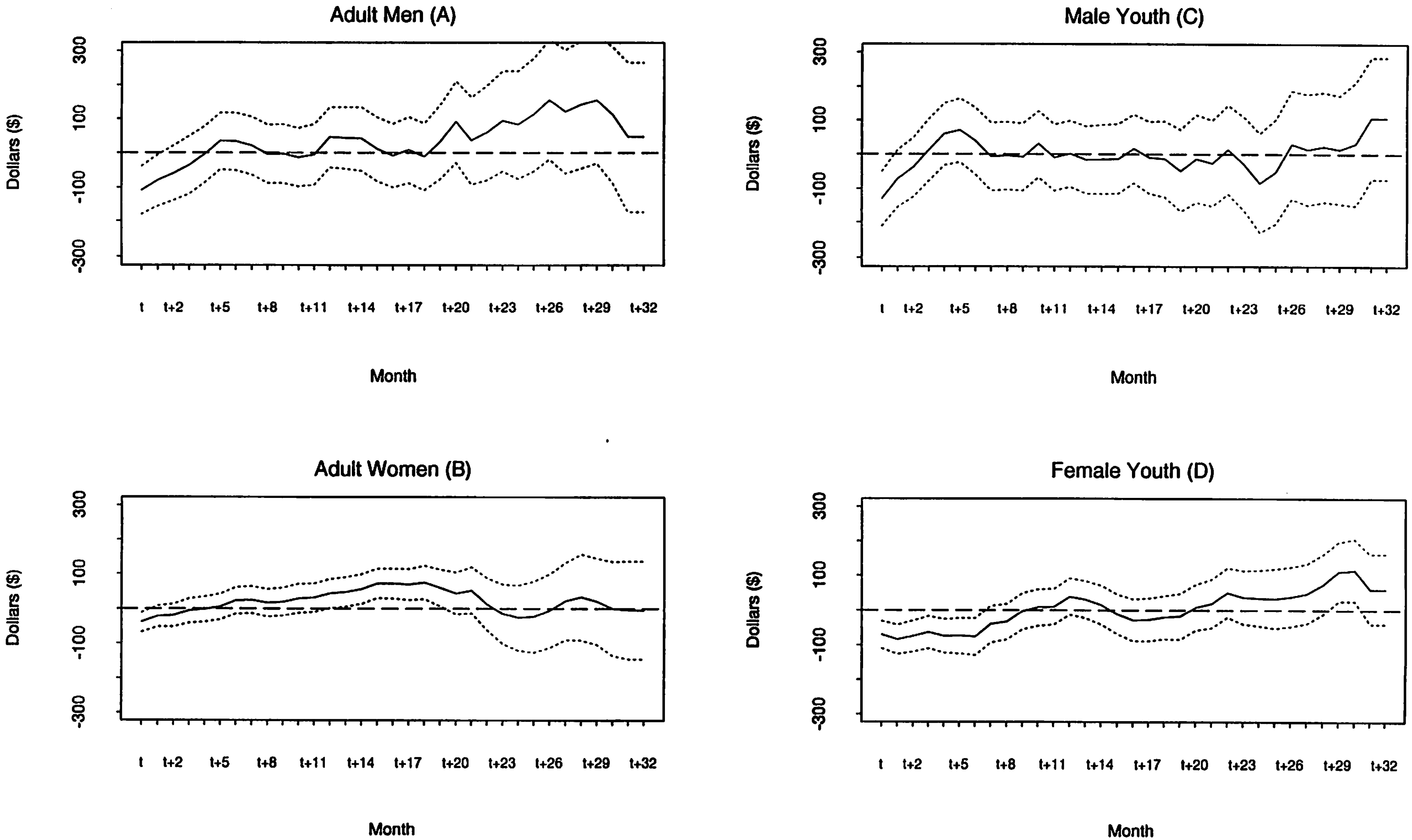
Figure 5-1

EXPERIMENTAL ESTIMATES OF THE MONTHLY EFFECT OF JTPA PROGRAM

Dependent Variable: Pooled Monthly Earnings

Classroom Training Treatment Stream

All Available Person-Months



1. The dependent variable in an OLS regression is self-reported monthly earnings. The sample consists of all person-months in the 33 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 one percent of earnings values are dropped in each month in both the treatment and control groups.
2. Experimental estimates are the sum of the coefficient on treatment status and the coefficients on the interactions between treatment status and month after random assignment.
3. Standard error bars indicate +/- 2 Eicher-White robust standard errors about the mean.

1 Appendix

This appendix consists of two subsections. The first provides information on the data we use from the NJS and on the construction and characteristics of our analysis sample. The second describes in detail the procedure used to construct each table and figure in the paper.

1.1 The NJS Data and Our Analysis Sample

Between November, 1987, and September, 1989, 20601 applicants accepted into JTPA ($D = 1$) at one of the 16 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 18 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 17689 persons reported data during the follow-up period (12069 treatments and 5620 controls). Data from the follow-up surveys on earnings, employment and other outcomes was converted by 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 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 17689 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, 2 year college, 4 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.

Table A-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 non-missing values for all of the demographic characteristics displayed.

1.2 Construction of the Tables and Figures

Table 1-1 summarizes a more comprehensive table in Heckman, LaLonde and Smith (1999). Table 4-1 is based on the self-reported data on training receipt from the follow-up surveys. The table includes only persons who did not report missing values for monthly receipt of classroom training, hours per month in classroom training, amount paid for classroom training, and monthly employment over the 19 months after the month of random assignment.

Figure 4-1 is based on the monthly self-reported training receipt data from the follow-up surveys. It includes only persons with valid data on training receipt for all 33 months. This rectangularity restriction reduces the sample sizes to 137 treatments and 63 controls for adult males, 254 treatments and 133 controls for adult females, 175 treatments and 98 controls for male youth and 324 treatments and 159 controls for female youth.

Figure 4-2 plots the results obtained by estimating 19 monthly linear probability models for each demographic group for the month of random assignment and the 18 following months. The dependent variable is an indicator for the start of a new classroom training spell, and the independent variables include demographic characteristics (from the BIF data), lagged employment and training variables (from the monthly data), training center indicators, and a treatment group indicator. The figure plots the coefficients on the treatment group indicator for each group. To be included in the sample for these regressions, persons must have valid data on classroom training receipt and employment for all 19 months. These restrictions result in sample sizes of 916 adult men, 2022 adult women, 483 male youth, and 892 female youth.

Figure 5-1 plots estimates of Δ_s based on OLS regressions for each demographic group of self-reported monthly earnings on treatment status, month after random assignment, the interaction of treatment status and month after random assignment, calendar month, and various demographic characteristics. The estimate of Δ_s for month s is the sum of the coefficient on treatment status and the coefficient on the interaction of treatment status and month s after random assignment. Person-months are the units of observation in these regressions. Person-months with earnings in the top 1 percent of each random assignment $*$ demographic group $*$ month s cell are dropped from the sample, as are those with missing values for earnings or demographic characteristics. The final sample sizes are 28134 adult men, 63758 adult women, 16775 male youth, and 32501 female youth. Standard errors are robust Eicker-White estimates.

The first half of Table 5-1 presents the estimated returns to JTPA access, R_0, discounted at the constant rate $\delta = 1/(1+r)$ with $r = 0.03$ although we present values for $r = 0$ and $r = 0.10$. The lower half provides the estimated internal rates of return which, when used as the value for r, reduce to zero the estimated returns. Estimates are calculated separately for different assumptions regarding the duration of the program effect (33 months, 5 years, or 10 years). Table 5-1 is based on the estimates of Δ_s plotted in Figure 4-1 and on self-reported tuition payments. Costs in months more than 33 months after random assignment are assumed to be zero. Returns are discounted continuously as shown:

$$R_0 = \sum_{s=0}^{32}\left(\int_s^{s+1} e^{-rt}(\Delta_s - \mu_s)\, dt\right) + \int_{33}^{M} e^{-rt}\overline{\Delta}_s\, dt$$

where $s = 0$ is the month of random assignment, μ_s is defined in Section 5 of the main text, M is the expected duration of the program effect (in months), and $\overline{\Delta}_s$ is the mean of Δ_s over months 20 to 32 after random assignment. Estimates of the internal rates of return indicate the value of r which resolves R_0 to zero.

Table 6-1 relies on adjustments to the estimates of Δ_s and μ_s to estimate the returns to JTPA training, R_1, discounted at the constant rate $\delta = 1/(1+r)$ with $r = 0.03$, $r = 0.10$ and $r = 0$. We focus on the estimates for $r = 0.03$. Also reported in Table 6-1 are the estimated internal rates of return which, when used as the value for r, reduce to zero the estimated returns. Estimates are calculated separately for different assumptions regarding the duration of training effect (33 months, 5 years, or 10 years). The method for calculating R_1 is identical to that for R_0 after substituting the adjusted values of $\Delta_s - \mu_s$ and $\overline{\Delta}_s$. Internal rates of return are estimated similarly.

Table 7-1 presents estimated bounds on the training effect under assumptions described in the text. Person-months with earnings in the top 1 percent of each demographic group * month s cell are dropped from the sample, as are those with missing values for earnings. The remaining sample includes 102749 person-months from the experimental treatment group and 48173 person-months from the experimental control group. Bounds on the returns are constructed for the discounted stream of individual monthly bounds. Let $\overline{v}_s$ be the estimated upper bound on for month s for any combination of the above assumptions. Then an upper bound $\overline{R}_1$ on discounted return to training is

$$\overline{R}_1 = \sum_{s=0}^{32}\left(\int_s^{s+1} e^{-rt}\overline{v}_t dt\right) + \int_{33}^{M} e^{-rt}\overline{\overline{v}}_t dt,$$

where $s = 0$ is the month of random assignment, M is the expected duration of the program effect (in months), and $\overline{\overline{v}}_s$ is the mean of $\overline{v}_s$ over months 20 to 32 after random assignment. The standard error of $\overline{R}_1$ is constructed by discounting the standard errors of $\overline{v}_s$ and assuming $\overline{\overline{v}}_s$ to have a standard error of the mean. We calculate the right hand side of (7.1) using all person months r after training.

Table 8-1 presents the results of the nonexperimental strategies discussed in Section 8 of the main text. These estimates are taken from an OLS regression for each demographic group of self-reported monthly earnings on indicators for training status, calendar month, month after random assignment, and various demographic characteristics using person-months as the unit of observation. Only person-month observations from the experimental treatment group are included in the estimation. Person-months with earnings in the top 1 percent of each demographic group $*$ month s cell are dropped from the sample, as are those with missing values for earnings or for one of the regressors. These restrictions reduce the sample from the initial count of 154.118 person-months in the classroom training treatment stream to the levels reported in Table 8-1. Separate sets of training status indicators are used to estimate effects for training spells of 1 to 4 months or more than 5 months in duration. Standard errors for these estimates are constructed using robust Eicker-White estimates of the variance-covariance matrix.

Table 8-2 presents the estimated returns to JTPA training, R_1, discounted at the constant rate $\delta = 1/(1+r)$ with $r = 0.03$. These estimates are based on the values of Δ_d and Δ_a from Table 8-1. Estimates are calculated separately for different assumptions regarding the duration of the program effect (33 months, 5 years, or 10 years). Separate estimates are presented for training spells:of 3 and 6 months duration. In calculating these returns, trainees are assumed to begin training during the month of random assignment and to train continuously for 3 or 6 months. Returns are discounted continuously. Private returns are calculated according to

$$R_1 = \int_0^m e^{-rt}(\Delta_d - E(c \mid D=1, T=1))\, dt + \int_{m+1}^{M} e^{-rt}\Delta_a\, dt$$

where m is the total duration of training (3 or 6 months), and M is the expected duration of the training effect (in months). Note that in estimating R_1 for trainees with 6 months of training, the first integral must be split to allow two values for Δ_d and c. Social returns are calculated by subtracting from the private returns the estimated marginal costs incurred by the training provider ($3183) adjusted upwards by 1.5 to reflect the deadweight cost of taxation. Standard errors of the returns are similarly discounted and are constructed using the standard error estimates from Table 8-1.

Table 8-3 is constructed identically to Table 8-1, except that the control group is substituted for the treatment group. The corresponding sample sizes are reported at the top of Table 8-3.

Table 8-4 is based on monthly earnings regressions for the 19 months after random assignment among members of the experimental treatment group. This restriction is taken because the severe attrition in observations for the second follow-up survey makes results for later months extremely sensitive to sample composition. In order to isolate the effect of training completion, the sample is further restricted to those who either receive no training or those who complete their training within 12 months of random assignment. Person-months with earnings in the top 1 percent of each demographic group $*$ month s cell are dropped from the sample, as are those with missing values for earnings or for one of the regressors. Monthly regressions are performed according to the methods described in the text above and a selection-corrected monthly estimate of training effect $\tilde{v}_t$ for each method is derived for months 0 through 18. For succeeding months, the training effect $\tilde{\tilde{v}}_t$ is taken to be the mean of $\tilde{v}_t$ for the months 13-18, by which time any training is completed. The associated 33 month, 5 year and ten year returns are calculated by

$$\tilde{R}_1 = \sum_{s=0}^{12}(\int_s^{s+1} e_t^{-rt}\tilde{v}_t dt) + \int_{13}^{M} e^{-rt}\tilde{\tilde{v}}_t dt,$$

where M is the expected duration of the training effect (in months). Standard errors for returns are based on the computed standard errors of the monthly training effects.

TABLE A-1

DEMOGRAPHIC CHARACTERISTICS OF TRAINEES AND NON-TRAINEES
Classroom Training Treatment Stream
19 Month Rectangular Sample
16 Experimental Sites

	Adult Male Trainees			Adult Male Non-trainees			Adult Female Trainees			Adult Female Non-Trainees		
	Treatments	Controls	Prob(t>|T|)[1]	Treatments	Controls	Prob(t>|T|)	Treatments	Controls	Prob(t>|T|)	Treatments	Controls	Prob(t>|T|)
Total Sample Size	363	89	NA	381	236	NA	952	272	NA	745	544	NA
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 Non-trainees**			**Female Youth Trainees**			**Female Youth Non-Trainees**		
	Treatments	Controls	Prob(t>|T|)	Treatments	Controls	Prob(t>|T|)	Treatments	Controls	Prob(t>|T|)	Treatments	Controls	Prob(t>|T|)
Total Sample Size	210	60	NA	167	114	NA	430	141	NA	304	211	NA
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

[1]T-tests are of the null hypothesis that means of the treatment and control samples are equal within demographic and training groups.