

**Using Data on Applicants to Training Programs to Measure
the Program's Effect on Earnings**

Glen Cain
Department of Economics
Institute for Research on Poverty
University of Wisconsin-Madison

Steven Bell
Abt Associates, Inc.
Bethesda, Maryland

Larry Orr
Abt Associates, Inc.
Bethesda, Maryland

Winston Lin
Abt Associates, Inc.
Bethesda, Maryland

August 1993

The research was supported by the W. E. Upjohn Institute for Employment Research, to whom we are most grateful. We are also grateful to Michael Tilkin for research assistance, to the Internal Revenue Service for data, and to Thomas Fraker and James Walker for helpful comments on an earlier draft.

Abstract

Using data from the AFDC Homemaker-Home Health Aide Demonstration, a welfare training program which had an experimental design, a method of evaluation is developed for possible application in a nonexperimental setting, where random assignment is not available. Two techniques to control for selection bias are pursued. One uses applicants as the source of one or more comparison groups, which effectively controls for the self-selection process of volunteering. The other uses the subjective ranking of suitability of applicants by program administrators; although crudely measured, this variable takes into account administrative selection procedures. The subjective ranking is added to the more conventional list of independent variables. Two nonexperimental applicant groups, those who were screened out and those who dropped out, appear sufficiently similar to the control group in predicted earnings in the two postprogram years to indicate that the procedure has promise. IRS data are used to measure earnings in the years that preceded, overlapped, and followed the training program.

Using Data on Applicants to Training Programs to Measure the Program's Effect on Earnings

I. DEALING WITH SELECTION BIAS IN NONEXPERIMENTAL EVALUATIONS

The main purpose of this paper is to propose and examine a method of evaluation research in a nonexperimental setting, in which the advantage of random assignment to treatment and control groups is not available. Another purpose is to use earnings data from the Internal Revenue Service to measure the program's impact over a longer postprogram period than is commonly available with survey data. The IRS data, in conjunction with data from a previous evaluation of the training program, permit measures of the impact of the program using nonexperimental comparison groups as alternatives to a randomly assigned control group.

In nonexperimental evaluations it is widely recognized that persons in comparison groups are likely to differ from program participants in unobserved ways that influence earnings, as a result of the process used to select participants in the program. Although labor economists have applied increasingly sophisticated econometric methods to deal with this problem of selectivity bias, there remains no agreement on which nonexperimental methods, if any, yield valid evaluations.¹

Our proposed method is to model the selection process to take account of the systematic factors that determine who receives the treatment and who does not. By including the variables that systematically determine selection in the model that estimates the treatment effect on the outcome, all excluded variables may be assumed to be, on average, independent of the treatment variable. If the assumption holds, the estimate of the treatment effect should be unbiased.

Our application of the method uses data from an experimentally designed evaluation of the AFDC Homemaker-Home Health Aide Demonstration, a training program for women who were AFDC recipients. Information was collected on characteristics of 11,000 program applicants. The

women who were selected for the program were trained as homemaker-home health aides to infirm elderly persons in demonstration projects in seven states in the period from 1983 to 1986.²

Selection bias has two major sources, self-selection and administrative selection. We suggest below several reasons why an evaluation that uses as a comparison group applicants who did not subsequently participate in the training program may reduce the selection bias relative to a comparison group taken from an external data source, such as the Current Population Survey of the Bureau of the Census.

In our study the women who applied for training were, as are most applicants, self-selected. Thus, those who did not participate in the training program had in common with the trainees this attribute of volunteering. In some cases volunteering may be a response to particular earnings and employment experiences of the applicants. For example, it has been widely documented that the average earnings of trainees tend to show a "pre-program dip" in the survey year before the training program begins. (See Ashenfelter and Card, 1985, and the studies they cite.) Another similarity among the applicants is their location in the same local labor market.

Comparisons between program participants and persons from external data sources are vulnerable to the criticism that particular circumstances of the time and place of the program's setting and the particular personal characteristics of the participants will cause the participants to have quite different postprogram labor market experiences from those of a comparison group chosen from an external source. The information available in survey sources is unlikely to permit matching on these particularities.

Selection from the applicant group into the program, which is usually determined by program administrators, is a second source of selection bias. In our data a subjectively determined ranking of the applicants in terms of their "suitability" or "potential" for succeeding in the training program was used by program administrators as a criterion for selecting applicants into the training program.³ This

ranking variable offers an unusual opportunity to model the administrator's selection process in a way that accounts for the subjective factors that determine selection in addition to the use of various objective measures that are conventionally available in survey data.

Applicants who were screened as eligible for the training program were randomly assigned to a treatment (trainee) group and a control group. About 16 percent of those selected for the treatment group did not attend, a group we label as No Shows. Those who entered the program are called Participants. The program consisted of four to eight weeks of training and up to a year of subsidized work experience providing home care to elderly patients.

The random selection process for Controls requires that their outcomes (primarily employment and earnings) be compared to the experiences of the combined group of Participants and No Shows. Although the participants in training programs are ultimately the group of interest, there is no accepted way of matching the Controls with the Participants. Expressed differently, it would be necessary to model the selection process that distinguished the No Shows from the Participants to be able to interpret the Controls as a valid comparison group for the Participants. We do not, however, have specific information about the self-selection that determined the decision of the No Shows to pass up the training. Arguments can be made either for negative selection, such as having low motivation, or for positive selection, such as having found a preferred job. As we show below, the mean values of the subjective rankings and of a variety of objective characteristics among the No Shows, Participants, and Controls are similar.

The remaining applicants in the sample fall into three groups: those who were rejected by the screening process (Screen Outs), those who dropped out before the screening process was completed, and a relatively small number whom we termed "not elsewhere classified" (NEC). In this paper we combine the latter two groups and label them Drop Outs. (Since the NEC group was not selected for the experiment and has no record of being screened out, we presume that they dropped out before an

administrative decision about their status was made.) Clearly, it is the Screen Outs for whom the subjective ranking of their suitability for the program provides a direct measure, although not a full measure, of the determinants of selection. In contrast, we have no specific information about why the residual group dropped out, and, like the No Shows, their reasons for dropping out may reflect negative or positive selection factors.

II. THE DATA AND THE REGRESSION MODEL FOR ESTIMATION

The program outcome of interest (dependent variable) for this analysis is annual earnings, obtained from the Internal Revenue Service for the years 1984 to 1988. These years cover most of the period of the training program, which, given the different timings of admission periods across the seven states, span the last months in 1983 and the first months in 1986. Importantly, the IRS data for 1987 and 1988 give us two full years of postprogram experience for all trainees. To protect the confidentiality of the data, the IRS provides mean earnings (and standard deviations) for grouped data consisting of a minimum of 10 persons. The individual records from the training project were linked to the IRS data by social security numbers.

For reasons of convenience in data processing, our grouped data, 1,006 units, contain between 10 and 19 persons, and the average group size is 11.1. All groups are homogeneous with respect to the six applicant groups (including those who were "not elsewhere classified"), and most are homogeneous with respect to race (Hispanic-White, Black, and White) and the subjective ranking variable. Four values of the subjective rank of an applicant were used by the program administrators: 1 = most suitable for acceptance, 2 = suitable, 3 = less suitable, and 4 = least suitable. Homogeneity of a characteristic within a group maximizes the variance in the mean values of that characteristic across all 1,006 groups, which will increase the reliability of the estimated effect (coefficient) of that characteristic on the earnings outcome.

The IRS earnings are the only earnings information we have for the two nonexperimental groups, Screen Outs and Drop Outs. For the experimental groups we determined that the IRS earnings and the earnings from the survey carried out for the evaluation of the experiment are quite similar for periods of overlap. For 1985, the most complete year for an overlap between the two data sources, the mean survey earnings exceeded the mean IRS earnings by 2 percent.

A notable finding in our study is that the IRS is a low-cost source of earnings data for long periods of postprogram experience. Inadequate lengths of posttraining experiences have been a serious weakness in both experimental and nonexperimental evaluation studies.

The estimation model we use for evaluation has the following general form:

$$\text{EARNNS} = A + \underline{X}\underline{B} + \underline{Z}\underline{C} + DS + u,$$

where EARNNS is the mean earnings of the grouped data for the applicants; \underline{X} is a vector of "objective" variables that represent productivity or earnings traits (and possible selection traits); \underline{Z} is a vector of five applicant categories defined as dummy variables: Participants, Controls, No Shows, Screen Outs, and Drop Outs, all defined above; S is the subjective ranking of the applicant; A, \underline{B} , \underline{C} , and D are parameters to be estimated; and u is an assumed well-behaved error term. Note that the variables EARNNS, X, Z, and S are all mean values for the data groups of 10 to 19 applicants. (The Z values are, however, all equal to 1 or 0.)

Table 1 shows the sample sizes for the five applicant groups in terms of the number of persons and in terms of the number of data groups that give the mean values for the IRS-reported earnings and other variables. The full sample for estimating the EARNNS model consists of the 1,006 observations for the data groups.

Two features of the nonexperimental data limit the effective size of the sample for testing comparison groups for a nonexperimental evaluation. First, there are only 61 No-Show groups, 84

TABLE 1
Sizes of Applicant Groups

Applicant Group	Number of Persons	Number of Data Groups			All Cohorts
		Cohort 1 1983	Cohort 2 Jan. 1984-June 1984	Cohort 3 July 1984-May 1985	
Participant	3,912	124	90	145	359
Control	4,625	147	107	166	420
No Show	725	21	16	24	61
Screen Out	931	--	--	84	84
Drop Out ^a	909	--	--	82	82
Total	11,102	292	213	501	1006

^aOf the 82 dropout groups, 63 were originally classified as "drop-outs" and 19 as "not elsewhere classified (NEC)".

Screen-Out groups, and 82 Drop-Out groups, compared to 420 Control groups and 359 Participant groups.⁴ Second, all data for the Screen Outs and Drop Outs come from the last of the three intake periods for the program, July 1984 to May 1985. (This group of applicants will be referred to as Cohort 3.) None are from the 1983 (Cohort 1) or January-June 1984 (Cohort 2) periods, which yielded cohorts of applicants who had considerably higher earnings than those in Cohort 3. Because the period of application is another characteristic of the selection process, we will confine most of our analysis to Cohort 3.⁵

Table 2 shows selected descriptive statistics of the applicant groups, based on the 1,006 observations of data groups. The Participants and Controls are similar in education, race, and subjective rank scores. The Controls would be shown to be even more similar to the combined group of Participants and No Shows, because the screened-in applicants were randomly assigned to either the combined groups or to the Control group. The Screen Outs and the Drop Outs have the lowest average educational attainment and the lowest average subjective rank, recognizing that rank 1 is most suitable. (Mean values of all the measured characteristics of the applicant groups are available from the authors.)

The results from estimating the EARNNS model are addressed to three main questions. First, which among the three categories--No Shows, Screen Outs, and Drop Outs--works best as an alternative to the experimental control group in estimating the earnings impact of the training program? In simplest terms, does the postprogram earnings record of any of the alternative groups so closely match that of the Controls that it shows promise as a comparison group in a nonexperimental design? Second, to what extent does the addition of the subjective ranking variable, S, improve the comparison with the Controls of these alternative groups? Third, what new information about the earnings differentials between the trainees and the other groups, including the Controls, is provided by the IRS data covering the extended postprogram period?

TABLE 2**Selected Characteristics of Applicant Groups**

Applicant Group	Cohort 3				All Cohorts					
	% Education < 12 ^a	% White	% Black	Average Subjective Rank ^b	% Education < 12 ^a	% White	% Black	Average Subjective Rank ^b	% Subjective Rank Unreported ^c	% Subjective Rank = 1, 2 ^d
Participant	44	28	56	2.3	44	26	60	2.2	8	80
Control	46	28	55	2.3	45	25	59	2.2	9	78
No Show	50	17	58	2.3	47	17	54	2.3	10	77
Screen Out	50	38	51	3.1	50	38	51	3.1	36	31
Drop Out ^e	52	30	53	3.0	52	30	53	3.0	63	22

^aPercent with less than a high school degree.

^bAverage score of subjective ranking by administrators for suitability for program: 1 is most suitable; 5 is least; 3 is the number assigned to unreported rankings.

^cPercent with subjective rank unreported.

^dPercent with highest rankings (= 1, 2).

^eOf the 82 dropout groups, 63 were originally classified as "drop-outs" and 19 as "not elsewhere classified (NEC)".

Note that *S*, the subjective rank, may be presumed to have been a key determinant of selection into the Screen Out group, but *S* does not directly determine selection into the No Shows or Drop Outs. Again, we do not have explicit information on the selection process for these two groups. Although our empirical results reported below show that the postprogram earnings of the Controls and the No Shows are similar, we do not have a strong a priori theoretical justification for these results, which may merely reflect this particular program or sample. By contrast, similarity in postprogram earnings between the Controls and Screen Outs, in conjunction with success in using the *S* variable, has greater practical significance because of its strong theoretical rationalization; specifically, that modeling the selection process is a valid and practical method for obtaining unbiased estimates of treatment effects in a nonexperimental design.

The model above is estimated for each of five years, 1984 to 1988. Only the last two years, 1987 and 1988, are, however, strict tests of the Participants' earnings performance in the posttraining, unsubsidized, period. As will be shown below, earnings of the trainees are substantially higher in the early years, but these results are not necessarily a measure of enhanced earnings capacities in an unsubsidized labor market. We also show that these early years are indeed a period of transitorily low earnings for the groups that did not enter the training program.

The modest sample size of data groups for No Shows, Screen Outs, and Drop Outs is one limitation of our estimations. Another is the weaknesses in a key variable, *S*, measuring the subjective rank of an applicant. First, the metric for *S* is limited to only four values, 1 to 4. Clearly, a wider variation would be desirable. Note that a two-value rating designating acceptance or rejection would be useless if it were strictly adhered to. Second, the rating is not recorded for many applicants. In our data a missing rating is scored 3, after setting the old ratings of 3 and 4 to equal 4 and 5. A score of 3 effectively assigns a middle (or "neutral") value to a missing observation. The *S* variable is thus measured with error. To the extent that the error is random, the explanatory power and true

coefficient of S in the EARNNS model are biased down. As shown in Table 2, the problem of missing values is much worse among the categories for nonexperimental comparison groups. In the three categories that were screened in--Participants, Controls, and No Shows--less than 10 percent of persons have missing values, while the Screen Outs have missing values in 36 percent of the cases, and 63 percent of the Drop Outs have missing values for S .

A final shortcoming of the S variable, which is not unexpected, is the concentration of low values (designating suitability) for Participants, Controls, and No Shows, and the concentration of relatively high values of S (not suitable) for Screen Outs and Drop Outs. About 79 percent of the group units for Participants, Controls, and No Shows have a mean value of 1 or 2, but only 27 percent of Screen Outs and Drop Outs have these low values. It is clear that these data, useful as they are for providing some information about modeling selection, hardly give the method of subjective evaluation a fair test of its role in modeling selection. Since the original evaluation design for this program was that of a controlled experiment, it is not surprising that obtaining S ratings for the Screen Outs was not emphasized. The scarcity of ratings for the Drop Outs reflects the fact that limited contact was obtained with this group.

III. ESTIMATION RESULTS THAT TEST FOR ALTERNATIVE COMPARISON GROUPS: TABLES 3 TO 7

Tables 3-7 contain the main results for examining our three objectives: (1) using applicants to obtain a valid comparison group in nonexperimental evaluations of training programs; (2) modeling the program administrators' selection of program participants from among the pool of applicants; specifically, using the administrators' subjective rankings to eliminate selection bias in estimating the effect of the program; and (3) using IRS data on earnings to measure the long-term impact of the program. We argue below that some degree of success is obtained in each of the three objectives, and

that there is useful information in the results even if one were pessimistic about using applicant groups in a nonexperimental evaluation.

Tables 3-7 show selected regression models that estimate IRS-recorded annual earnings for each of the five years, 1984 to 1988. The sample for these regressions consists of only Cohort 3, those who applied for the program between July 1984 and May 1985. As noted above, this is the only cohort period with data for Screen Outs and Drop Outs. Thus, 1984 is a year before the training program began for most of the sample. In 1985 some portion of time was spent in training and in both subsidized and unsubsidized employment for some of the Participants. In 1986 most of the trainees were in the unsubsidized labor market for most of the year, although some who entered the program as late as May 1985 were in subsidized jobs in the first part of 1986. The years 1987 and 1988 were a period of unsubsidized employment for all the trainees.

The outcomes for 1987 and 1988 provide the sharpest focus on the three objectives posed above, but several comments about the results for all five years are useful. The annual earnings of these women are very low indeed, especially in the early years. In 1984, when all the women in the sample were recipients of AFDC for at least part of the year, the average earnings reported to the IRS were only \$776, and the average for the Control group was \$749. In the next four years earnings rose sharply for every applicant group. Among the Controls average earnings rose to \$1,495 in 1985, \$2,473 in 1986, \$3,139 in 1987, and \$3,896 in 1988. The figures for No Shows, Screen Outs and Drop Outs are similar. Both the distinction between transitory and permanent income levels and the impermanence of AFDC status are evident. The overall low level of earnings and the trend in earnings reveal the sharply depressed economic situation of these women at or near the time of applying for the program. Matching the trainee group from sources other than the applicants would be difficult.

The average earnings reported for 1984 in Table 3 are considerably higher for Participants than for Controls, \$898 compared to \$749. Even the weighted average of the 1984 earnings for Participants and No Shows, with No Shows as 15.5 percent of the combined group, is \$869, which is \$120 (or 16 percent) higher than the mean for the Controls. Random assignment should have made the preprogram earnings for Controls and the combined Participants plus No Shows similar. The reason for the difference is that some of the Participants benefited from subsidized employment in the last several months of 1984. Thus, 1984 earnings are not a "pure" preprogram measure of earnings for the Participants.

Column (1) in each table lists the five applicant groups. In the regression models, dummy variables designate the groups, with Controls as the omitted group. Column (2) reports the mean earnings for each of the five applicant groups in the year for each table. Column (3) reports the coefficients of the four dummy variables for the applicant groups in a regression with no other independent variables. This is a baseline regression that gives the same information as column (2), but (3) provides a comparison for how the annual earnings of the five applicant groups change in the regression models shown in columns (4) to (6), when various sets of explanatory variables are included. (No column with 1984 earnings as an independent variable is listed in Table 3, because 1984 earnings form the dependent variable in this table.)

By specifying Controls as the omitted group in the regressions, we interpret each coefficient of the other groups as the group's earnings differential relative to the Controls. Thus, the relatively large negative coefficient for Screen Outs in column (3) in Tables 3-7 (for each year) shows consistent evidence of "creaming" by the program administrators--that is, selecting (relatively) high-earnings women for the chance to enter the training program and rejecting low-earnings women.

Column (4) shows the coefficients of the dummy variables for the applicant groups in a regression that includes a number of worker characteristics that are typical of those used to estimate

TABLE 3

**Regression Results for Cohort 3 for 1984, One Year before the Training
Program Started for Most Participants^a**

(Dependent variable = total earnings in 1984; mean = \$776; standard errors in parentheses)

Applicant Group and Other with Independent Variables (1)	Mean Earnings (2)	Coefficients of Dummy Variables ^b (3)	Coefficients in Regression with Conventional Independent Variables ^c (4)	Coefficients in Regression Subjective Ranking ^d (5)
Controls	\$749 ^e (63)	--	--	--
Participants	898	149 (93)	127 (92)	126 (92)
No Shows	710	-39 (179)	-29 (177)	-29 (177)
Screen Outs	592	-157 (110)	-69 (126)	-91 (133)
Dropouts/NEC	823	74 (110)	104 (126)	86 (131)
Subjective Rank				24 (48)
N ^f	499	499	493	493
R ²		.02	.08	.08

^aFor Cohort 3, the application and program assignment took place between July 1984 and May 1985.

^bRegression with only applicant group variables.

^cConventional independent variables: race, state location, education, marital status, age, number of children, worked for pay before, highest wage attained. (See Appendix for full definitions and coefficients.)

^dIncludes all independent variables in column (4) plus subjective ranking of administrator.

^eIntercept term (and standard error in parentheses) from regression (3), with Controls as omitted dummy variable.

earnings with survey data.⁶ Their intended role in an evaluation analysis is to account for productivity differences (or determinants of earnings) that are not attributable to the training program. The list of variables, referred to in the table as "conventional" explanatory variables, includes the race (or ethnicity), age, marital status, and education of the woman, the number of dependent children she has, her state of residence, whether the woman had ever been employed before, and her highest wage obtained in past employment. These eight types of explanatory variables are, in fact, specified with 22 right-hand-side variables. The full list of variables is shown in Appendix Table A.1, along with their means and estimated coefficients in a representative earnings regression from Table 7.

A surprising finding in Tables 3-7 is that the coefficients in column (4) for No Shows, Screen Outs, and Drop Outs are not very different (and sometimes smaller in absolute value) from the corresponding coefficients for these groups in column (3). Evidently, the conventional productivity variables have little systematic effect on narrowing the earnings differential between Controls and Screen Outs (or between the Controls and the other two comparison groups), and, therefore, do not appear to account for the selection bias affecting earnings.

The coefficients of the Participant group are discussed later, when the findings for the experimental design of the analysis are considered. At this point we note only that the coefficients of the Participant dummy variable do not much change as explanatory variables are added, as indeed they should not, given random assignments. The small changes that do occur may reflect the nonrandom selection (or separation) of No Shows from the Participants.

Column (5) in Tables 4-7 shows the coefficients of the dummy variables for the applicant groups in a regression that adds 1984 earnings to the variables used in column (4). Earnings in 1984 are for the period before the training program began for most of the women in Cohort 3. Although not always available in evaluation analyses, a pretraining earnings variable is recognized to be an important predictor of a person's productive capacity. Earnings in 1984 are, indeed, a powerful

TABLE 4

**Regression Results for Cohort 3 for 1985, a Year of Subsidized
Employment for Most Participants^a**

(Dependent variable = total earnings, 1985; mean = \$2,240; standard errors in parentheses)

Applicant Group and Other Independent Variables (1)	Mean Earnings (2)	Coefficients of Dummy Variables ^b (3)	Coefficients in Regression with Conventional Independent Variables ^c (4)	Coefficients in Regression with 1984 Earnings ^d (5)	Coefficients in Regression with Subjective Ranking ^e (6)
Controls	\$1,495 ^f (89)	--	--	--	--
Participants	4,273	2,778 (131)	2,729 (129)	2,608 (95)	2,610 (95)
No Shows	1,484	-11 (251)	139 (248)	167 (182)	168 (182)
Screen Outs	1,154	-341 (155)	-399 (177)	-333 (130)	-263 (137)
Dropouts/NEC	1,477	-18 (155)	-135 (177)	-234 (130)	-177 (135)
1984 Total Earnings				.95 (.05)	.96 (.05)
Subjective Rank					-78 (49)
N ^f	499	499	493	493	493
R ²		.56	.60	.79	.79

^aFor Cohort 3, the application and program assignment took place between July 1984 and May 1985.

^bRegression with only applicant group variables.

^cConventional independent variables: race, state location, education, marital status, age, number of children, worked for pay before, highest wage attained. (See Appendix for full definitions and coefficients.)

^dIncludes all independent variables in column (4) plus 1984 total earnings.

^eIncludes all independent variables in column (5) plus subjective ranking.

^fIntercept term (and standard errors in parentheses) from regression (3), with Controls as omitted dummy variable.

TABLE 5

Regression Results for Cohort 3 for 1986, a Year of Subsidized Employment for Many Participants^a
 (Dependent variable = total earnings, 1986; mean = \$2,175; standard errors in parentheses)

Applicant Group and Other Independent Variables (1)	Mean Earnings (2)	Coefficients of Dummy Variables ^b (3)	Coefficients in Regression with Conventional Independent Variables ^c (4)	Coefficients in Regression with 1984 Earnings ^d (5)	Coefficients in Regression with Subjective Ranking ^e (6)
Controls	\$2,473 ^f (110)	--	--	--	--
Participants	3,838	1,365 (162)	1,283 (160)	1,158 (132)	1,161 (132)
No Shows	2,295	-178 (310)	-155 (307)	-126 (253)	-125 (253)
Screen Outs	1,929	-544 (191)	-619 (219)	-550 (181)	-471 (191)
Dropouts/NEC	2,139	-334 (192)	-517 (219)	-620 (180)	-555 (187)
1984 Total Earnings				.99 (.07)	.99 (.07)
Subjective Rank					-88 (69)
N ^f	499	499	493	493	493
R ²		.22	.29	.52	.52

^aFor Cohort 3, the application and program assignment took place between July 1984 and May 1985.

^bRegression with only applicant group variables.

^cConventional independent variables: race, state location, education, marital status, age, number of children, worked for pay before, highest wage attained. (See Appendix for full definitions and coefficients.)

^dIncludes all independent variables in column (4) plus 1984 total earnings.

^eIncludes all independent variables in column (5) plus subjective ranking.

^fIntercept term (and standard errors in parentheses) from regression (3), with Controls as omitted dummy variable.

TABLE 6

Regression Results for Cohort 3 for 1987, the First Postprogram Year for All Participants^a
 (Dependent variable = total earnings, 1987; mean = \$3,261; standard errors in parentheses)

Applicant Group and Other Independent Variables (1)	Mean Earnings (2)	Coefficients of Dummy Variables ^b (3)	Coefficients in Regression with Conventional Independent Variables ^c (4)	Coefficients in Regression with 1984 Earnings ^d (5)	Coefficients in Regression with Subjective Ranking ^e (6)
Controls	\$3,139 ^f (127)	--	--	--	--
Participants	3,935	796 (187)	641 (178)	572 (161)	578 (161)
No Shows	2,951	-188 (358)	-267 (341)	-236 (308)	-235 (307)
Screen Outs	2,627	-512 (222)	-582 (245)	-503 (221)	-374 (232)
Dropouts/NEC	3,046	-93 (221)	-351 (243)	-435 (219)	-327 (228)
1984 Total Earnings				.96 (.09)	.96 (.09)
Subjective Rank					-147 (84)
N ^f	495	495	489	489	489
R ²		.07	.21	.36	.36

^aFor Cohort 3, the application and program assignment took place between July 1984 and May 1985.

^bRegression with only applicant group variables.

^cConventional independent variables: race, state location, education, marital status, age, number of children, worked for pay before, highest wage attained. (See Appendix for full definitions and coefficients.)

^dIncludes all independent variables in column (4) plus 1984 total earnings.

^eIncludes all independent variables in column (5) plus subjective ranking.

^fIntercept term (and standard error in parentheses) from regression (3), with Controls as omitted dummy variable.

TABLE 7

Regression Results for Cohort 3 for 1988, the Second Postprogram Year for All Participants^a
 (Dependent variable = total earnings, 1988; mean = \$3,989; standard errors in parentheses)

Applicant Group and Other Independent Variables (1)	Mean Earnings (2)	Coefficients of Dummy Variables ^b (3)	Coefficients in Regression with Conventional Independent Variables ^c (4)	Coefficients in Regression with 1984 Earnings ^d (5)	Coefficients in Regression with Subjective Ranking ^e (6)
Controls	\$3,896 ^f (148)	--	--	--	--
Participants	4,626	730 (217)	597 (210)	530 (194)	540 (192)
No Shows	3,683	-213 (416)	-203 (403)	-172 (371)	-169 (368)
Screen Outs	3,376	-520 (256)	-646 (287)	-560 (265)	-300 (277)
Dropouts/NEC	3,765	-131 (257)	-373 (287)	-462 (264)	-248 (272)
1984 Total Earnings				1.02 (.11)	1.02 (.11)
Subjective Rank					-292 (100)
N ^f	498	498	492	492	492
R ²		.05	.17	.29	.31

^aFor Cohort 3, the application and program assignment took place between July 1984 and May 1985.

^bRegression with only applicant group variables.

^cConventional independent variables: race, state location, education, marital status, age, number of children, worked for pay before, highest wage attained. (See Appendix for full definitions and coefficients.)

^dIncludes all independent variables in column (4) plus 1984 total earnings.

^eIncludes all independent variables in column (5) plus subjective ranking.

^fIntercept term (and standard error in parentheses) from regression (3), with Controls as omitted dummy variable.

predictor with these grouped data, even four years later, in 1988. The regression coefficient in Tables 4 to 7 shows that for each additional one dollar of earnings in 1984, the current year's earnings are expected to be one dollar higher, and the t-ratio for the coefficient is about 10, which is remarkably high. Evidently, the use of group means as the observational unit is responsible for the strong and stable relationship between past and current earnings. No such stability between a prior year's earnings and later year's earnings would be found in a regression with individual persons as the units of observation.

Finally, column (6) shows the coefficients of the dummy variables for the applicant groups in a regression that adds the subjective ranking, which is not usually available and which offers a unique opportunity for modeling the selection process among applicants to the various groups. Given the presence of the other explanatory variables, subjective ranking has a large and statistically significant effect (in the expected direction) on 1987 and 1988 earnings, and smaller and only marginally significant effects on earnings in 1985 and 1986. (It has essentially a zero effect on 1984 earnings.) Column (6) is critical for testing whether any of the nonexperimental comparison groups is notably improved as an alternative to the experimental Controls when the subjective ranking is "held constant." We see below that this improvement occurs, notably for the Screen Outs, for whom the subjective ranking is particularly relevant.

Consider Tables 5 and 6, which cover the two years, 1987 and 1988, that are entirely posttraining years. Among the three applicant groups that are alternatives to the Control group, the No Shows (who were selected to participate in the training program) have earnings that are most similar to the Controls. The earnings of the No Shows are moderately less than those for the Controls, after including all available independent variables in the regression: \$235 less in 1987 and \$169 less in 1988. Neither difference is statistically significant, but there may be a practical difference in the

percentage differences: 4 percent less in 1988 and 7 percent less in 1987, particularly if differences of even this small amount were to persist over many years.

The earnings of the No Shows are also more similar to those of the Controls for each of the other years, 1984 to 1986, than are the earnings of the Screen Outs or Drop Outs. As noted above, however, we are reluctant to use the earnings similarity between the No Shows and Controls to claim that No Shows provide a nonexperimental comparison group that would be generally valid, because the theoretical basis for the similarity is not firm, and the results we find may be peculiar to our data set. It is true that the No Shows are like the Control and Participant groups in having been selected into the program, but we have no specific information on their reasons for dropping out. The subjective ranking variable and 1984 earnings (for the preprogram year) are virtually the same, so there is little scope for modeling the selection process in the no-show decision.

Our attempt to model the selection process for the Screen Outs shows mixed results in terms of securing a nonexperimental comparison group, although on balance we believe that the results are quite promising. Although the (negative) earnings differences between Screen Outs and Controls are still of practical significance by 1987 and 1988 -- \$374, or 12 percent, less in 1987, and \$300, or 8 percent, less in 1988--the use of the subjective ranking variable sharply reduces the earnings differences. Compare the coefficients for the Screen Out dummy variable in column (6), when the subjective ranking is included, with its coefficients in column (5), when the subjective ranking variable is not included. The Control-Screen Out differential is reduced by 26 percent in 1987 (from \$503 to \$374) and by 46 percent in 1988 (from \$560 to \$300). We suggest that if even this crudely measured subjective variable has this much success in explaining the differences between applicants who are selected to participate in the program from those who are not selected, there is reason to believe that a more thorough measurement and modeling of the selection process would provide a valid comparison group in a nonexperimental design.

The subjective ranking variable has both practical and statistical significance in the 1987 and 1988 regressions. For each one unit of the ranking, from 1 to 2 or from 4 to 5, for example, the mean earnings of a group are predicted to decline by \$147 in 1987 (a 5 percent decline relative to the Control group mean) and by \$292 in 1988 (a 7 percent decline).⁷ As shown in Table 1, the mean subjective ranking for the Controls was 2.3 and that for Screen Outs was 3.1. (Recall that the value of 3 is assigned to missing values of the ranking, which serves to increase slightly the mean for the Controls and to decrease considerably the mean for the Screen Outs.)

The Drop Out comparison group is shown to be quite similar to the Screen Outs. Furthermore, even though only 37 percent of the Drop Outs were assigned a subjective ranking by the program administrators, using this variable also reduced the difference between the Control earnings and the Drop Out earnings by as much in percentage terms as it did for the Screen Outs. Thus, there is also promise in considering this nonexperimental group as a comparison group, particularly if some information on a subjective ranking can be obtained. Based on the data in this study, the Screen Outs and Drop Outs could be combined, if increasing sample size was important.

IV. ESTIMATION RESULTS OF PROGRAM IMPACTS: TABLES 8 AND 9

Table 8 concentrates on the estimates of the earnings effect of the training program when the Screen Outs or No Shows are used as comparison groups, using the observations for all cohorts in addition to those for Cohort 3. For Cohort 3 the regressions are similar to those in Tables 3-7 except that separate regressions are fit with just Screen Outs and Participants or with just No Shows and Participants. Using separate regressions for each pair of applicant groups in Table 8 allows all the independent variables in the model to affect EARNNS additively for just the two applicant groups, whereas in Tables 3-7 the independent variables are additive across all the applicant groups. To the extent that there are interactions between the independent variables and the applicant-group dummy

TABLE 8**Program Participation Effect: Earnings Differences between Participants
as Compared to Screen Outs and No Shows**(Dependent variable = total earnings; standard errors in parentheses)^a

	Cohort 3 ^b		All Cohorts ^c	
	Screen Outs (1)	No Shows (2)	Year	No Shows (3)
1984: Preprogram year	270 (158)	66 (217)	1984: Mostly in-program year	1,272 (164)
1985: In-program year	2,969 (175)	2,047 (236)	1985: Mix of in- and postprogram years	1,763 (210)
1986: Mix of in- and postprogram years	1,698 (233)	1,094 (303)	1986: Mostly postprogram year	675 (244)
1987: Postprogram year	1,004 (252)	859 (345)	1987: Postprogram year	716 (255)
1988: Postprogram year	767 (299)	626 (403)	1988: Postprogram year	613 (285)
Sample Sizes ^d	224-228	164-165		411-414

^aAll coefficients are taken from a regression with all appropriate independent variables.^bThe regressions for Cohort 3 include 1984 total earnings except for the regression for 1984.^cThe regressions for all cohorts exclude 1984 total earnings as an independent variable because 1984 earnings are an outcome of the training program for Cohorts 1 and 2.^dSome years had a few observations deleted because of obvious errors in the earnings or an excessive amount of missing values among the independent variables.

variables, the results from Tables 3-7 and Table 8 (for Cohort 3) will differ. The differences turn out to be minor.

Nevertheless, an important conceptual issue is illustrated by Table 8, which is that in a nonexperimental design only the Participants are the policy-relevant group to which comparisons are made. Recall that in the experimental design the Controls are compared with the combined groups of Participants and No Shows. In a nonexperimental design there are no Controls, and the No Shows are considered as a potential comparison group. In Table 8 the figures for Screen Outs and for No Shows express differences relative to the earnings of Participants, where positive values measure an excess or gain of Participants.

The results for Cohort 3 in Table 8, as in Tables 3-7, show positive effects of the training program that are very high in 1985, when many of the Participants are in subsidized employment. The earnings gain of the Participants declines over the next three years, but even by 1988 the gain is statistically and practically significant and larger than the gains relative to the Control group, which will be discussed in Table 9.⁸

There are data for Participants and No Shows for all three cohorts, and the EARNNS regressions for these data are shown in column (3) of Table 8, using the regression with all the independent variables except for 1984 earnings. Recall that 1984 earnings for Participants in Cohort 1, and to a lesser extent in Cohort 2, are partly attributable to the subsidized employment component of the training program. Thus, the earnings advantage of the Participants in 1984, \$1,272 in column (3) of Table 8, reflects the subsidized earnings and explains why the Participant advantage in (mainly) preprogram earnings for 1984 is so much less for Cohort 3. The earnings gains of the Participants for 1985 to 1988 are also less than those for Cohort 3, shown in column (2), but the year-by-year decline in the gain and the values for 1987 and 1988 are not sharply different between columns (2) and (3).

TABLE 9

Experimental Effects: Earnings Differences between Controls and Combined Participants and No Shows (P+NS), Cohort 3 and All Cohorts, by Year

(Dependent variable = total earnings; standard errors in parentheses)

Year; Selected Independent Variables	Cohort 3		All Cohorts	
	Coefficients in Simple Regression ^a	Coefficients in Regression with All Independent Variables ^b	Coefficients in Simple Regression ^a	Coefficients in Regression with All Independent Variables ^c
1984: Intercept (P+NS)	749	--	1,138 (74)	--
Subjective Rank	122 (92)	104 (91)	1,136 (104)	1,143 (77)
		-1 (64)		-136 (84)
1985: Intercept (P+NS)	1,495	--	2,271 (77)	--
Subjective Rank	2,382 (153)	2,237 (120)	1,608 (109)	1,603 (102)
1984 Total Earnings		-65 (84)		-229 (71)
		.98 (.08)		
1986: Intercept (P+NS)	2,473	--	3,039 (83)	--
Subjective Rank	1,146 (167)	948 (137)	741 (78)	718 (108)
1984 Total Earnings		-49 (97)		-200 (75)
		1.02 (.09)		
1987: Intercept (P+NS)	3,139	--	3,746 (94)	--
Subjective Rank	655 (185)	433 (160)	400 (133)	345 (119)
1984 Total Earnings		-166 (113)		-251 (83)
		1.05 (.12)		
1988: Intercept (P+NS)	3,896	--	4,374 (102)	--
Subjective Rank	595 (217)	408 (195)	524 (144)	480 (128)
1984 Total Earnings		-263 (137)		-283 (89)
		1.14 (.15)		
Sample sizes	n = 326-331 ^d		n = 826-829 ^d	

^aIn the simple regression, the intercept = Control mean of total earnings. (No standard error is shown.)

^bIndividual variables include: race, state, education, marital status, age, children, worked before, highest wage before, subjective rank, and 1984 total earnings (except in the 1984 regression). Intercept term is not shown.

^cIncludes all in ^b except 1984 total earnings. Intercept term is not shown.

^dSome years had a few observations deleted because of obvious errors in the earnings or an excessive amount of missing values among the independent variables.

Table 9 concentrates on the classical experimental results between the Control group and the "treatment" group, where the treatment group is composed of Participants and No Shows, the two groups randomly assigned to the training program. Again, comparisons are shown for Cohort 3 and all cohorts, and again the results for Cohort 3 are similar to those that can be obtained from Tables 3-7. In Table 9 separate regressions are estimated with just the two applicant groups, now treating Participants and No Shows (P+NS) as a single group. Controls are the omitted group in the regressions, so the positive coefficients of the (P+NS) dummy variable reflect the earnings advantage of the (P+NS) group.

Results from a simple regression, using only a single dummy variable for (P+NS), are shown along with the multiple regression using all the independent variables, including the subjective ranking variable, S, and 1984 earnings for Cohort 3, when the 1984 earnings can be considered to be (mainly) preprogram earnings. It is interesting that the subjective ranking is consistent in sign and usually statistically significant as a predictor of earnings, even though the variation of S within the screened-in groups is relatively small. The following three findings from Table 9 are especially noteworthy.

Finding 1. The program gains in Table 9 are smaller than the Participant-Control differences reported earlier, which reflects the fact that the No Shows earned less than the Participants in each year. We have no way of knowing whether the earnings advantage of Participants relative to the No Shows is entirely attributable to the gains from training or partially attributable to a positive selection bias for the Participants. Assuming that the smaller estimated gain for the (P+NS) group is unbiased and assuming that there is no effect of the training program or of the subsequent experiences of the trainees on the No Shows, we can attribute all of the (P+NS) gain to the training program by multiplying it by the reciprocal of the ratio, $P/(P+NS)$. For both Cohort 3 and all cohorts, the Participants were approximately 85.5 percent of the total, so the multiplier is 1.17. Thus, all estimated coefficients of the (P+NS) dummy variable in Table 9 could be multiplied by 1.17 to obtain an upwardly adjusted measure of the earnings effect of the training program.

Because Table 8 shows a direct comparison between alternative comparison groups and Participants, which, again, would be the only comparison available in a nonexperimental evaluation, the upwardly adjusted estimates of the earnings gains from training from Table 9 may be considered the appropriate comparison to make with Table 8. For example, the 1988 gains from the training program in Table 9 are \$480 for all cohorts and \$408 for Cohort 3, and multiplied by 1.17 these would be \$562 and \$477. From Table 8, the gains using No Shows as a comparison group are \$613 for all cohorts and \$647 for Cohort 3, and the gain using Screen Outs is \$767 for Cohort 3. Generally, the upwardly adjusted estimates of the earnings effect of the training program from Table 9 are smaller than the estimates in Table 8. This is consistent with the findings in Tables 3-7, which show the earnings of the Screen Outs and No Shows to be less than the earnings of Controls.

Finding 2. Another result from Table 9, again consistent with the previous tables, is the smaller earnings gain from training in the later periods, 1987 and 1988, when all subsidized employment had ended for Cohort 3, and from 1986 to 1988, when all or most of the subsidized employment had ended for all cohorts. It is likely that many training programs bring about increased earnings from job placement effects as well as from increases in earnings capacity (or gains in human capital), and we should expect that the job placement effects on earnings will decay with time, in contrast with human capital effects, which should hold fairly steady over time. If this interpretation of program effects is correct, there is a clear need to obtain employment and earnings information for a reasonably long period after the program ends--at least two or three years.

Finding 3. Finally, Table 9 shows modest but noticeably larger estimates of the earnings effect of the training program in the simple regression than in the multiple regression, especially for Cohort 3. This is unexpected if assignments were truly random and the samples are large. Again, this result was somewhat in evidence in Tables 3-7, where there was a rather consistent decline in the coefficient of Participants when comparing the simple regression result (in column 3) with the results in columns 4, 5, and 6, particularly for the years from 1986 to 1988. Several explanations are possible, including

sampling variability or other sources of an initial selection that favored the Participants. We conclude that close attention to the selection process is important in any evaluation analysis.

V. CONCLUSIONS

The need for nonexperimental methods to evaluate government programs for employment, training, welfare reform, education, and other such programs is indisputable. Most of these programs will not be conducted as controlled experiments, even when they are demonstration programs. This paper offers a number of findings that lend some optimism to the search for valid nonexperimental methods. The key theoretical concept is that if the selection process in determining program participation can be modeled, unbiased estimates of the program effect are attainable.

Using an evaluation study of a training program that had an experimental design, we pursued two strategies that suggest the feasibility of accounting for the selection process. One is to use applicants as the source of one or more comparison groups, which effectively controls for the self-selection of volunteering. Using applicant data also controls for the specific time and location of the program. A second strategy is to use information on the selection procedures of the program administrators -- information that includes the subjective ranking of the applicants in terms of their suitability for the program in addition to many objective criteria. Although crudely measured and not carefully designed to capture wide variation in suitability, the variable produced results that show what might be achieved if there were a dedicated effort to measure administrative selection procedures. When the subjective ranking of the applicants was added to the more conventional list of independent variables, the two nonexperimental applicant groups, Screen Outs and Drop Outs, appeared relatively similar to the Controls in predicted earnings in the two postprogram years, 1987 and 1988. (See Tables 6 and 7.)

A final promising result from our study is the success in using IRS data for measuring the earnings of the applicant groups for years that preceded, overlapped, and, most important, followed the period of the training program.

APPENDIX TABLE A.1
Cohort 3 Regression for 1988 (Table 7, Column (6))
An Illustrative Example of Complete Regression Specification Shown in Column (6)
of Tables 4-7

(Dependent variable = total earnings, 1988; mean = \$3,989)

Independent Variable	Mean (Standard Deviation)		Coefficient (Standard Error)	
1. Intercept	--		6,217	(1,319)
2. Participant (dummy variable)	.29	(.45)	540	(192)
3. No Show (dummy variable)	.05	(.21)	-169	(368)
4. Screen Out (dummy variable)	.17	(.37)	-300	(277)
5. Drop Out (dummy variable)	.16	(.37)	-248	(272)
6. Control (omitted)	.33	(.47)		
7. Black %	.55	(.49)	-315	(360)
8. Hispanic %	.10	(.28)	15	(513)
9. White % (omitted)	.29	(.44)		
10. Unknown or Other Race %	.06	(.22)	-322	(680)
11. New York %	.15	(.21)	-968	(607)
12. New Jersey %	.19	(.23)	-116	(607)
13. Ohio %	.14	(.25)	524	(569)
14. Texas %	.16	(.20)	1,816	(5,976)
15. Kentucky %	.13	(.23)	-974	(673)
16. South Carolina %	.07	(.18)	-598	(660)
17. Arkansas % (omitted)	.12	(.19)		
18. State Missing	.03	(.15)	2,001	(5,305)
19. Education Years of Schooling < 12, %	.47	(.23)	-1,971	(646)
20. Education Years of Schooling = 12, %	.33	(.19)	-1,480	(722)
21. Education Years of Schooling > 12, % (omitted)	.16	(.14)		
22. Education Years of Schooling, Missing %	.04	(.15)	-2,229	(4,406)
23. Marital Status: Single, %	.33	(.25)	582	(622)
24. Married, Widowed, Separated, Divorced, % (omitted)	.50	(.29)		
25. Marital Status, Missing, %	.18	(.29)	-1,116	(5,949)
26. Average number of dependent children	2.01	(.39)	68	(208)
27. Ever worked for pay, %	.88	(.32)	-30	(535)
28. Average maximum wage previously earned (wage ≤ \$25)	3.74	(2.14)	24	(51)
29. Maximum previous wage, missing, %	.20	(.36)	-319	(513)
30. Maximum previous wage reported as > \$25, %	.17	(.18)	555	(543)
31. Maximum previous wage, ≤ \$25, % (omitted)	.63	(.32)		
32. Average age	29.9	(3.3)	-33	(33)
33. Total earnings in 1984	776	(821)	1.02	(.11)
34. Average subjective ranking ("potential")	2.5	(.9)	-292	(100)

Endnotes

¹The literature on these issues is quite large. Recent studies include LaLonde (1986), Fraker and Maynard (1987), Heckman and Hotz (1989) and Friedlander and Robins (1992).

²See Bell and Orr (1988) and Bell and Orr (forthcoming) for analyses of the experimental design of this program and for evidence that the program substantially increased the earnings of the trainees in the postprogram period covered by their survey data.

³Eligibility was limited to individuals who had received AFDC continuously over the previous three months and who were not employed as homemaker-health aides in that period. Some states also required access to a car, a drivers license, and/or a minimum level of education or reading ability. The administrators' control over screening for suitability for the program was in addition to these eligibility criteria.

⁴Five of the group units had severe outlier values for earnings for (usually) one year, and these outliers were dropped when estimating the model for these years. Three were Participants and two were Screen Outs. Two other group units were borderline cases of outliers, and these were handled "conservatively," in the sense that they were not included if they served to make our results look more "positive."

⁵It is interesting to note that the national unemployment rate was at its highest annual level, 9.6 percent, in 1983 and declined to 7.5 percent in 1984 and 7.2 percent in 1985. The lower earnings of Cohort 3 may reflect a pool of applicants who, because they were not working in a somewhat tighter labor market, relative to 1983, had a lower earnings capacity than the two earlier cohorts.

⁶A few observations were dropped in these regressions because of excessive missing values in the variables used. When we used only the lesser number of observations for regressions in column (3), we obtain virtually the same results.

⁷The simple linear functional form for S , the subjective ranking variable, was found to be adequate to estimate its effect. Using a set of dummy variables for S did not improve the fit of the regression.

⁸We have also estimated the model with Screen Outs and Participants in which all three cohorts of Participants are included and in which the cohort group is controlled by two dummy variables. The results are very close to those shown in column (1) of Table 3. The dummy variables for Cohort 1 and Cohort 2 are positive and significant, but the basis for these effects are solely the Participant earnings, because the Screen Outs are all from Cohort 3.

References

- Ashenfelter, Orley, and David Card. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." Review of Economics and Statistics 67: 648-660.
- Bell, Stephen H., and Larry L. Orr. 1988. "Screening (and Creaming?) Applicants to Job Training Programs: The AFDC Homemaker-Home Health Aide Demonstrations." Paper presented to the Annual Research Conference of the Association for Public Policy Analysis and Management. Bethesda, Maryland: Abt Associates, Inc.
- _____. Forthcoming (1994). "Is Subsidized Employment Cost-Effective for Welfare Recipients? Experimental Evidence from Seven State Demonstrations." Journal of Human Resources 29 (1, Winter).
- Fraker, Thomas M., and Rebecca A. Maynard. 1987. "The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs." Journal of Human Resources 22: 194-227.
- Friedlander, Daniel, and Philip K. Robins. 1992. "Estimating the Effects of Employment and Training Programs: An Assessment of Some Nonexperimental Techniques." New York: Manpower Demonstration Research Corporation. October.
- Heckman, James L., and V. Joseph Hotz. 1989. "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." Journal of the American Statistical Association 84: 862-874.
- LaLonde, Robert J. 1986. Evaluating the Econometric Evaluations of Training Programs with Experimental Data." American Economic Review 76: 604-620.