

Contract No.: K-4279-3-00-80-30
MPR Reference No.: 8140-520

**National Job Corps
Study: Methodological
Appendixes on the
Short-Term Impact
Analysis**

Final Report

February 9, 2000

Peter Z. Schochet

Submitted to:

U.S. Department of Labor
Employment and Training Administration
Office of Policy and Research
Room N-5637
200 Constitution Ave., NW
Washington, DC 20210

Submitted by:

Mathematica Policy Research, Inc.
(Prime Contractor)
P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535

Project Officer:

Daniel Ryan

Project Director:

John Burghardt

Principal Investigators:

Terry Johnson
Charles Metcalf

In conjunction with:

Battelle Human Affairs Research
Centers (Subcontractor)
4000 N.E. 41st St.
Seattle, WA 98105

Decision Information Resources, Inc.
(Subcontractor)
610 Gray St., Suite 200
Houston, TX 77002

ACKNOWLEDGMENTS

This report reflects the contributions of several people. It was skillfully produced by Cathy Harper and carefully edited by Walter Brower. John Burghardt provided helpful comments.

TABLES

Table		Page
A.1	EFFECTIVE RESPONSE RATES TO THE 12-MONTH FOLLOW-UP INTERVIEW, BY RESEARCH STATUS AND KEY SUBGROUP	9
A.2	CUMULATIVE DISTRIBUTION OF THE NUMBER OF MONTHS BETWEEN 12 MONTHS AFTER RANDOM ASSIGNMENT AND COMPLETION OF THE 12-MONTH INTERVIEW FOR THOSE IN THE IN-PERSON AREAS, BY RESEARCH STATUS	14
A.3	INTERVIEW MODE FOR CASES WHO COMPLETED THE 12-MONTH INTERVIEW, BY RESEARCH STATUS AND GENDER	15
A.4	REASONS FOR NONCOMPLETION OF THE 12-MONTH INTERVIEW, BY RESEARCH STATUS AND GENDER	17
A.5	EFFECTIVE RESPONSE RATES TO THE 30-MONTH FOLLOW-UP INTERVIEW, BY RESEARCH STATUS AND KEY SUBGROUP	19
A.6	CUMULATIVE DISTRIBUTION OF THE NUMBER OF MONTHS BETWEEN 30 MONTHS AFTER RANDOM ASSIGNMENT AND COMPLETION OF THE 30-MONTH INTERVIEW FOR THOSE IN THE IN-PERSON AREAS, BY RESEARCH STATUS	22
A.7	INTERVIEW MODE FOR CASES WHO COMPLETED THE 30-MONTH INTERVIEW, BY RESEARCH STATUS AND GENDER	24
A.8	REASONS FOR NONCOMPLETION OF THE 30-MONTH INTERVIEW, BY RESEARCH STATUS AND GENDER	25
B.1	DATA ITEM RESPONSE FOR KEY OUTCOME MEASURES USED IN THE 30-MONTH IMPACT ANALYSIS, BY RESEARCH STATUS AND GENDER	30
C.1	PROBABILITIES THAT ELIGIBLE APPLICANTS WERE SELECTED TO THE CONTROL AND PROGRAM RESEARCH GROUPS, BY SAMPLING STRATA	51
C.2	COMPARISON OF THE CHARACTERISTICS OF RESPONDENTS AND THE FULL SAMPLE OF RESPONDENTS AND NONRESPONDENTS TO THE 30-MONTH INTERVIEW, BY RESEARCH STATUS	58

TABLES *(continued)*

Table	Page
D.1 CONTROL VARIABLES INCLUDED IN THE REGRESSION MODELS TO OBTAIN REGRESSION-ADJUSTED IMPACT ESTIMATES	83
D.2 IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES FOR THE FULL SAMPLE	87
D.3 IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES FOR MALES	89
D.4 IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES FOR FEMALES	91
D.5 IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES FOR 16- AND 17-YEAR-OLDS	93
D.6 IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES FOR 18- AND 19-YEAR-OLDS	95
D.7 IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES FOR 20- TO 24-YEAR-OLDS	97
D.8 IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES FOR RESIDENTIAL DESIGNEES	99
D.9 IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES FOR NONRESIDENTIAL DESIGNEES	101

INTRODUCTION

In a series of appendixes, this report discusses methodological issues related to the 30-month impact analysis for the National Job Corps Study. The appendixes are intended to complement the 30-month impact report (Schochet et al. 1999), which presents short-term impacts of Job Corps on key participant outcomes during the 30 months after random assignment.

This report contains the following five appendixes:

1. **"The 12-Month and 30-Month Interviews."** The outcome measures for the 30-month impact analysis were constructed using follow-up interview data collected 12 and 30 months after random assignment. This appendix provides a detailed discussion of the design of the follow-up interviews and examines response rates.
2. **"The Treatment of Missing Values and Outliers."** This appendix describes our procedure for treating missing values and outliers for the outcome measures used in the 30-month impact analysis.
3. **"The Calculation of Sample Weights and Standard Errors."** This appendix discusses the calculation of sample weights used in the 30-month impact analysis to obtain unbiased impact estimates that could be generalized to the study population. The appendix also discusses the calculation of standard errors of the impact estimates.
4. **"Regression-Adjusted Impact Estimates."** This appendix discusses impact estimates obtained using multivariate regression procedures. These regression-adjusted impact estimates are compared to the simple differences-in-means estimates that are presented in the 30-month impact report.
5. **"The Adjustment for Crossovers."** This brief appendix describes procedures that were used to adjust the impact estimates for the small number of control group members who enrolled in Job Corps during the period when they were not supposed to enroll.

APPENDIX A

THE 12-MONTH AND 30-MONTH INTERVIEWS

A. INTRODUCTION

Impact estimates over the 30 months after random assignment were obtained by comparing the outcomes of program group members (who could enroll in Job Corps) and control group members (who could not). The outcome measures for the analysis were constructed primarily from interview data collected 12 and 30 months after random assignment. This appendix discusses the design and implementation of the follow-up interviews.

Baseline interview data were also used to construct outcome measures covering the period between the random assignment and baseline interview dates. The design and implementation of the baseline interview is discussed in detail in Schochet (1998a). However, we summarize features of the baseline interview because it is necessary to understand the survey design for the baseline interview to understand the survey design for the follow-up interviews.

B. SURVEY DESIGN

1. Design of the Baseline Interview

Baseline interviewing took place between mid-November 1994 and July 1996. Detailed tracking information (contained in program intake forms sent to MPR as part of the random assignment process) was used to help locate youths. The Office of Management and Budget (OMB) approved the offering of a \$10 incentive fee to control group members and hard-to-locate program group members to induce them to complete the baseline interview.

After sample members had been randomly assigned, they were contacted by telephone as soon as possible (usually the same day) to increase the proportion of interview respondents who did not know their research status prior to the interview.

At the end of May 1995, we began attempting in-person interviews with sample members not reachable by telephone. We waited until May to conduct these interviews so that enough sample

members had been released into the field to make it cost-effective to hire field interviewers. In-person interviews were attempted only with sample members who lived in randomly selected areas when they applied to Job Corps, because it would have been extremely expensive to conduct in-person interviews nationwide.¹ About two-thirds of randomized youths in the study population lived in areas selected for in-person interviewing when they applied to Job Corps.²

Sample members in the selected areas were released into the field for in-person interviewing if they could not be reached by telephone within 45 days after random assignment. During the post-45-day period, in-person and telephone interviews were attempted with these youths. However, during the post-45-day period, neither telephone nor in-person interviews were attempted with youths who lived in the areas not selected for in-person interviewing. Consequently, the sample interviewed within 45 days is a nationally representative *random* sample of eligible applicants who could be interviewed by telephone within 45 days. The sample interviewed after 45 days is a

¹In order to define areas for in-person interviewing, we divided the country into three types of areas, on the basis of adjoining groups of counties: (1) those in which about 1,000 Job Corps students resided in 1993 (*superdense* areas), (2) those in which about 600 Job Corps students resided in 1993 (*dense* areas), and (3) those in which about 300 students resided in 1993 (*nondense* areas). The "optimal" number of each type of area to select was calculated to maximize the precision of the impact estimates, subject to the cost of conducting interviews in each type of area and a fixed interview budget. On the basis of this procedure, we randomly selected all 16 superdense areas, 18 of the 29 dense areas, and 29 of the 75 nondense areas for in-person interviewing. All control group members designated for nonresidential slots on the Supplemental ETA-652 form, however, were eligible for in-person interviews to increase the precision of impact estimates for the small nonresidential program component.

²The figures for control group members (72 percent) and for program research group members (66.5 percent) differ because sampling rates to the research sample differed for various population subgroups.

nationally representative *clustered* sample of those who could be reached after 45 days. Both groups combined represent all persons in the study population.³

Baseline interviews were no longer attempted for sample members in the selected areas if they did not complete the interview within nine months of random assignment. However, as discussed in the next subsection, these youths were eligible for 12-month follow-up interviews.

2. Design of the 12-Month Interview

The 12-month interview was conducted between March 1996 and September 1997. With OMB approval to offer a finder's fee or an incentive payment to hard-to-locate sample members, we offered a \$10 inducement to program group members who were not at a Job Corps center and to all control group members. Interviews were attempted with youths between 12 and 27 months after their random assignment dates. Interviews completed between months 27 and 30 were 30-month interviews.

The target sample for the 12-month follow-up interview included (1) all sample members selected for in-person interviews at baseline (whether or not they completed a baseline interview), and (2) those not eligible for in-person interviews at baseline who completed the baseline interview by telephone within 45 days after random assignment. Thus, youths who resided in areas not selected for in-person interviews and who did not complete a baseline interview by telephone were not eligible for 12-month (and subsequent) interviews. In addition, we did not attempt follow-up interviews with 77 people selected for the study sample (40 program group and 37 control group

³We selected the 45-day cutoff after analyzing the cumulative telephone response rates by time since random assignment for the early cohort of sample members. The 45-day cutoff was chosen because telephone response rates increased slowly after this period. Furthermore, we did not want to extend the cutoff date, because we did not want to delay in-person interviewing in the in-person areas.

members), because these youths were found to have enrolled in Job Corps prior to random assignment. Consistent with our decision to include in the study only youths who had not previously attended Job Corps, these program readmits were removed from the study sample.⁴ Finally, 39 sample members (21 program and 18 control) were confirmed to have died. In total, 14,725 youths (9,017 program and 5,708 control) were released for 12-month interviews.

We completed 12-month interviews with 326 youths (187 program and 139 control) in the in-person areas who had not completed a baseline interview. An abbreviated baseline interview was administered to these “combo” cases at the end of the 12-month interview.

For the 12-month interview, we attempted interviews by telephone first and, if unsuccessful, attempted them in person. In contrast to the in-person interviewing at baseline, there was *no* clustering of in-person interviews in the follow-up interviews. In-person interviewing started in May 1996, after a sufficient number of youths had been released into the field.

3. Design of the 30-Month Interview

The 30-month interview was conducted between September 1997 and February 1999. A \$10 incentive fee was offered to all those in the target sample. Interviews were attempted with youths until 45 months after their random assignment dates. Interviews completed after then were treated as 48-month interviews.

⁴Because the study design excluded people who had previously enrolled in Job Corps, and because we believed Job Corps staff could identify these youths, Job Corps staff were not supposed to send information on program readmits to MPR for random assignment. However, in fact, staff were not able to identify all readmits, and information was mistakenly sent to MPR for some of these cases. After sample intake ended, we used historical information on center enrollees to identify those in our sample who enrolled in Job Corps prior to random assignment. Because information on the program readmits was sent *prior* to random assignment, there are no differences in the proportion or characteristics of readmits in the program and control groups; thus, we excluded these youths from the study.

A 30-month interview was attempted with all sample members who completed either the baseline or the 12-month interview, except for 16 youths who were confirmed to have died since their last interview. In total, 14,671 youths (8,983 program and 5,688 control) were released for 30-month interviews. The 493 respondents to the 30-month interview who completed a baseline interview but not the 12-month interview were asked about their experiences since the baseline interview.

As with the 12-month interview, we attempted 30-month interviews by telephone first and, if unsuccessful, attempted them in person to youths in *all* areas. In-person interviewing started in October 1997 and concluded in February 1999.

C. INTERVIEW RESPONSE ISSUES

This section discusses response rates to the baseline and follow-up interviews, the mode of completion of the follow-up interviews, and reasons for noncompletion of the follow-up interviews.

1. The Baseline Interview

As discussed in detail in Schochet (1998a), the response rate to the baseline interview for sample members in all areas was 93.1 percent. Interviews were completed with 14,327 of the 15,386 youths in the research sample, and most interviews were completed by telephone soon after random assignment. Furthermore, the difference in completion rates between the program and control groups was only 1.5 percentage points (93.8 percent program, 92.3 control). The response rate for sample members in the areas selected for in-person interviewing--the *effective* response rate--was 95.2 percent (95.9 percent program, 94.3 percent control). This is the relevant response rate for the study, because "nonrespondents" in the nonselected areas consisted of both those who would and those who would not have completed baseline interviews in the post-45-day period if given the

chance. Therefore, “true” respondents and nonrespondents can be identified only in the selected areas.

Response rates to the baseline interview were high for all key subgroups. Item nonresponse was infrequent for nearly all data items.

2. The 12-Month Interview

We completed 12-month interviews with 13,383 of the 14,725 youths released for 12-month interviews. For those in the in-person areas only, we completed 9,421 of the 10,448 interviews attempted. As Table A.1 shows, the effective response rate to the 12-month interview (that is, the response rate in the in-person areas) was 90.2 percent (91.4 percent program, 88.4 percent control).^{5,6} Nearly 98 percent of those who completed the 12-month interview also completed the full baseline interview.

The effective response rate to the 12-month interview differed only slightly across key youth subgroups (Table A.1). These response rates were calculated using ETA-652 and ETA-652 Supplement data, which are available for *both* interview respondents and nonrespondents, and refer to youth characteristics at the time of application to Job Corps. The response rate was slightly higher for females than males (92 percent, compared to 89 percent) and for younger sample members than

⁵As mentioned above, the effective response rate is the percentage of sample members in areas selected for in-person interviews at baseline who completed a 12-month interview. This is the relevant response rate for the study, because we did not attempt follow-up interviews with youths who were not selected for in-person interviews at baseline and who did not complete a baseline interview by telephone within 45 days after random assignment.

⁶The response rates exclude the program readmits and youths who died.

TABLE A.1

EFFECTIVE RESPONSE RATES TO THE 12-MONTH FOLLOW-UP INTERVIEW,
BY RESEARCH STATUS AND KEY SUBGROUP

Subgroup	Effective Response Rate		
	Program Group	Control Group	Combined Sample
Full Sample	91.4	88.4	90.2
Demographic Characteristics			
Gender			
Male	90.8	86.8	89.1
Female	92.2	91.0	91.8
Age at Application			
16 to 17	92.2	90.5	91.5
18 to 19	90.9	87.6	89.6
20 to 21	91.4	87.6	89.8
22 to 24	90.3	84.2	87.9
Race/Ethnicity			
White, non-Hispanic	89.9	87.0	88.7
Black, non-Hispanic	91.8	89.4	90.9
Hispanic	91.2	85.9	89.0
Other	94.6	90.6	92.9
Region			
1	90.0	88.8	89.5
2	94.3	85.8	90.7
3	90.4	89.1	89.9
4	91.2	88.0	89.9
5	90.4	88.3	89.6
6	91.9	87.0	89.9
7/8	92.3	92.9	92.6
9	90.1	87.0	88.8
10	93.2	87.9	91.0
Size of City of Residence			
Less than 2,500	93.9	91.7	93.0
2,500 to 10,000	91.8	89.7	91.0
10,000 to 50,000	92.7	86.8	90.4
50,000 to 250,000	90.5	89.2	90.0
250,000 or more	91.2	87.9	89.8
PMSA or MSA Residence Status			
In PMSA	91.1	86.6	89.3
In MSA	91.2	89.3	90.4
In neither	93.3	91.5	92.6

TABLE A.1 (continued)

Subgroup	Effective Response Rate		
	Program Group	Control Group	Combined Sample
Density of Area of Residence			
Superdense	91.2	88.1	90.0
Dense	91.1	88.5	90.0
Nondense	92.2	88.7	90.8
Lived in Areas with a Large Concentration of Nonresidential Females			
Yes	92.2	89.3	90.9
No	90.9	87.6	89.6
Legal U.S. Resident			
Yes	91.4	88.2	90.1
No	89.5	96.1	92.0
Job Corps Application Date			
11/94 to 2/95	90.3	88.7	89.7
3/95 to 6/95	92.6	88.9	91.1
7/95 to 9/95	91.9	89.1	90.7
10/95 to 12/95	90.5	86.5	88.8
Fertility and Family Status			
Fertility			
Had dependents	90.9	90.7	90.9
Had no dependents	91.6	87.8	90.0
Family Status			
Family head	91.6	89.8	90.9
Family member	92.6	88.9	91.1
Unrelated individuals	88.7	85.8	87.6
Education			
Completed the 12th grade	92.4	89.6	91.3
Did not complete the 12th grade	91.2	88.1	89.9
Welfare Dependence			
Public Assistance			
Received AFDC	92.5	89.9	91.5
Received other assistance	90.8	88.9	90.0
Did not receive	91.1	87.5	89.6
Health			
Had Any Health Conditions That Were Being Treated			
Yes	90.9	86.8	89.4
No	91.7	88.5	90.4

TABLE A.1 (continued)

Subgroup	Effective Response Rate		
	Program Group	Control Group	Combined Sample
Crime			
Arrests			
Arrested in past three years	89.8	87.0	88.7
Not arrested in past three years	91.7	88.4	90.4
Convictions			
Ever convicted or adjudged delinquent	91.1	88.6	90.0
Never convicted or adjudged delinquent	91.4	88.3	90.1
Baseline Interview Completion Status			
Completion Status			
Completed within 45 days			
Lived in in-person areas	92.8	90.9	92.0
Did not live in in-person areas ^a	93.2	91.5	92.6
Completed between 46 and 270 days	86.5	82.7	85.1
Did not complete	69.4	55.4	62.6
Anticipated Program Enrollment Information			
Residential Designation Status			
Resident	91.1	87.6	89.7
Nonresident	92.7	91.2	92.1
CCC/Contract Center Designation^b			
CCC center	91.6	88.5	90.3
Contract center	91.2	88.4	90.0
Performance Level of Designated Center^b			
High or medium high	90.8	88.8	90.0
Medium low or low	91.7	88.2	90.3
Size of Designated Center^b			
Large or medium large	90.6	88.1	89.6
Medium small or small	91.6	88.6	90.4
Sample Size	6,206	4,242	10,448

SOURCE: ETA-652 and ETA-652 Supplement data.

- NOTE:
1. The effective response rate is the response rate for those sample members who were eligible for a baseline interview after 45 days after random assignment. These youths lived in randomly selected (in-person) areas at application to Job Corps.
 2. The following cases were excluded from the calculations: (1) 39 cases (21 program group and 18 control group members) who were confirmed to have died, and (2) 77 cases (40 program group and 37 control group members) who were determined to have enrolled in Job Corps prior to random assignment.

TABLE A.1 (continued)

^a Figures pertain to those not in the in-person areas who completed baseline interviews within 45 days after random assignment. These youths were eligible for follow-up interviews.

^b Figures are obtained using data on OA counselor projections about the centers that youths were likely to attend.

older ones (92 percent for those 16 and 17 years old, compared to 88 percent for those 22 and older). In addition, for those who lived in less populated areas, those who completed high school, those never arrested, and likely nonresidential students, response rates were slightly higher than those of their counterparts. Because of these slight subgroup differences in response rates, sample weights for the 12-month interview sample were adjusted to help reduce the potential bias in the impact estimates due to interview nonresponse (see Appendix C).

It is noteworthy that among those who completed baseline interviews within 45 days after random assignment, the response rate for those who lived in the in-person areas was similar to the rate for those who did not (Table A.1). This is an expected result, because the in-person areas were randomly selected.

Most interview respondents completed the 12-month interview soon after their 12-month release date (Table A.2). The average 12-month interview was completed two months after the 12-month release date (that is, in month 14), and more than three-quarters of 12-month interviews were completed by month 15. Only about 8 percent of interviews were completed after month 18. The distributions of completion times were similar for program and control group members.

As shown in Table A.3, about 93 percent of interviews were completed by telephone in MPR's phone center through computer-assisted telephone interviewing (CATI). About 7 percent of interviews were conducted in the field (6 percent in person and 1 percent when the field interviewer had the youth call the MPR phone center). About 9 percent of program group interviews were conducted while the youth was in a Job Corps center. These figures are similar by research status and by gender.⁷

⁷We conducted (1) 24 interviews with youths who were living in a school or college, (2) 288 interviews with youths in jail (166 program and 122 control), (3) 46 interviews with youths living in halfway houses or residential treatment centers, (4) 62 interviews with youths in the military, (5) 77 interviews with youths in a group home, and (6) 30 interviews with homeless youths.

TABLE A.2

CUMULATIVE DISTRIBUTION OF THE NUMBER OF MONTHS BETWEEN 12 MONTHS
AFTER RANDOM ASSIGNMENT AND COMPLETION OF THE 12-MONTH INTERVIEW
FOR THOSE IN THE IN-PERSON AREAS, BY RESEARCH STATUS
(Percentages)

Number of Months	Program Group	Control Group	Combined Sample
-3 to 0 ^a	1.1	1.3	1.2
0 to less than .5	26.7	30.8	28.4
.5 to 1	15.0	15.2	15.1
1 to 2	21.8	20.3	21.2
2 to 3	12.9	10.5	12.0
3 to 4	7.1	6.9	7.0
4 to 5	4.2	3.8	4.1
5 to 6	2.9	2.8	2.8
6 to 12	7.3	7.6	7.4
12 to 15 ^b	0.9	0.8	0.8
Average Number of Months	2.1	2.0	2.0
Number of Respondents to the 12-Month Interview	5,673	3,748	9,421

SOURCE: 12-month follow-up interview data.

NOTE: The in-person areas are randomly selected areas in which youths were eligible for baseline interviews after 45 days after random assignment. Youths not in the in-person areas who did not complete baseline interviews within the 45-day period were not eligible for follow-up interviews.

^a Youths in the in-person areas who did not complete the baseline interview within 9 months (270 days) after random assignment but who were located before 12 months after random assignment were administered the 12-month interview and an abbreviated baseline interview.

^b Cases who were located after the 15-month period (that is, between 27 and 30 months after random assignment) were administered 30-month interviews.

TABLE A.3

INTERVIEW MODE FOR CASES WHO COMPLETED THE 12-MONTH INTERVIEW,
BY RESEARCH STATUS AND GENDER
(Percentages)

Interview Mode	Program Group			Control Group		
	Males	Females	Total	Male	Females	Total
Telephone Center	93.1	94.3	93.6	92.5	94.4	93.2
In the Field	6.9	5.7	6.4	7.5	5.6	6.8
Over the telephone	1.1	0.8	1.0	1.4	1.3	1.4
In person	5.9	4.9	5.5	6.1	4.3	5.5
Interview Conducted While Respondent Was at a Job						
Corps Center	8.9	8.0	8.5	0.6	0.4	0.6
Number of Respondents to the 12-Month Interview	4,710	3,583	8,293	3,274	1,816	5,090

SOURCE: 12-month follow-up interview data.

Most interview nonrespondents were youths who could not be located, although some were youths who were located but refused to complete the interview (Table A.4). Our survey staff were unable to locate about 86 percent of program group nonrespondents and about 81 percent of control group nonrespondents. The refusal rate was higher for control group than program group members (18 percent, compared to 12 percent), probably because some control group members, having been denied access to Job Corps, did not want to be part of the study. Interestingly, the difference in the refusal rates by research status was due to differences for males but not for females. Only a very small number of nonrespondents had partial interviews or were unavailable because they were incarcerated or in the military.

3. The 30-Month Interview

The sample of those who completed 30-month interviews was the primary analysis sample used in the 30-month impact report. Thus, obtaining sufficiently high response rates to the 30-month interview was crucial for obtaining credible estimates of the short-term impacts of Job Corps on key participant outcomes.

We completed 30-month interviews with 11,787 of the 14,671 youths released for 30-month interviews. For those in the in-person areas only, we completed 8,257 of the 10,405 interviews attempted, resulting in an effective response rate of 79.4 percent (80.7 percent program, 77.4 percent control).⁸ About 96 percent of those who completed the 30-month interview also completed the 12-month interview. In addition, about 98 percent completed the full baseline interview; the remaining 2 percent were “combo” cases who did not complete the full baseline interview but completed the

⁸The response rates exclude the program readmits and those who died.

TABLE A.4

REASONS FOR NONCOMPLETION OF THE 12-MONTH INTERVIEW,
BY RESEARCH STATUS AND GENDER
(Percentages)

Reasons for Noncompletion	Program Group			Control Group		
	Males	Females	Total	Male	Females	Total
Unable to Locate	88.4	82.0	85.9	80.0	84.1	81.1
Refusal	8.5	18.1	12.2	18.2	15.3	17.5
Incarcerated and Unavailable	0.2	0.0	0.1	0.0	0.0	0.0
In Military and Unavailable	2.0	0.0	1.2	1.5	0.0	1.1
Break-Off or Partial Interview	0.2	0.0	0.1	0.2	0.0	0.0
Other	0.7	0.0	0.4	0.0	0.6	0.3
Number of Nonrespondents to the 12-Month Interview	447	277	724	455	163	618

SOURCE: 12-month follow-up interview data.

NOTE: The following cases were excluded from the calculation: (1) 39 cases (21 program group and 18 control group members) who were confirmed to have died, and (2) 77 cases (40 program group and 37 control group members) who were determined to have enrolled in Job Corps prior to random assignment.

abbreviated baseline interview as part of the 12-month interview. Thus, complete baseline and follow-up data are available for most youths in the 30-month sample.

The effective response rate to the 30-month interview was fairly high across all key youth subgroups, although there were some subgroup differences (Table A.5). The pattern of subgroup findings closely follows the pattern of findings for the 12-month interview. The response rate was higher for females than for males (84 percent, compared to 76 percent) and for younger sample members than for older ones (81 percent for those 16 and 17 years old, compared to 77 percent for those 22 and older). In addition, the response rate was about 6 percentage points higher for those who lived in less populated areas than for those who lived in more populated areas. Furthermore, it was slightly higher for (1) those who completed high school, (2) those never arrested or convicted, (3) those who lived with family members, (4) those with health problems, (5) those with children, and (6) likely nonresidential students than for their counterparts. There were few differences by race/ethnicity and region.

Because of these subgroup differences in response rates, sample weights for the 30-month interview sample were adjusted to help reduce the potential bias in the impact estimates due to interview nonresponse (see Appendix C). These adjusted weights were used to calculate all impact estimates.

Most interview respondents completed the 30-month interview soon after it was due to be completed (Table A.6). The average 30-month interview was completed 2.4 months after the 30-month release date (that is, in month 32.4 after random assignment), and about 70 percent of interviews were completed within 3 months of release (that is, between months 30 and 33). Less than 13 percent of interviews were completed after month 36. The distributions of completion times were similar for program and control group members. The fact that most interviews were conducted

TABLE A.5

EFFECTIVE RESPONSE RATES TO THE 30-MONTH FOLLOW-UP INTERVIEW,
BY RESEARCH STATUS AND KEY SUBGROUP

Subgroup	Effective Response Rate		
	Program Group	Control Group	Combined Sample
Full Sample	80.7	77.4	79.4
Demographic Characteristics			
Gender			
Male	77.9	74.3	76.3
Female	84.2	82.7	83.7
Age at Application			
16 to 17	81.5	79.6	80.7
18 to 19	79.9	77.4	78.9
20 to 21	81.2	75.5	78.9
22 to 24	79.5	72.4	76.8
Race/Ethnicity			
White, non-Hispanic	80.1	77.4	79.0
Black, non-Hispanic	80.7	78.0	79.6
Hispanic	80.1	75.3	78.1
Other	86.1	78.0	82.8
Region			
1	82.1	75.8	79.5
2	79.2	69.7	75.2
3	77.7	77.7	77.7
4	79.6	77.5	78.8
5	80.3	79.5	79.9
6	80.3	75.8	78.5
7/8	84.7	81.6	83.5
9	82.1	78.3	80.5
10	84.4	81.0	83.0
Size of City of Residence			
Less than 2,500	84.5	84.2	84.4
2,500 to 10,000	87.0	83.3	85.6
10,000 to 50,000	81.6	80.6	81.2
50,000 to 250,000	79.0	77.7	78.5
250,000 or more	79.8	75.0	77.8
PMSA or MSA Residence Status			
In PMSA	79.2	74.2	77.2
In MSA	80.6	79.0	80.0
In neither	86.1	83.2	85.0

TABLE A.5 (continued)

Subgroup	Effective Response Rate		
	Program Group	Control Group	Combined Sample
Density of Area of Residence			
Superdense	79.9	75.1	78.0
Dense	79.7	77.1	78.6
Nondense	83.6	82.9	83.3
Lived in Areas with a Large Concentration of Nonresidential Females			
Yes	81.5	77.9	79.9
No	80.1	77.0	78.9
Legal U.S. Resident			
Yes	80.8	77.4	79.4
No	75.6	82.4	78.1
Job Corps Application Date			
11/94 to 2/95	80.0	74.7	77.9
3/95 to 6/95	82.5	80.4	81.6
7/95 to 9/95	80.3	76.4	78.7
10/95 to 12/95	79.7	76.9	78.6
Fertility and Family Status			
Fertility			
Had dependents	82.6	81.2	82.0
Had no dependents	80.3	76.5	78.8
Family Status			
Family head	81.4	79.1	80.5
Family member	81.8	77.9	80.2
Unrelated individuals	77.9	74.9	76.7
Education			
Completed the 12th grade	83.0	80.1	81.9
Did not complete the 12th grade	80.1	76.8	78.8
Welfare Dependence			
Public Assistance			
Received AFDC	82.0	78.3	80.5
Received other assistance	81.0	78.0	79.8
Did not receive	80.0	76.8	78.7
Health			
Had Any Health Conditions That Were Being Treated			
Yes	85.6	79.8	83.4
No	80.8	77.4	79.4

TABLE A.5 (continued)

Subgroup	Effective Response Rate		
	Program Group	Control Group	Combined Sample
Crime			
Arrests			
Arrested in past three years	76.8	76.4	76.6
Not arrested in past three years	81.3	77.5	79.8
Convictions			
Ever convicted or adjudged delinquent	77.5	72.5	75.4
Never convicted or adjudged delinquent	81.0	77.6	79.6
Anticipated Program Enrollment Information			
Residential Designation Status			
Resident	80.1	76.2	78.5
Nonresident	82.8	82.1	82.5
CCC/Contract Center Designation^a			
CCC center	80.5	78.9	79.8
Contract center	80.6	78.0	79.5
Performance Level of Designated Center^a			
High or medium high	81.1	77.6	79.7
Medium low or low	80.3	78.6	79.6
Size of Designated Center^a			
Large or medium large	80.0	76.0	78.4
Medium small or small	80.9	79.3	80.3
Sample Size	6,182	4,223	10,405

SOURCE: ETA-652 and ETA-652 Supplement data.

- NOTE:
1. The effective response rate is the response rate for those sample members who were eligible for a baseline interview after 45 days after random assignment. These are youths who lived in randomly selected (in-person) areas at application to Job Corps.
 2. The following cases were excluded from the calculations: (1) 39 cases (21 program group and 18 control group members) who were confirmed to have died since their previous interview, and (2) 77 cases (37 control group and 40 program group members) who were determined to have enrolled in Job Corps prior to random assignment.

^a Figures are obtained using data on OA counselor projections about the centers that youths were likely to attend.

TABLE A.6

CUMULATIVE DISTRIBUTION OF THE NUMBER OF MONTHS BETWEEN 30 MONTHS
AFTER RANDOM ASSIGNMENT AND COMPLETION OF THE 30-MONTH INTERVIEW
FOR THOSE IN THE IN-PERSON AREAS, BY RESEARCH STATUS
(Percentages)

Number of Months	Program Group	Control Group	Combined Sample
-3 to 0 ^a	1.6	2.6	2.0
0 to .5	25.5	26.0	25.7
.5 to 1	16.0	15.1	15.7
1 to 2	16.6	15.8	16.3
2 to 3	10.2	10.0	10.1
3 to 4	7.3	7.4	7.3
4 to 5	6.2	5.9	6.1
5 to 6	3.9	4.9	4.3
6 to 12	11.6	11.2	11.4
12 to 15 ^b	1.2	1.1	1.1
Average Number of Months	2.5	2.4	2.4
Number of Respondents to the 30-Month Interview	4,988	3,269	8,257

SOURCE: 30-month follow-up interview data.

NOTE: The in-person areas are randomly selected areas in which youths were eligible for baseline interviews after 45 days after random assignment. Youths not in the in-person areas who did not complete baseline interviews within the 45-day period were not eligible for follow-up interviews.

^a Youths in the in-person areas who did not complete the 12-month interview within 27 months after random assignment but who were located before 30 months after random assignment were administered the 30-month interview.

^b Cases who were located between 45 and 48 months after random assignment were administered 48-month interviews.

quickly and that most 30-month respondents also completed 12-month interviews suggests that recall error did not have a large effect on item responses and that recall error did not differ substantially across sample members.

About 86 percent of interviews were completed by telephone in MPR's phone center (Table A.7). About 14 percent were conducted in the field (8 percent in person, 5 percent when the field interviewer had the youth call the MPR phone center, and 1 percent when the field interviewer called the youth). These figures are similar by research status and by gender. The proportion of interviews completed in the field was higher for the 30-month interview than for the 12-month interview because it was more difficult to locate and interview youths by phone at the 30-month interview. Only a small fraction (about 1 percent) of program group interviews were conducted while the youth was in a Job Corps center, because nearly all program group enrollees had already left Job Corps.⁹

Reasons for noncompletion of the 12-month and 30-month interviews were similar (Table A.8). Our survey staff were unable to locate about 80 percent of the 30-month nonrespondents (83 percent program and 79 percent control). The refusal rate to the 12-month and 30-month interviews were similar, and they remained higher for control group males than for program group males. Among male nonrespondents, nearly 3.5 percent did not complete the interview because they were in jail and unavailable, and an additional 1 percent were in the military and unavailable.

⁹We conducted (1) 21 interviews with youths who were living in a school or college, (2) 459 interviews with youths in jail (261 program and 198 control), (3) 27 interviews with youths living in halfway houses or residential treatment centers, (4) 93 interviews with youths in the military, (5) 22 interviews with youths in a group home, and (6) 27 interviews with homeless youths.

TABLE A.7

INTERVIEW MODE FOR CASES WHO COMPLETED THE 30-MONTH INTERVIEW,
BY RESEARCH STATUS AND GENDER
(Percentages)

Interview Mode	Program Group			Control Group		
	Males	Females	Total	Male	Females	Total
Telephone Center	85.6	87.6	86.5	84.4	87.9	85.7
In the Field	14.4	12.3	13.5	15.7	12.1	14.3
Interviewer called youth	1.5	1.3	1.4	1.0	1.0	1.0
Interviewer had youth use a cell phone to call the phone center	4.0	5.2	4.5	5.1	5.1	5.1
In person	8.9	5.9	7.5	9.6	6.0	8.2
Interview Conducted While Respondent Was at a Job Corps Center	1.2	1.0	1.1	0.5	0.4	0.5
Number of Respondents to the 30-Month Interview	4,028	3,283	7,311	2,811	1,665	4,476

SOURCE: 30-month follow-up interview data.

TABLE A.8

REASONS FOR NONCOMPLETION OF THE 30-MONTH INTERVIEW,
BY RESEARCH STATUS AND GENDER
(Percentages)

Reasons for Noncompletion	Program Group			Control Group		
	Males	Females	Total	Male	Females	Total
Unable to Locate	83.7	80.8	82.7	78.4	79.7	78.7
Refusal	10.6	17.5	12.9	16.7	19.6	17.4
Incarcerated and Unavailable	3.3	0.9	2.5	3.4	0.7	2.7
In Military and Unavailable	1.6	0.0	1.1	0.8	0.0	0.6
Break-Off or Partial Interview	0.5	0.7	0.6	0.4	0.0	0.3
Other	0.4	0.0	0.3	0.4	0.0	0.3
Number of Nonrespondents to the 30-Month Interview	1,052	542	1,594	829	276	1,105

SOURCE: 30-month follow-up interview data.

NOTE: The following cases were excluded from the calculations: (1) 39 cases (21 program group and 18 control group members) who were confirmed to have died since their previous interview, and (2) 77 cases (37 control group and 40 program group members) who were determined to have enrolled in Job Corps prior to random assignment.

APPENDIX B

THE TREATMENT OF MISSING VALUES AND OUTLIERS

A. INTRODUCTION

Three categories of outcome measures were constructed for the 30-month impact analysis: (1) education and training in Job Corps and elsewhere; (2) employment and earnings; and (3) nonlabor market outcomes, including the receipt of public assistance benefits, involvement with the criminal justice system, use of alcohol and illegal drugs, health, fertility, custodial responsibility for children, marital status, living arrangements, and mobility. The 30-month impact report describes the specific outcome measures used in the analysis, our reasons for selecting these measures, and our basic procedure for constructing them. This appendix discusses in more detail the construction of key outcome measures and examines the prevalence of missing values and outliers.

B. THE PREVALENCE OF MISSING VALUES

Table B.1 displays the proportion of the 30-month sample with nonmissing values for selected outcome measures. The figures are presented separately for program and control group members, and are presented for the full sample and by gender.

Data item nonresponse was uncommon for most outcome measures used in the 30-month impact analysis. Indicators of the occurrence of key events are rarely missing. For example, item nonresponse was typically less than 3 percent for indicators of (1) participation in Job Corps and other education and training programs (such as GED, high school, or vocational schools); (2) educational attainment (such as the receipt of GED and vocational trade certificates and highest grade completed); (3) employment and characteristics of the most recent job; (4) the receipt of various forms of public assistance benefits; (5) arrests, arrest charges, convictions, and incarcerations for convictions; (6) alcohol and various types of illegal drug use; (7) health status; (8) fertility; and (9) marital status and living arrangements.

TABLE B.1
 DATA ITEM RESPONSE FOR KEY OUTCOME MEASURES
 USED IN THE 30-MONTH IMPACT ANALYSIS,
 BY RESEARCH STATUS AND GENDER
 (Percentages)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Job Corps Experiences						
Enrolled in a Job Corps Center						
All months	98.2	99.4	98.7	NA	NA	NA
Quarter 1	96.9	98.8	97.7	NA	NA	NA
Quarter 5	98.4	98.9	98.6	NA	NA	NA
Quarter 10	97.6	97.9	97.7	NA	NA	NA
Months Between Random Assignment and Center Enrollment ^a	94.7	97.1	95.7	NA	NA	NA
Months Enrolled ^a	91.4	94.2	92.6	NA	NA	NA
Months Between Date Left Job Corps and the 30-Month Interview ^a	94.6	96.4	95.4	NA	NA	NA
Participated in Academic Classes or Vocational Training ^a	96.9	98.7	97.7	NA	NA	NA
Total Hours in Academic Classes and Vocational Training ^a	87.8	90.1	88.8	NA	NA	NA
Took Academic Classes ^a	97.2	98.8	97.9	NA	NA	NA
Total Hours in Academic Classes ^a	91.6	93.4	92.4	NA	NA	NA
Took Vocational Training ^a	97.2	98.9	97.9	NA	NA	NA
Total Hours in Vocational Training ^a	91.9	94.0	92.8	NA	NA	NA
Participation in Other Job Corps ^a Activities						
World of Work	96.4	98.0	97.1	NA	NA	NA
Progress/Performance Evaluation Panels	96.9	98.4	97.5	NA	NA	NA
Health Classes	96.8	97.9	97.3	NA	NA	NA
Parenting Skills Classes	97.4	98.9	98.0	NA	NA	NA
Social Skills Training	96.3	97.4	96.8	NA	NA	NA
Cultural Awareness Classes	96.3	97.7	96.9	NA	NA	NA
Alcohol and Other Drugs of Abuse Program	97.4	98.9	98.0	NA	NA	NA
Education and Training in Job Corps and Elsewhere						
Enrolled in a Program, by Period						
Ever during the 30 months	99.1	99.8	99.4	98.9	99.4	99.1
Quarter 1	96.9	98.4	97.6	97.0	97.8	97.3
Quarter 5	97.0	98.3	97.6	98.2	99.0	98.5
Quarter 10	97.2	98.0	97.6	98.9	99.2	99.0
Number of Programs Attended	98.0	98.5	98.2	95.6	96.9	96.1
Percentage of Weeks in Programs	88.0	91.4	89.5	92.0	93.7	92.6

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Hours per Week in Programs						
All months	85.7	89.3	87.3	89.6	90.9	90.0
Quarter 1	90.8	93.9	92.2	93.4	94.5	93.8
Quarter 5	93.3	94.8	94.0	95.8	96.2	96.0
Quarter 10	95.7	96.1	95.8	96.5	96.9	96.7
Attended Programs Other than Job Corps, by Type						
Any	98.7	99.4	99.0	98.5	99.0	98.7
High school ^b	97.8	97.6	97.8	97.8	97.8	97.8
ABE or ESL ^b	98.2	97.7	98.0	97.9	98.2	98.0
GED ^b	97.7	98.2	97.9	97.8	98.1	97.9
Vocational/technical school	99.2	98.9	99.1	98.5	99.3	98.8
Two-year college	99.4	98.9	99.2	98.6	98.9	98.7
Four-year college	99.3	98.8	99.1	98.7	98.8	98.7
Percentage of Weeks in Programs Other than Job Corps, by Type						
	97.4	98.1	97.8	96.1	96.8	96.4
Hours per Week in Programs Other than Job Corps, by Type						
Any	91.2	92.9	92.0	89.4	90.8	89.9
High school	94.2	95.1	94.6	93.6	94.8	94.1
GED	95.7	95.7	95.7	94.4	94.6	94.5
Vocational/technical school	98.2	97.9	98.0	96.9	98.2	97.4
Two-year college	99.0	98.6	98.8	98.2	98.2	98.2
Took Academic Classes	44.5	42.9	43.8	44.5	43.3	44.1
Weeks in Academic Classes	39.5	39.0	39.2	41.8	40.8	41.4
Hours per Week in Academic Classes	33.9	34.6	34.2	38.3	36.5	37.6
Took Vocational Training	44.8	43.3	44.2	45.6	44.1	45.1
Percentage of Weeks in Vocational Training						
	42.6	41.7	42.2	45.4	43.8	44.8
Hours per Week in Vocational Training						
	38.5	37.9	38.2	44.5	42.6	43.8
Degrees, Diplomas, and Certificates Received						
GED certificate ^b	99.4	99.3	99.4	99.3	99.5	99.3
High school diploma ^b	98.8	99.0	98.9	99.1	98.6	98.9
Vocational/technical certificate	99.4	99.5	99.4	99.5	99.6	99.6
College degree (two-year or four-year)	99.6	99.7	99.7	99.7	99.7	99.7
Highest Grade Completed at 30 Months	99.9	99.9	99.9	99.8	99.8	99.8
Employment and Earnings						
Employed, by Period						
Quarter 1	97.4	98.1	97.7	97.6	98.6	98.0
Quarter 5	98.2	98.9	98.5	98.5	98.7	98.5
Quarter 8	97.4	98.5	97.9	97.5	98.6	97.9
Quarter 10	98.8	99.0	98.9	98.4	99.3	98.8
Ever during the 30 months	100.0	99.8	99.9	99.9	99.9	99.9
Number of Jobs	100.0	99.8	99.9	99.9	99.9	99.9

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Percentage of Weeks Employed, by Period						
Quarter 1	94.8	96.5	95.6	95.9	97.0	96.3
Quarter 5	96.0	97.1	96.5	96.5	97.1	96.7
Quarter 8	96.4	97.9	97.1	96.8	98.3	97.3
Quarter 10	97.6	98.4	98.0	97.3	98.7	97.8
All months	89.3	93.2	91.1	90.5	93.3	91.5
Hours per Week Employed, by Period						
Quarter 1	93.5	95.4	94.4	94.2	95.5	94.7
Quarter 5	94.1	95.3	94.6	94.2	95.8	94.8
Quarter 8	94.9	96.4	95.6	94.8	97.2	95.7
Quarter 10	96.1	97.1	96.6	95.9	97.8	96.6
All months	86.6	90.4	88.3	87.4	90.6	88.6
Earnings per Week, by Period						
Quarter 1	92.9	94.8	93.8	93.5	94.5	93.9
Quarter 5	92.9	94.3	93.5	93.3	94.2	93.6
Quarter 8	94.1	95.9	94.9	94.4	96.8	95.3
Quarter 10	95.2	96.4	95.7	95.0	97.1	95.7
All months	84.3	88.1	86.0	85.1	87.7	86.1
Characteristics of the Most Recent Job in Quarter 10 for Those Employed						
Number of months on job	99.2	99.5	99.3	99.4	99.7	99.5
Usual hours worked per week	99.4	99.0	99.2	99.4	99.2	99.3
Hourly wage	98.3	97.9	98.1	98.3	98.1	98.2
Weekly earnings	98.3	97.9	98.1	98.3	98.1	98.2
Occupation	99.7	99.8	99.7	99.7	99.8	99.7
Type of employer	97.7	96.7	97.3	97.3	96.2	96.9
Fringe benefits available						
Health insurance	97.1	98.0	97.5	97.3	97.7	97.4
Paid sick leave	96.7	97.6	97.1	97.0	96.7	96.9
Paid vacation	97.5	98.6	98.0	97.9	98.2	98.0
Retirement or pension benefits	93.3	91.7	92.6	93.1	91.2	92.5
Employed or in an Education or Training Program, by Period						
Quarter 1	97.4	98.4	97.8	96.4	97.4	96.8
Quarter 5	97.6	98.6	98.1	97.9	98.5	98.1
Quarter 10	97.8	98.5	98.1	98.6	99.0	98.7
Percentage of Weeks in Any Activity	82.2	87.8	84.7	85.4	89.5	86.9
Hours per Week in Any Activity						
Quarter 1	86.4	90.2	88.1	88.8	90.9	89.6
Quarter 5	88.5	90.9	89.6	90.9	92.7	91.5
Quarter 10	93.2	94.7	93.9	94.6	96.4	95.2
All months	75.6	81.0	78.1	78.9	83.1	80.5
Receipt of Public Assistance						
Received AFDC/TANF, Food Stamps, SSI/SSA, or GA Benefits, by Period						
All months	96.0	98.8	97.2	95.5	98.4	96.6
Months 1 to 12	96.5	98.3	97.3	95.7	97.8	96.5

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Months 13 to 24	95.4	98.3	96.7	94.7	98.1	96.0
Months 25 to 30	97.3	98.8	98.0	97.1	98.6	97.7
Number of Months Received Benefits	96.0	98.8	97.2	95.5	98.4	96.6
Amount of Benefits Received	77.1	77.9	77.4	75.7	77.8	76.5
Received AFDC/TANF Benefits, by Period						
All months	96.7	98.9	97.7	96.4	98.7	97.2
Months 1 to 12	97.1	98.9	97.9	97.1	98.5	97.6
Months 13 to 24	96.6	98.8	97.6	96.3	98.0	96.9
Months 25 to 30	98.5	99.2	98.8	98.3	98.4	98.3
Number of Months Received AFDC/TANF Benefits	94.9	97.1	95.9	94.9	95.7	95.2
Amount of AFDC/TANF Benefits Received	84.2	84.7	84.4	84.1	84.3	84.1
Received Food Stamp Benefits, by Period						
All months	98.1	99.3	98.7	97.8	98.9	98.2
Months 1 to 12	98.7	99.1	98.9	98.4	98.7	98.5
Months 13 to 24	98.2	99.3	98.7	97.7	99.0	98.2
Months 25 to 30	98.7	99.2	98.9	98.6	99.2	98.8
Number of Months Received Food Stamp Benefits	98.1	99.3	98.7	97.8	98.9	98.2
Amount of Food Stamp Benefits Received	90.1	90.7	90.4	88.9	90.8	89.6
Received SSI/SSA Benefits	98.4	99.3	98.8	98.4	99.3	98.7
Number of Months Received SSI/SSA Benefits	98.1	99.1	98.5	98.2	98.8	98.4
Amount of SSI/SSA Benefits Received	96.5	97.9	97.2	97.0	97.8	97.3
Received GA Benefits	97.5	98.3	97.9	97.5	98.1	97.7
Number of Months Received GA Benefits	97.5	98.1	97.7	97.1	98.0	97.4
Amount of GA Benefits Received	97.2	97.8	97.5	96.8	97.7	97.1
Covered by Public Health Insurance						
At 12 months	92.7	97.5	94.8	93.2	97.0	94.6
At 30 months	94.6	98.7	96.5	95.5	97.8	96.4
Received WIC Benefits (for females only)	NA	99.3	99.3	NA	99.5	99.5
Number of Months Received WIC Benefits (for females only)	NA	98.3	98.3	NA	98.3	98.3
Lived in Public Housing						
At 12 months	98.3	98.4	98.3	97.2	98.0	97.5
At 30 months	98.0	98.9	98.4	98.1	98.4	98.2
Received UI Benefits	99.7	99.9	99.8	99.8	99.9	99.8
Number of Weeks Received UI Benefits	99.5	99.8	99.6	99.5	99.8	99.6
Amount of UI Benefits Received	99.5	99.7	99.6	99.5	99.6	99.6
Received Child Support	99.6	99.7	99.6	99.8	99.7	99.7
Amount of Child Support Received	99.4	98.4	98.9	99.7	98.9	99.4
Received Income from Friends	99.2	99.6	99.4	99.5	99.4	99.5
Amount of Income Received from Friends	96.3	95.9	96.2	97.0	95.7	96.5
Received Other Income	99.3	99.7	99.4	99.4	99.1	99.3

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Amount of Other Income Received	98.1	98.6	98.3	98.8	98.5	98.7
Involvement with the Criminal Justice System						
Arrested or Charged with a Delinquency or Criminal Complaint, by Period						
Months 1 to 12	98.4	99.4	98.8	99.3	99.1	99.2
Months 13 to 24	98.3	99.4	98.8	99.1	99.2	99.1
Months 25 to 30	98.1	99.4	98.7	99.0	99.2	99.0
All months	98.7	99.5	99.1	99.4	99.2	99.3
Number of Arrests	97.9	99.3	98.5	98.8	99.2	98.9
Months Until First Arrested	97.9	99.3	98.5	98.8	99.2	98.9
Most Serious Charge for Which Arrested	98.1	99.2	98.6	98.7	98.9	98.8
Arrested for:						
Murder or assault	97.1	98.6	97.8	97.3	98.7	97.8
Robbery	97.0	98.6	97.8	97.3	98.7	97.8
Burglary	97.1	98.6	97.8	97.4	98.7	97.9
Larceny, vehicle theft, or other property crimes	97.3	98.6	97.9	97.5	98.7	97.9
Drug law violations	97.0	98.6	97.8	97.4	98.7	97.9
Other personal crimes	97.0	98.7	97.8	97.3	98.7	97.8
Other miscellaneous crimes	97.4	98.7	98.0	97.6	98.8	98.1
Convicted, Pled Guilty, or Adjudged Delinquent						
Made a Deal or Plea-Bargained	97.3	98.7	97.9	97.6	98.4	97.9
Most Serious Charge for Which Convicted						
Convicted of:	97.6	98.8	98.1	97.9	98.8	98.2
Murder or assault	97.1	98.6	97.8	97.3	98.7	97.8
Robbery	97.0	98.6	97.8	97.3	98.7	97.8
Burglary	97.1	98.6	97.8	97.4	98.7	97.9
Larceny, vehicle theft, or other property crimes	97.3	98.6	97.9	97.5	98.7	97.9
Drug law violations	97.0	98.6	97.8	97.4	98.7	97.9
Other personal crimes	97.0	98.7	97.8	97.3	98.7	97.8
Other miscellaneous crimes	97.4	98.7	98.0	97.6	98.8	98.1
Served Time in Jail for Convictions						
Weeks Spent in Jail	96.6	98.8	97.6	97.0	98.9	97.7
Put on Probation or Parole	97.9	98.9	98.3	98.2	98.9	98.5
Tobacco, Alcohol, and Illegal Drug Use						
Smoked Cigarettes						
At 12 months	99.7	99.8	99.8	99.9	99.8	99.8
At 30 months	99.6	100.0	99.8	99.8	99.8	99.8
Consumed Alcoholic Beverages						
At 12 months	99.7	99.7	99.7	99.9	99.8	99.9

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Used Marijuana, Hashish, or Hard Drugs						
At 12 months	99.5	99.8	99.7	99.8	99.6	99.7
At 30 months	99.5	99.9	99.6	99.6	99.6	99.6
Used Marijuana or Hashish						
At 12 months	99.6	99.8	99.7	99.9	99.8	99.8
At 30 months	99.6	99.9	99.7	99.8	99.8	99.8
Used Hard Drugs						
At 12 months	99.6	99.8	99.7	99.8	99.6	99.7
At 30 months	99.7	99.9	99.8	99.8	99.7	99.8
Snorted Cocaine Powder						
At 12 months	99.7	99.8	99.8	99.9	99.8	99.8
At 30 months	99.6	99.9	99.8	99.7	99.7	99.7
Smoked Crack Cocaine or Freebased						
At 12 months	99.7	99.8	99.7	99.9	99.7	99.8
At 30 months	99.7	99.9	99.8	99.7	99.7	99.7
Used Speed, Uppers, or Methamphetamines						
At 12 months	99.6	99.8	99.7	99.9	99.7	99.8
At 30 months	99.6	99.9	99.8	99.7	99.7	99.7
Used Hallucinogenic Drugs						
At 12 months	99.6	99.8	99.7	99.9	99.8	99.9
At 30 months	99.6	99.9	99.7	99.7	99.7	99.7
Used Heroin, Opium, Methadone, or Downers						
At 12 months	99.7	99.8	99.7	99.9	99.8	99.9
At 30 months	99.6	99.9	99.8	99.7	99.7	99.7
Used Other Drugs						
At 12 months	99.6	99.8	99.7	99.9	99.8	99.8
At 30 months	99.6	99.9	99.7	99.7	99.6	99.7
Shot or Injected Drugs with a Needle or Syringe						
At 12 months	99.6	99.8	99.7	99.9	99.8	99.9
At 30 months	99.6	99.9	99.7	99.7	99.6	99.7
In Alcohol or Drug Treatment	99.3	99.7	99.5	99.7	99.4	99.6
Weeks in Alcohol or Drug Treatment	99.1	99.8	99.4	99.1	99.5	99.2
Health						
Health Status						
At 12 months	99.7	99.8	99.7	99.8	99.6	99.7
At 30 Months	99.6	99.9	99.7	99.8	99.7	99.7
Had Serious Physical or Emotional Problems That Limited the Amount of Work or Other Regular Activities That Could Be Done						
At 12 months	99.7	99.7	99.7	99.7	99.7	99.7
At 30 months	99.7	99.8	99.7	99.7	99.6	99.6

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Fertility, Marriage, and Living Arrangements						
Had New Children	99.4	99.8	99.6	99.6	99.8	99.7
Number of New Children	99.4	99.8	99.6	99.6	99.7	99.6
Had Children out of Wedlock	99.4	99.8	99.6	99.6	99.7	99.7
Pregnant at 30 Months (for females)	NA	99.1	99.1	NA	99.3	99.3
Lived with All Children ^c	97.5	99.7	98.9	98.4	99.4	98.9
Time Spent with Noncustodial Children ^d	96.0	98.1	96.4	97.4	94.2	97.0
Provided Support for Noncustodial Children ^d						
Any (such as food, toys, and money)	92.2	94.9	92.7	95.4	91.4	94.9
Money	92.2	96.1	93.0	95.4	92.8	95.1
Household Membership	98.2	98.7	98.4	98.6	99.0	98.8
Whether Youth Is the Household Head	99.2	99.5	99.3	99.4	99.7	99.5
Number in Household	98.9	99.3	99.1	99.2	99.6	99.4
Marital Status at 30 Months	99.7	99.9	99.8	99.8	99.9	99.8
Mobility						
Distance in Miles Between Zip Codes of Residence at Application to Job Corps and at the 30-Month Interview	96.7	98.1	97.3	96.5	97.7	97.0
Lived in Same State at Application to Job Corps and at the 30-Month Interview	96.7	98.1	97.3	96.5	97.7	97.0
Sample Size	4,028	3,283	7,311	2,811	1,665	4,476

SOURCE: Baseline, 12-month, and 30-month follow-up interview data for those who completed 30-month interviews.

NOTE: All figures are unweighted.

^aData pertain to program group members who enrolled in Job Corps.

^bData pertain to those without a high school credential at random assignment.

^cData pertain to those with children.

^dData pertain to parents who did not live with all their children.

NA = not applicable.

Missing values were somewhat more common for measures of *time* spent in key activities, because these measures were constructed using activity start and end dates, which sample members sometimes could not recall. Furthermore, data item nonresponse was more common for time measures covering longer periods than for those covering shorter periods. For example, the measures of quarterly hours employed were missing for about 5 percent of cases per quarter, whereas the measure of hours employed covering the entire 30-month period was missing for about 11 percent of cases.

Measures of the amount of benefits that were received from the main public assistance programs (AFDC/TANF and food stamps) were missing for about 10 to 16 percent of all cases, primarily because some recipients did not remember or know the average monthly benefit amount that they received during a particular welfare spell.

Measures pertaining to academic and vocational training experiences were missing for more than one-half of sample members because of a problem in the skip logic in the CATI program for the 30-month follow-up questionnaire. The error was corrected in April 1998, and thus the measures of academic and vocational training experiences are missing for about 55 percent of the 30-month sample who completed 30-month interviews before then. Impacts on the academic education and vocational training outcome measures were estimated using the sample of those interviewed after the error was corrected.¹⁰

Data item nonresponse did not differ by research status or by gender.

¹⁰The skip logic error affected program and control group members equally. Thus, the impact estimates on these outcomes are likely to be unbiased, although they may not be representative of all those in the study population.

C. THE TREATMENT OF MISSING VALUES AND OUTLIERS

In this section, we discuss the treatment of missing values and outliers for key outcome measures used in the 30-month impact analysis. We begin with a detailed discussion of our approach for addressing these issues for the employment and earnings outcomes. We then provide a briefer description of similar procedures that were used for the other two categories of outcome measures.

1. Employment and Earnings

We constructed the key employment and earnings outcome measures using a weekly employment timeline for each youth. We used the timelines to determine the jobs held by sample members in each week during the 30-month (130-week) follow-up period, and used job start and end dates to construct them. Positive integers were used to signify that the youth was employed in a week, and a blank code signified that the youth was not working. If the reported *day* the job started or ended was missing, we set the day to "15." However, if the month or year was missing, then the relevant timeline entries were set to "missing" (using alphabetic codes). A timeline entry could have multiple codes. For example, a code of "1B" signified that the youth was working on the first job reported in the survey--job 1--in that week, but also that we were unsure whether the youth was working on job 2. A code of "13" signified that the youth was employed in jobs 1 and 3; a code of "A" signified that we were unsure whether the youth was working on job 1, and so on.

Next, we describe our approach for constructing key employment-related outcome measures defined over specific periods: employment rates, weeks employed, hours employed, and earnings. We conclude with a brief discussion of the construction of variables describing the characteristics of the most recent job in quarter 10.

a. Employment Rates

Employment rates by quarter after random assignment were key outcome measures for the impact analysis. We calculated these rates using the employment timeline for each youth. For each quarter, we created an indicator variable that was set to “1” if the youth worked for at least 1 week during the quarter, “0” if the youth never worked and had no missing job codes, and to “missing” otherwise. The quarterly employment rates were calculated as the weighted average of these employment indicator variables.

The missing values in the employment rate measures were due primarily to missing job start and end dates. We did not impute missing values for these outcomes. Thus, the raw employment rate measures were used in the impact analysis.

b. Weeks Employed

The percentage of weeks employed in a quarter was also a key outcome measure for the impact analysis. This measure was constructed for each youth by dividing the number of weeks worked in the quarter by 13 (the number of weeks in a quarter). The number of weeks that a youth was employed was created by summing the weeks that the youth’s employment timeline had positive codes. The variable was set to “0” if the youth was not employed each week, and it was set to “missing” if *any* timeline entry had a missing code but no positive code (for example, the variable was set to “missing” if a code was “A” but would *not* have been set to missing if a code was “1B,” because the youth was known to have been working in Job 1).

Importantly, nearly all missing values for the measures of weeks employed were for youth *whom we knew worked*, but for whom we did not know for how long, because job start and end dates were missing. In contrast, variables for weeks worked were never missing for those who did not work, because they were set to “0” for these youths. Consequently, we were concerned that the

mean value for the variables for the number of weeks worked were biased downwards (because the variables contain “too many zeroes” or “too few positive values”) for *both* program and control group members. This problem could lead to biased impact estimates.

To address this concern, we used the following two steps to impute missing values for the time employed measures for those who we knew were employed:

1. We calculated the weighted mean number of weeks worked for those with positive values by gender, age, and race.
2. Workers with missing values were assigned the appropriate mean value according to their gender, age, and race.

The imputation procedure was performed *separately* for program and control group members.

This procedure is appealing, because the mean value of the adjusted weeks worked variable is equivalent to the product of (1) the proportion of those employed, and (2) the mean number of weeks worked for employed youths who originally had positive variable values. We refer below to this imputation procedure as the *zero-correction* imputation procedure.

It is noteworthy that we estimated impacts on the percentage of weeks employed by quarter using both the adjusted and unadjusted variables. As expected, the mean values for both the program and the control groups were higher using the adjusted measures, but the impact estimates were very similar. For example, in quarter 10, the average percentage of weeks employed using the adjusted measure was 55.7 percent for the program group and 53.8 percent for the control group (an impact of 1.9 percentage points). Using the unadjusted measure, the average percentage of weeks employed was 55.4 percent for the program group and 53.6 percent for the control group (also an impact of 1.9 percentage points). We present the impact estimates using the adjusted measures in the impact report.

c. Hours Employed per Week

In order to calculate hours employed measures, we constructed for each youth an hours timeline that covered the 130-week follow-up period. A timeline entry signified the total number of hours that a youth worked in all jobs during the week. We created the hours timelines using the employment timelines and survey information on the number of hours per week that employed youths usually worked on their jobs. A timeline entry in a given week was set to "missing" if the employment timeline had a missing job code in that week. For example, we set the variable to "missing" if we found a code of "A" or "1B" (because we were unsure whether the youth worked in job 2 and, hence, whether to include hours worked in job 2). Total hours worked in a week was topcoded at 84 (12 hours worked per day for 7 days).

Using a regression approach, we imputed missing values for the variable on the number of hours per week that the youth usually worked on a job.¹¹ For those with positive values, we regressed usual hours worked on a set of control variables (that included demographic characteristics and other features of the job--the hourly wage, occupation, and available fringe benefits) using ordinary least squares procedures.¹² Separate models were estimated for program and control group members. For missing cases, we computed predicted usual hours worked using the parameter estimates from the regression models. These predicted values were used in place of the missing values when we constructed the hours timelines.

The hours employed outcome measures were obtained using the hours timelines. To calculate hours worked over a given period, we summed across entries in the hours timeline. The measures were set to "missing" if the hours timeline had any missing entries over the period.

¹¹The usual hours worked variable was missing for about 1 percent of jobs.

¹²The regression R² values were about .12.

The raw hours employed measures were then adjusted using the zero-correction procedure to impute missing values for employed youths. These adjusted measures were used in the 30-month impact analysis.

d. Earnings

The earnings measures were constructed using a weekly earnings timeline for each youth. A timeline entry was calculated by (1) multiplying, for each job youth held during the week, the number of hours worked in the week and the hourly wage; and (2) summing these products over all jobs. The employment and hours timelines and hourly wage information were used to construct the earnings timelines. A timeline entry was set to "0" if the youth did not work in the week, and was set to "missing" if the relevant hours timeline entry was missing. However, a timeline entry was not set to "missing" if the hourly wage was missing, because missing hourly wages were imputed using the regression approach described above for imputing usual hours worked per week.^{13,14}

We hand-checked cases that reported hourly wages less than \$2.50 (about 2.5 percent of jobs) and greater than \$15 (about 1.2 percent of jobs). We looked at verbatim job descriptions and other job characteristics to determine whether outlier values were valid. About 85 percent of cases were determined to be valid.

We used several methods to treat hourly wages that we considered to be invalid in order to check the robustness of study findings. For example, (1) we imputed outliers using the regression model (which was our final approach), (2) we set outliers to missing, and (3) we set outliers less than \$2.50 to \$2.50 and outliers greater than \$15 to \$15. These procedures produced very similar impact estimates because of the small number of outliers.

¹³About 2 percent of jobs had missing wage information.

¹⁴The R² values from the wage regressions were about .18.

Earnings over a given period were calculated by summing across entries in the earnings timeline. Earnings were set to “0” for those who did not work during the period and to “missing” if any earnings timeline entry was missing during the period.

The earnings measures were then adjusted to impute missing values for workers using the zero-correction imputation procedure. In the 30-month impact report, we present estimated earnings impacts using the adjusted earnings measures. However, because earnings were the key outcome measure for the impact analysis, we estimated earnings impacts using various earnings constructs to test the sensitivity of study findings to alternative assumptions about how to treat missing values and outliers. As discussed, we constructed earnings measures using various assumptions about how to treat hourly-wage-rate outliers. In addition, we estimated impacts using adjusted earnings measures obtained using the zero-correction procedure and unadjusted measures. These procedures yielded very similar impact estimates. For example, the impact per eligible applicant on earnings per week in quarter 10 was \$12.9 (\$180.6 for the program group and \$167.7 for the control group) using the adjusted earnings measure. The impact was also \$12.9 using the unadjusted earnings measure, although as expected, earnings levels were slightly smaller for both research groups (\$178.7 for the program group and \$165.8 for the control group).

e. Characteristics of the Most Recent Job in Quarter 10

In the 30-month impact report, we present differences in the average characteristics of jobs held by program and control group members during quarter 10, including the hourly wage, job tenure, usual hours worked per week, weekly earnings, occupations, types of employers, and available fringe benefits. This analysis used information on the most recent job held by sample members during the 10th quarter after random assignment. The most recent job was identified by searching for the most

recent positive job code in the employment timeline between weeks 118 and 130. For ties, we selected the job that the youth had held the longest.

The outcomes describing the characteristics of the most recent job were conditional on having been employed in quarter 10. Thus, we did not impute missing values, because we did not have the “zero” problem discussed above. We treated outliers in hourly wage rates using the same procedures described above.

2. Education and Training

The procedures used to construct key education and training outcomes were very similar to those used to construct the employment-related outcomes. Using enrollment dates, we created weekly timelines signifying whether or not youths were enrolled in Job Corps or other education and training programs during each week of the follow-up period. These timelines were used to construct period-specific measures of participation in all education and training programs, participation in specific types of programs, and weeks spent in these programs.

The education and training timelines were also used along with information about usual hours per week spent in programs to construct weekly hours timelines.¹⁵ We used regression procedures to impute the small number of missing values for the variable on usual hours per week spent in programs.¹⁶ Weekly hours in the timelines were topcoded at 48 hours. Period-specific measures of hours spent in education and training programs were constructed using the hours timelines.

¹⁵We assumed that youths in Job Corps spent 40 hours per week in education and training.

¹⁶The control variables used in the regression models included demographic characteristics and other characteristics of the education or training program (such as the type of program and whether the youth took academic classes or vocational training). The regression R^2 values were about .13. About 1 percent of programs had missing values.

Cases with missing values for the measures on time spent in education and training programs were primarily those who we know participated in programs, but for whom there were missing program start and end dates. Thus, we used the zero-correction procedure to impute missing values for these program participants. Separate imputation procedures were performed for different types of programs. These adjusted measures were used in the 30-month impact analysis.

We also created a weekly timeline signifying whether or not the youth was in academic classes during each week of the follow-up period, and another signifying whether or not the youth was in vocational training. By applying the procedures described above to these timelines, we constructed measures of time spent in academic classes and vocational training.¹⁷

We did not impute missing values for outcomes pertaining to the receipt of degrees, diplomas, or certificates (for example, GED certificates, high school diplomas, vocational certificates, and college degrees). However, as discussed in the 30-month impact report, we constructed several measures of highest grade completed, because of inconsistencies in responses across interviews.

3. Nonlabor Market Outcomes

We constructed outcome measures on the receipt of public assistance benefits using very similar procedures to those used for the employment-related outcomes. We created monthly timelines on the receipt of various forms of public assistance benefits (AFDC/TANF, food stamps, General Assistance, SSI/SSA, WIC, and UI) and used these timelines to construct measures of participation in these programs. For those who received benefits, we used the zero-correction imputation procedure to impute missing values for the number of months that benefits were received.

To construct measures of the amount of benefits received, we used the welfare timelines and information on the monthly amount of benefits received for each spell of receipt. We used

¹⁷The academic and vocational training hours timeline entries were each topcoded at 48 hours.

regression procedures to impute missing benefit amounts for AFDC/TANF, food stamp, and SSI/SSA spells.¹⁸ The control variables used in the models included gender, age, household composition, fertility history, region of residence, and employment and earnings measures. We also identified outliers in usual monthly benefit amounts by hand-checking very large and very small values. We compared potential outliers with published statistics on monthly benefit amounts by household size, household composition, and state. We imputed outlier values using the regression models.

For the other nonlabor market outcomes, we did not adjust for missing values for any of the constructed *binary* (0/1) or categorical outcome measures. For example, we did not impute missing values for indicators of arrests, convictions, health status, marital status, or the presence of children. However, we used the zero-correction imputation procedure to impute missing *continuous* variables that were conditional on other variables. For example, we imputed missing values for the time spent in jail for those whom we know were incarcerated. Similarly, we imputed missing values for the time spent in drug or alcohol treatment for those whom we know were treated.

¹⁸The regression R^2 values were about .30 for the AFDC/TANF benefit amount models and about .10 for the food stamp and SSI/SSA benefit amount models. About 15 percent of AFDC/TANF spells and 8 percent of food stamp spells had missing benefit amounts.

APPENDIX C

THE CALCULATION OF SAMPLE WEIGHTS AND STANDARD ERRORS

A. INTRODUCTION

This technical appendix describes the calculation of sample weights that were used in the 30-month impact analysis to obtain unbiased estimates of program impacts that could be generalized to the study population. Sample weights were needed to account for the sample and survey designs and for interview nonresponse. This appendix also discusses procedures for constructing standard errors of the impact estimates, which were used to conduct tests of the statistical significance of the impact estimates.

B. CALCULATION OF SAMPLE WEIGHTS

For several reasons, youths in the study population had different probabilities of being included in the follow-up interview samples. First, youths had different probabilities of being assigned to the program and control groups, because sampling probabilities differed for various population subgroups. Second, as discussed in Appendix A, youths selected to the research sample had different probabilities of being included in the baseline interview sample, because (1) baseline interview attempts continued in the post-45-day period for sample members who lived in randomly selected areas only, and (2) youths in different types of areas (superdense, dense, and nondense) had different probabilities of being eligible for post-45-day baseline interviews. All youths in the selected in-person areas were eligible for follow-up interviews. However, only youths in the nonselected areas who completed baseline interviews within 45 days after random assignment were eligible for 12- or 30-month follow-up interviews.

Next, we discuss how sample weights were constructed to account for these design features. We conclude the section with a discussion of our approach for adjusting the weights to account for the effects of nonresponse to the follow-up interviews.

1. Weights to Account for the Sample Design

Groups of youths in the study population had different probabilities of being selected to the research sample. Table C.1 displays selection probabilities by research status for youths in those subgroups for which sampling rates were constant. The sampling rates to the control group are displayed by gender and by whether the youth lived in one of the 57 areas sending the largest number of nonresidential students to Job Corps.¹⁹ The sampling rates to the *program research* group are displayed by residential designation status obtained from the special study (ETA-652 Supplement) form. The control and *program research* group sampling rates are displayed also for youths who were sent for random assignment before and after August 16, 1995. This is because the probabilities that youths were assigned to the research sample were increased for likely nonresidential students at that time to compensate for the lower-than-expected flow of eligible applicants and the higher-than-expected program no-show rate during the first several months of sample intake.

The sampling probabilities displayed in Table C.1 were adjusted for the following sample members:

- Four youths in the *program research* group who were also randomly assigned to the program nonresearch group.²⁰ The selection probabilities for each of these youths is $2p$, where p is the relevant sampling probability from Table C.1 for each youth.

¹⁹Sampling rates were higher in these 57 areas to meet sample size targets for nonresidential students.

²⁰This occurred as the result of a small error in our random assignment program. Our computer program was designed to check whether each youth sent for random assignment had been previously randomly assigned and to randomly assign only new cases. However, our computer program did not check whether duplicate information on a youth was present *within* a batch of information sent to MPR for random assignment purposes. Once identified, this problem was corrected.

TABLE C.1

PROBABILITIES THAT ELIGIBLE APPLICANTS WERE SELECTED
TO THE CONTROL AND PROGRAM RESEARCH GROUPS,
BY SAMPLING STRATA
(Percentages)

	Sampling Probability	
	Random Assignment Date Before 8/16/95	Random Assignment Date on or After 8/16/95
Control Group		
Females in areas from which a low concentration of nonresidential Job Corps female students come	5	5
Females in 57 areas from which a high concentration of nonresidential Job Corps female students come	8	9
Males in areas from which a low concentration of nonresidential Job Corps female students come	8	8
Males in 57 areas from which a high concentration of nonresidential Job Corps female students come	8	9
Program Research Group		
Residential designees	10.7	11.1
Nonresidential designees	15.4	17.0
Number in Sample Universe	47,288	33,595

- 27 youths who were recruited by the Florida employment service office in Hialeah (FLESHI) and who were randomized to the research sample after March 27, 1995. A large proportion of youths recruited by FLESHI in early 1995 were assigned to the control group, and FLESHI staff expressed concern to Region 4 senior staff about the negative effects the evaluation was having on their reputation. To help smooth the flow of control group members who were recruited by FLESHI for the remainder of the sample intake period, all youths sent for random assignment after March 27, 1995, had the *same* probability of being assigned to the control group (and the same probability of being assigned to the program research group). Hence, all youths in a batch sent for random assignment were randomized together rather than in separate strata. The uniform sampling rates were set as the average of all the sampling probabilities of all FLESHI youths who were sent for random assignment prior to March 28, 1995. The sampling rates to the control group were set as follows: (1) 7.63 percent for those sent for random assignment between March 28, 1995, and August 15, 1995; and (2) 8.05 percent for those sent for random assignment after August 15, 1995. The sampling rates to the program research group were set as follows: (1) 11.62 percent for those sent for random assignment between March 28, 1995, and August 15, 1995; and (2) 12.04 percent for those sent for random assignment after August 15, 1995.

The sample design weight for a youth was constructed to be inversely proportional to the probability of selection to the research group to which the youth was selected.

2. Weights to Account for the Survey Design

In this section, we first discuss selection probabilities to the baseline interview sample. These probabilities are needed to construct the selection probabilities to the follow-up interview samples. Second, we discuss the selection probabilities to the 12- and 30-month interview samples, and the construction of weights that account for both the sample and survey designs.

a. Selection Probabilities to the Baseline Interview Sample

As discussed in detail in Appendix A, baseline interviews were attempted by telephone with all youths in the research sample during the first 45 days after random assignment. However, only youths in randomly selected areas who were not reachable by telephone within the 45-day period

were eligible for telephone or in-person interviews during the post-45-day period.²¹ To select these areas, we divided the country into 16 superdense, 29 dense, and 75 nondense areas. We then selected all 16 superdense, 18 dense, and 29 nondense areas as those where youths would be eligible for post-45-day interviewing. We selected different proportions of superdense, dense, and nondense areas for in-person interviewing to maximize the precision of the impact estimates, subject to the cost of conducting interviews in each type of area and a fixed interview budget.

The within-45-day sample is a random sample of those in the study population reachable by telephone within 45 days. The post-45-day sample, however, is a *clustered* sample of those in the study population reachable by telephone after 45 days. Thus, the post-45-day sample is underrepresented in the baseline sample relative to their numbers in the study population, and those in superdense, dense, and nondense areas have different representations in the post-45-day sample.

The probability that a youth was selected to the baseline interview sample was calculated by multiplying the probability the youth was selected into the research sample (as described above) by a factor f defined as follows:

- $f = 1$ if the youth completed a baseline interview within the first 45 days after random assignment
- $= 1$ if the youth lived in a superdense area at application to Job Corps
- $= 1$ if the youth was in the control group and was designated for a nonresidential slot on the Supplemental ETA-652 form
- $= 18/29$ if the youth completed a baseline interview between 45 and 270 days after random assignment and lived in a dense area at application to Job Corps

²¹Control group members designated for nonresidential slots on the Supplemental ETA-652 form, however, were eligible for post-45-day interviews regardless of where they lived. This design feature was adopted to increase the precision of impact estimates for the small nonresidential program component.

= 29/75 if the youth completed a baseline interview between 45 and 270 days after random assignment and lived in a nondense area at application to Job Corps

The factor f can be interpreted as the conditional probability that an eligible applicant was in the baseline sample given that the applicant was selected into the research sample.

b. Selection Probabilities to the 12- and 30-Month Follow-Up Interview Samples

As discussed, the following two groups of youths were eligible for 12-month interviews:

1. All youths in the randomly selected areas slated for in-person interviewing at baseline (whether or not they completed a baseline interview)
2. Youths not in the in-person areas at baseline who completed baseline interviews within 45 days after random assignment

Thus, selection probabilities to the 12-month interview sample were the *same* as selection probabilities to the baseline interview (ignoring the effects of interview nonresponse). The 300 youths in the in-person areas who completed the 12-month interview but not the full baseline interview were assigned the same selection probabilities to the 12-month sample as those who completed baseline interviews between 45 and 270 days after random assignment.

Selection probabilities to the 30-month interview were identical to the selection probabilities to the 12-month interview. It is noteworthy that we did not attempt 30-month interviews for those who did not complete either a baseline or a 12-month interview, because we expected very low 30-month completion rates for this group. However, ignoring the effects of nonresponse, the selection probabilities to the 12- and 30-month samples were theoretically equivalent.

The primary weights used in the 30-month impact analysis were adjusted for interview nonresponse (as discussed in the next section). However, to test the sensitivity of our estimates, we

also conducted the analysis using unadjusted weights, which were constructed to be inversely proportional to the selection probabilities to the 30-month sample. For both the program and control groups, the weights were scaled to sum to the size of the study population--80,883 eligible applicants.

3. The Adjustment of Weights to Account for Nonresponse to the 30-Month Interview

The main analysis sample for the 30-month impact analysis included the 11,787 youths (7,311 program group and 4,476 control group members) who completed 30-month interviews. The effective response rate (that is, the response rate in the in-person areas) to the 30-month interview was 79.4 percent (80.7 percent for the program group and 77.4 percent for the control group). Because more than one in five youths did not complete the interview, control group members in the analysis sample may not be fully representative of all control group members (respondents and nonrespondents), and the sample of program group members may not be fully representative of all program group members. If not corrected, the effects of interview nonresponse could lead to two problems:

1. *The impact estimates could be biased.* This would occur if the average baseline characteristics of control and program group members in the 30-month analysis sample differed.
2. *The impact estimates might not be generalizable to the study population.* This would occur if the average characteristics of respondents and nonrespondents differed (regardless of whether or not the average characteristics of program group and control group respondents were similar).

In this section, we assess the effects of nonresponse to the 30-month interview on estimated impacts during the 30 months after random assignment and discuss our approach for adjusting for these effects. As discussed in the impact report, the sample of those who completed the 12-month

interview was also used in the impact analysis to test the robustness of our findings. We used similar methods to adjust the 12-month and 30-month weights. Furthermore, the effective response rate to the 12-month interview was 90.2 percent (91.4 percent for the program group and 88.4 percent for the control group), so that potential effects of nonresponse were more serious for the 30-month sample than for the 12-month sample. Consequently, we do not present the results from the 12-month nonresponse analysis.

a. Assessing the Effects of Nonresponse

Our basic approach for assessing the effects of nonresponse was to compare the characteristics of respondents to the full sample of respondents and nonrespondents by using ETA-652 and ETA-652 Supplement data. These data were collected at program intake and thus were available for all interview respondents *and* nonrespondents. For the analysis, we selected data items that we believed were correlated with whether a youth was a respondent and with key study outcome measures. We did not use baseline interview data, because these data were not available for 30-month nonrespondents who did not complete the baseline interview.

The analysis was performed using *only* the 10,405 sample members who lived in the areas selected for in-person interviews at baseline. Youths in the nonselected areas were excluded from the analysis, because “nonrespondents” in these areas consisted of both those who would and those who would not have completed baseline interviews in the post-45-day period if given the chance. Therefore, “true” nonrespondents can be identified only in the selected areas. This sample of nonrespondents, however, is representative of nonrespondents nationwide. The analysis sample contains 8,257 respondents to the 30-month interview (3,269 control group and 4,988 program group members) and 2,148 nonrespondents (954 control group and 1,194 program group members).

We used standard statistical tests to assess the similarity of respondents and the full sample of respondents and nonrespondents in the in-person areas. We used univariate t-tests to compare variable means for binary and continuous variables and chi-squared tests to compare variable distributions for categorical variables.²² In addition, we conducted a more formal multivariate analysis to test the hypothesis that key variable means and distributions are *jointly* similar. For this analysis, we estimated logit regression models where the probability a person was a respondent versus a nonrespondent was regressed on a set of youth characteristics. Chi-squared (log-likelihood) tests were used to assess whether the explanatory variables in the models were jointly statistically significant.

There are some differences in the characteristics of respondents to the 30-month interview and the full sample of respondents and nonrespondents (Table C.2). For example, females and younger sample members were significantly more likely than their counterparts to complete an interview. In addition, response rates were significantly higher (1) for those in less populated areas than for those in more populated areas (such as PMSAs, MSAs, or superdense areas), (2) for those with children at program application than for those without children, (3) for those who had completed high school at program application than for those without a high school degree, (4) for those never convicted prior to application than for those convicted, and (5) for nonresidential designees than for residential designees. Furthermore, the explanatory variables in the logit models are jointly statistically significant at the 1 percent level of significance for both program and control group members.

²²The test statistics to test for differences between respondents and the full sample are the same as those to test for differences between respondents and nonrespondents only.

TABLE C.2

COMPARISON OF THE CHARACTERISTICS OF RESPONDENTS AND THE FULL SAMPLE
OF RESPONDENTS AND NONRESPONDENTS TO THE 30-MONTH INTERVIEW,
BY RESEARCH STATUS
(Percentages)

	Control Group		Program Group	
	Respondents	Respondents and Nonrespondents	Respondents	Respondents and Nonrespondents
Demographic Characteristics				
Male	55.1***	62.7	56.2***	56.2
Age at Application				
16 to 17	40.9***	39.9	40.3	38.8
18 to 19	32.0	31.8	31.7	32.1
20 to 21	16.5	16.8	16.4	16.7
22 to 24	10.7	11.5	11.6	12.5
(Average age)	18.9***	19.0	18.9	19.0
Race/Ethnicity				
White, non-Hispanic	24.2	23.8	24.2*	23.6
Black, non-Hispanic	54.6	54.9	55.3	56.0
Hispanic	16.3	16.8	15.5	15.9
Other	4.8	4.6	4.9	4.5
Region				
1	5.4**	5.3	5.5**	5.3
2	8.4	9.0	8.7	8.5
3	14.4	14.9	13.1	13.8
4	21.3	21.2	22.0	21.9
5	9.9	9.8	9.8	9.9
6	13.2	13.4	14.1	14.3
7/8	11.7	11.3	12.5	12.2
9	10.5	10.2	9.4	9.7
10	5.2	4.9	4.9	4.5
Size of City of Residence				
Less than 2,500	5.9***	5.2	5.6***	5.1
2,500 to 10,000	7.7	7.0	8.6	7.6
10,000 to 50,000	15.8	14.6	16.0	15.4
50,000 to 250,000	18.2	17.4	17.9	17.9
250,000 or more	52.4	55.7	51.9	54.0
PMSA or MSA Residence Status				
In PMSA	42.3***	44.4	44.2***	45.2
In MSA	43.7	43.5	41.7	42.3
In neither	13.9	12.1	14.1	12.6

TABLE C.2 (continued)

	Control Group		Program Group	
	Respondents	Respondents and Nonrespondents	Respondents	Respondents and Nonrespondents
Density of Area of Residence				
Superdense	48.1***	50.8	50.7***	52.7
Dense	26.6	26.5	25.1	24.9
Nondense	25.3	22.7	24.2	22.4
Lived in 57 Areas with a Large Concentration of Nonresidential Females				
	40.1	45.0	37.5	40.5
Legal U.S. Resident				
	98.7	98.8	98.7	98.6
Job Corps Application Date				
11/94 to 2/95	21.0**	21.3	22.7*	22.3
3/95 to 6/95	31.0	29.2	29.6	28.3
7/95 to 9/95	28.0	28.6	27.8	28.6
10/95 to 12/95	20.0	20.9	19.9	20.9
Fertility and Family Status				
Had Dependents				
	17.8***	16.9	15.3	17.3
Family Status				
Family head	14.5*	14.3	14.0**	15.3
Family member	61.7	61.4	61.6	60.0
Unrelated person	23.9	24.3	24.4	24.7
Average Family Size				
	3.2*	3.2	3.2**	3.2
Education				
Completed the 12th Grade				
	22.5***	21.2	21.7**	21.6
Welfare Dependence				
Public Assistance Receipt				
Received AFDC	28.4	28.3	28.4	29.4
Received other assistance	14.8	14.2	15.4	15.1
Did not receive	56.8	57.5	56.2	55.5
Health				
Had Any Health Conditions That Were Being Treated				
	3.3	3.1	3.6	3.5

TABLE C.2 (continued)

	Control Group		Program Group	
	Respondents	Respondents and Nonrespondents	Respondents	Respondents and Nonrespondents
Crime				
Arrested in Past Three Years	11.5	11.9	11.0***	11.3
Ever Convicted or Adjudged Delinquent	5.3**	5.9	5.5*	5.7
Completion Status to Previous Interviews				
Baseline Interview Completion Status				
Completed within 45 days	91.5***	88.4	91.3***	89.2
Completed between 46 and 270 days	5.5	5.9	6.0	6.6
Did not complete	3.0	5.7	2.7	4.2
Completed the 12-Month Interview	95.4***	88.4	96.0***	91.4
Anticipated Program Enrollment Information				
Designated for a Nonresidential Slot	20.8***	21.2	15.2*	20.4
Designated for a CCC Center ^a	12.6	12.5	12.8	12.2
Designated for a High- or Medium-High-Performing Center ^a	46.1	46.3	46.8	47.3
Designated for a Large or Medium-Large Center ^a	36.0***	36.7	37.4	37.1
Sample Size	3,269	4,223	4,988	6,182

SOURCE: 30-month follow-up interview, ETA-652 and ETA-652 Supplement data.

NOTES: 1. The figures are calculated for those sample members who were eligible for a baseline interview after 45 days after random assignment. These youths lived in randomly selected (in-person) areas at application to Job Corps.

2. All figures are calculated using sample weights to account for the sample and survey designs.

TABLE C.2 (continued)

3. The following cases in the in-person areas are excluded from the calculations: (1) 72 cases (32 control group and 40 program group members) who died between random assignment and the 30-month interview date, and (2) 63 cases (31 control group and 32 program group members) who were determined to have enrolled in Job Corps prior to random assignment.

^a Figures are obtained using data on OA counselor projections about the centers that youths were likely to attend.

*Difference between respondents and the full sample is significant at the .10 level, two-tailed test.

**Difference between respondents and the full sample is significant at the .05 level, two-tailed test.

***Difference between respondents and the full sample is significant at the .01 level, two-tailed test.

The findings are very similar for program and control group members. Thus, it does not appear that there were large differences in the average baseline characteristics of respondents by research status.

c. The Adjustment of the Weights

Because of the differences between the characteristics of respondents and nonrespondents, we adjusted the 30-month weights to account for the effects of nonresponse. The weights were adjusted so that the weighted baseline characteristics of interview respondents were similar, on average, to those of the full population of respondents and nonrespondents. To be sure, there may have been unmeasured differences between respondents and nonrespondents for which we cannot control. Consequently, our procedure cannot account for the full effects of interview nonresponse. However, because of the relatively large number of data items in the ETA-652 and ETA-652 Supplement forms, we believe that our procedure can account for some important differences between respondents and nonrespondents.²³

To construct the adjusted weights, we estimated models where the probability that a youth in the in-person areas completed the 30-month interview was regressed on a set of control variables. We estimated the models using logit maximum likelihood techniques and estimated separate models for program and control group members.

²³Sample selection statistical procedures could be used to account for both measured and unmeasured differences between respondents and nonrespondents. However, to implement these procedures effectively, we would have had to find at least one "instrumental" variable that is correlated with interview response status but uncorrelated with unobservable factors associated with the outcome measures. As is often the case, we were unable to find credible instrumental variables. Consequently, we did not correct for potential nonresponse bias using these sample selection procedures.

We used the following four steps to construct the adjusted weights:

1. ***A predicted probability (propensity score) was created for each respondent and nonrespondent using estimates from the "best" logit model.*** The best logit model included only control variables with predictive power in the regression models. The control variables for the model using program group members included 0/1 indicator variables signifying (1) gender, (2) race, (3) age, (4) region, (5) whether the youth was a family member or family head, (6) whether the youth lived in a PMSA or MSA, (7) the size of city of residence, (8) high school completion status, and (9) application date to Job Corps. The models using control group members included 0/1 indicator variables signifying (1) gender; (2) race; (3) age; (4) region; (5) whether the youth needed a bilingual program in Job Corps; (6) whether the youth lived in a superdense, dense, or nondense area at application; (7) the size of city of residence; (8) high school completion status; (9) residential designation status; and (10) application date to Job Corps.²⁴
2. ***Youths were divided into six groups on the basis of the size of their predicted probabilities.*** The first group consisted of the 5 percent of youths with the largest predicted probabilities, and the second group consisted of the 15 percent of youths with the next-highest predicted probabilities. The other four groups were divided by quintiles of the predicted probability distribution. For example, the third group consisted of those whose predicted probabilities were between the 60th and 80th percentiles of the predicted probability distribution, and the fourth group consisted of those between the 40th and 60th percentiles, and so on. Cluster analytic techniques were used to determine these groupings.
3. ***The weighted 30-month interview response rate was calculated for each of the six propensity score groups.*** The response rates ranged from about .70 to .88 for the program group, and .60 to .88 for the control group. The variation in the response rates suggests that the control variables had some predictive power in explaining whether or not a youth was an interview respondent.

²⁴We did not include indicator variables signifying baseline completion status in the final models, because the response rate to the 30-month interview was much higher for those who completed full baseline interviews than for those who did not (81 percent, compared to 46 percent). Thus, the coefficient estimates on the baseline completion variables were much larger than those of the other control variables. Consequently, the addition of the baseline completion variables would largely determine the nonresponse adjustments to the sample weights. We do not believe that the differences between respondents and nonrespondents can be captured primarily by whether a sample member completed the baseline interview within 45 days, after 45 days, or not at all. Thus, we did not include these variables in the final models.

4. *The adjusted weight for a youth was then constructed to be proportional to the product of the unadjusted weight and the inverse of the response rate in that youth's propensity score group.* The weights for both the control and program groups were scaled to sum to 80,883 (the size of the study population).²⁵

Using these adjusted weights, there were no differences between the observable characteristics of respondents and the full sample of respondents and nonrespondents for both research groups (not shown). The adjusted weights were the primary weights used to construct all impact estimates presented in the 30-month impact report.

C. CALCULATION OF STANDARD ERRORS

Standard errors of the impact estimates were used to test the statistical significance of program impacts. The construction of these standard errors is complicated, because they must account for design effects due to unequal weighting of the data and due to the clustered portion of sample caused by the random selection of areas for post-45-day interviewing at baseline.

In this three-part section, we discuss how we calculated standard errors for the impacts presented in the 30-month impact report. In the first section, we discuss the estimation of standard errors for impacts per eligible applicant (that is, for the difference between the weighted mean outcomes of program and control group members). Second, we discuss the estimation of standard errors for impacts per Job Corps participant only. Finally, we discuss how we conducted chi-squared

²⁵The 30-month sample contains youths who completed 30-month interviews but who were not in the in-person areas at baseline. These youths were not included in the sample used to estimate the logit models. However, we constructed weights for these youths by calculating predicted probabilities using the parameter estimates from the logit models, and assigned these youths to one of the six groups discussed above on the basis of the size of their predicted probabilities. Each of these youths was then assigned the response rate in the appropriate propensity score group (which was created using only those who lived in the in-person areas at baseline).

tests to test for differences in the distributions of categorical outcome measures across the program and control groups.

1. Standard Errors for Impacts per Eligible Applicant

The impact per eligible applicant on a binary or continuous outcome was calculated by comparing the weighted mean outcomes of program and control group members. To obtain an expression for the standard error of this impact estimate, it is instructive to first express the mean outcome of the program group (or the control group) as follows:

$$(1) \quad \bar{y} = \tau \bar{y}_l + (1-\tau)[\theta_s \bar{y}_{2s} + \theta_d \bar{y}_{2d} + \theta_n \bar{y}_{2n}],$$

where:

\bar{y} = the overall weighted mean of the variable

\bar{y}_l = the weighted mean (using the sample design weights) of those in the 30-month sample who completed baseline interviews within 45 days after random assignment

\bar{y}_{2s} , \bar{y}_{2d} , \bar{y}_{2n}
= the weighted mean (using the sample design weights) of those in superdense, dense, and nondense areas, respectively, who (1) completed a baseline interview in the post-45-day period, or (2) did not complete a baseline interview, but completed a 12-month interview--"combo"cases. These two groups are labeled the "post-45-day" group.

θ_s , θ_d , θ_n
= the proportion of the post-45-day population in superdense, dense, and nondense areas, respectively

τ = the proportion of all potential baseline interview completers who would have completed the baseline interview within 45 days after random assignment

In order to use equation (1), we assume that the weight, τ , is the proportion of baseline interview completers and combo cases in the in-person areas who completed the baseline interview within 45 days after random assignment (which is about 88 percent). This assumes that baseline

interview nonrespondents (except for combo cases) were split *proportionally* between the within-45-day and post-45-day populations. As discussed in Schochet (1998a), this is a reasonable assumption, because the characteristics at program intake of baseline interview nonrespondents, within-45-day responders, and post-45-day responders were similar.

The *variance* of the difference between the mean outcome of program and control group members can be written using equation (1) as follows:

$$(2) \quad \text{var}(\bar{I}) = \tau^2 \text{var}(\bar{I}_1) + (1 - \tau)^2 [\theta_s^2 \text{var}(\bar{I}_{2s}) + \theta_d^2 \text{var}(\bar{I}_{2d}) + \theta_n^2 \text{var}(\bar{I}_{2n})],$$

where \bar{I} represents the difference between the program and control group means, and where the other parameters and subscripts were defined above. The standard error of the impact estimate is the square root of the variance expression in equation (2).

Next, we discuss the estimation of each of the variance components in equation (2).

a. Variance Estimate of the Impact for the Within-45-Day Sample

Because the two samples are independent, the variance of the impact estimate for the within-45-day sample is simply the sum of the variances of the program and control group means. Thus, the following equation can be applied separately to each of the two groups:

$$(3) \quad \text{var}(\bar{y}_1) = (1 - g) \text{deff}_{w_1} \frac{\sigma_1^2}{n_1},$$

where:

σ_1^2 = variance of the outcome measure in the within-45-day population

g = proportion of the population that is sampled (which is assumed in all analyses to be the average sampling rates to the research sample--7.4 percent for control group members and 11.6 percent for program group members)

n_1 = within-45-day sample size

$deff_{w_1}$ = design effect due to unequal sample design weights (w) (which equals $n_1 \sum w^2 / (\sum w)^2$, and that is due to the fact that various population subgroups had different probabilities of being selected to the research sample)

An unbiased estimate of the unknown σ^2 is calculated in the usual way, and this estimate is inserted in place of σ^2 in equation (3).

b. Variance Estimate of the Impact for the Post-45-Day Sample in Superdense Areas

All 16 superdense areas were selected as in-person areas. Thus, the post-45-day sample in the superdense areas is a random (not clustered) sample. Thus, the same procedure as discussed for the within-45-day sample can be used to estimate the variance of the impact for the post-45-day sample in the superdense areas.

c. Variance Estimate of the Impact for the Post-45-Day Sample in Dense and Nondense Areas

Program and control group members in the post-45-day sample in dense or nondense areas may not be independent, because these youths were selected from the *same* areas. For example, the average characteristics of program and control group members who lived in the same areas may be correlated, because they may have faced similar local economic conditions and because people with similar characteristics tend to cluster in the same geographic areas. Thus, the average outcome measures for the two groups in the same area may be correlated.

The variance of the post-45-day impact in dense or nondense areas can be written as follows:

$$(4) \quad var(\bar{I}_2) = \left[\sigma_{2w}^2 \left[\frac{(1-g_d)}{n_{2c}a} + \frac{(1-g_p)}{n_{2p}a} \right] + \frac{(1-f)\sigma_{2b}^2}{a} \right] deff_{2w}$$

where the subscripts c and p refer to the control and program groups, a is the number of dense (or nondense) areas selected for post-45-day baseline followup, f is the fraction of all dense (nondense) areas selected for post-45-day baseline followup, n_{2c} and n_{2p} are post-45-day program and control group sample sizes per dense (nondense) area, $deff_{2w}$ is the design effect due to unequal weighting (see the definitions in equation (3) above), and where the subscripts denoting dense or nondense areas have been dropped for notational simplicity.

The term σ_{2b}^2 in equation (4) represents the variance of \bar{I} across areas. In other words, it represents the *extent to which the impacts varied across areas*. The term captures both the between-area variance in the mean measure as well as the correlation of the group means within areas. The term σ_{2w}^2 represents the variance of the measure within areas.

An unbiased estimate of the variance expression in equation (4) is as follows:

$$(5) \quad \text{var}(\bar{I}) = \left[(1-f) \frac{s_b^2}{a} + s_w^2 \left[\frac{f(1-g_c)}{n_c a} + \frac{f(1-g_p)}{n_p a} \right] \right] deff_{2w}$$

where s_b^2 is the sample variance of the impacts between areas, s_w^2 is the (average) sample variance of the measure across youths within areas, and other subscripts are omitted for notational simplicity.

Because of small sample sizes, it is problematic to estimate the sample variance terms in equation (5) using post-45-day sample members only. This is because the response rate to the baseline interview was extremely high within the first 45 days after random assignment (89 percent) and only an additional 9 percent of the research sample in the in-person areas completed baseline interviews in the post-45-day period or were combo cases. Hence, the post-45-day sample is small. The 30-month sample contains only 157 post-45-day sample members (94 program and 63 control group members) who lived in the 18 selected dense areas and 170 post-45-day sample members (99

program and 71 control groups members) who lived in the 29 selected nondense areas. Hence, there were very few sample members in most of the selected dense and nondense areas, and there were none in several areas. Thus, the between-area and within-area variance estimates in the dense and nondense areas (that is, s_b^2 and s_w^2) would be imprecise if the post-45-day sample were used in the calculations.

To address this problem, we calculated the variance terms in the dense (and nondense) areas using the following two steps:

1. We estimated s_b^2 and s_w^2 in dense (nondense) areas using both the *within-45-day* and *post-45-day* samples who lived in the selected dense (nondense) areas.
2. Using the estimated variances in step (1), we calculated equation (5) using *post-45-day* sample sizes.

This procedure assumes that the between-area and within-area variance estimates are similar for the within-45-day and post-45-day populations. This assumption cannot be reliably tested, because of small post-45-day sample sizes. However, we believe that it is sufficiently accurate and that our procedure yields more reliable variance estimates than those that would be obtained using only the post-45-day samples in the calculations.

An estimate of the total variance of the impact estimate, that is, of the expression in equation (2), can then be calculated using the estimated variances for the within-45-day and post-45-day samples. Design effects were estimated by dividing this total variance estimate by an unbiased estimate of the variance of a simple random sample of the same size.

The total design effect for most measures based on the full baseline interview sample was about 1.08. Nearly all of the design effect was due to unequal sample weights. For two main reasons, only a small portion of the total design effect was due to clustering of the post-45-day sample. First, the

clustered portion of the sample in the dense and nondense areas was very small because of high baseline interview response rates within 45 days after random assignment. Second, impact estimates did not vary substantially across dense and nondense areas.

2. Standard Errors for Impacts per Job Corps Participant

In the 30-month impact report, we present estimated impacts per eligible applicant as well as estimated impacts per Job Corps participant. We obtained the impact per participant on an outcome measure by dividing the estimated impact per eligible applicant by the proportion of program group members who enrolled in Job Corps. In mathematical terms, the estimated impact per participant (I_p) can be expressed as follows:

$$(6) \quad I_p = \frac{I}{S},$$

where I is the estimated impact per eligible applicant and S is the Job Corps participation (show) rate among the program group.

The variance of I_p must account for *both* the variance of I and the variance of S , because both of these values were estimated from the sample. We used standard ratio estimator techniques to estimate the variance of the estimated impact per participant. Using a Taylor series approximation, we can write the variance of I_p as follows:

$$(7) \quad \text{var}(I_p) = \text{var}([I - I_p S]/S),$$

where I_{p0} is the “true” but unknown impact on participants, and where S_0 is the true but unknown show rate. Using the definition of the variance of the sum of two random variables, equation (7) yields the following expression:

$$(8) \quad \text{var}(I_p) = [\text{var}(I) + I_{p0}^2 \text{var}(S) - 2I_{p0} \text{cov}(I, S)] / S_0^2.$$

Equation (8) can be computed using the following procedure:

1. Replace I_{p0} by the estimated impact per participant, I_p , using equation (6).
2. Replace S_0 by the estimated show rate, S .
3. Calculate $\text{var}(S)$ using program group members and the techniques for obtaining a standard error of a variable mean, as discussed in Schochet (1998a).
4. Note that the covariance of I and S , $\text{cov}(I, S) = \text{cov}(\bar{y} - \bar{z}, S) = \text{cov}(\bar{y}, S)$, where \bar{y} is the mean outcome measure for program group members and \bar{z} is the mean outcome measure for control group members. Ignoring design effects due to clustering, the covariance term, $\text{cov}(\bar{y}, S)$, can be estimated using the program group as follows:

$$\text{cov}(\bar{y}, S) = (1-g) \sigma_{y,S} \sum w_i^2 / (\sum w_i)^2,$$

where w_i is the weight for the i^{th} program group member, g is the proportion of the study population that was sampled to the program group, and where:

$$\sigma_{y,S} = \sum w_i (y_i - \bar{y})(S_i - S) / \sum w_i$$

In this expression, y_i is the outcome for the i^{th} program group member, and S_i is 1 if the youth enrolled in Job Corps and zero otherwise.

The calculated t-statistics to test the statistical significance of the impacts per eligible applicant and the impacts per participant were nearly identical for all outcome measures. Thus, we draw the same conclusions about statistical significance for both sets of impact estimates. The results are so

similar because the estimation error in the show rate was very small as a result of the large program group sample size. Thus, the estimated show rate could almost be treated as a constant.

3. Significance Tests for Impacts on the Distribution of Categorical Variables

Thus far, we have discussed the construction of standard errors for binary and continuous variables. However, in the 30-month impact report, we also presented impacts on categorical variables (for example, the type of living arrangement at the 30-month interview or categories of total earnings over the 30-month period). To assess the statistical significance of these impact estimates, we used a modified chi-squared statistic to test whether the distribution of the categorical variables differed across the program and control groups. This test statistic was constructed by dividing the usual chi-squared statistic (appropriately weighted) by the average design effect across each level of the categorical variable (Scott and Rao 1981). We calculated this average design effect in two steps. First, using the methods from the previous section, we calculated the design effect for comparing the difference between group proportions for *each level* of the categorical variable. Second, we took a weighted average of these design effects.

Formally, we used the following equations to construct the chi-squared statistic:

$$(9) \quad \chi_{SR}^2 = \frac{\chi_w^2}{\bar{d}},$$

$$(10) \quad \chi_w^2 = \sum_{i=1}^2 \sum_{j=1}^J \frac{(n_i p_{ij} - n_i p_j)^2}{n_i p_j},$$

$$(11) \quad p_j = \frac{n_1 p_{1j} + n_2 p_{2j}}{n_1 + n_2},$$

$$(12) \quad \bar{d} = \frac{1}{(J-1)} \sum_{j=1}^J (1 - p_j) d_j,$$

where p_{ij} is the proportion of youths in group I who are in category j , n_i is the number of youths in group I , p_j is the proportion of the study population in category j , and d_j is the design effect for category j as described above. Under the null hypothesis of no difference between group distributions, the chi-squared statistic is distributed chi-squared with $(J-1)$ degrees of freedom.

The modified chi-squared test statistic is intuitive. The statistic decreases as the average design effect increases. Thus, the hypothesis of no difference between group proportions is rejected less often as the average design effect (that is, the average variance across the categories) increases.

APPENDIX D

THE ESTIMATION OF REGRESSION-ADJUSTED IMPACTS

A. INTRODUCTION

Many impact analysts report regression-adjusted impact estimates when using a random assignment design to evaluate the effectiveness of an intervention. Simple differences in the mean outcomes of program (treatment) and control group members yield unbiased estimates of program impacts in these evaluations. However, estimating impacts from multivariate models that control for other factors that affect the outcome measures can increase the precision of the estimated program impacts and the power of significance tests. In addition, the models can adjust for any random residual differences in the observable baseline characteristics of program and control group members.

As discussed in Appendixes A and C, the sample and survey designs for the National Job Corps Study are complex. It is fairly straightforward under this design to estimate program impacts that can be generalized to the study population using the simple differences-in-means estimation approach. Furthermore, because the 30-month analysis sample is large (about 7,300 program and 4,500 control group members), the impact estimates for the full sample and most key subgroups are relatively precise. However, it is much more difficult to obtain unbiased impact estimates using the regression approach, because of the large number of weighting cells (sampling strata). Thus, while the regression approach may increase the precision of the impact estimates relative to the simple differences-in-means approach, these efficiency gains may be offset by the difficulty in obtaining regression-adjusted impact estimates that are unbiased and that can be generalized to all eligible applicants in the study population.

This appendix compares impact estimates on key outcomes using the regression and differences-in-means approaches and discusses our reasons for presenting the differences-in-means estimates in the 30-month impact report. The appendix is in four sections. First, we discuss impact estimation

issues that account for the study design. Second, we discuss the control variables that were included in the regression models. Third, we present impact estimates and their standard errors on key outcome measures using the two approaches. Finally, we present our conclusions.

1. Impact Estimation Issues

As discussed in Appendix C, youths had different probabilities of being included in the follow-up interview samples, for two reasons:

1. Selection probabilities to the program research and control groups differed for various population subgroups.
2. For the baseline interview, only youths in randomly selected areas who could not be interviewed by telephone within 45 days after random assignment were eligible for telephone or in-person interviews during the post-45-day period. Furthermore, youths in different areas (superdense, dense, and nondense) had different probabilities of being eligible for post-45-day interviewing. Follow-up interviews were not attempted for those in the nonselected areas who did not complete baseline interviews within 45 days after random assignment.

This design yields 48 weighting cells (that is, strata with unique program research and control group probabilities of being included in the follow-up interview samples).²⁶

As discussed in Appendix C, it is straightforward to estimate unbiased program impacts using the differences-in-means approach, because sample weights can be used to account for the design features discussed above. The use of sample weights ensures that the weighted distributions of the outcomes of control group members are representative of the outcomes of those in the study

²⁶There are 16 cells based on the sample design, because sampling rates differed by gender, residential/nonresidential designation status, whether the case lived in one of the 57 heavily nonresidential areas, and time period. Within each of the 16 cells, there are 3 cells due to the survey design defined by (1) cases who completed baseline interviews within the 45-day period and cases in superdense areas who completed baseline interviews in the post-45-day period, (2) those in dense areas who completed baseline interviews in the post-45-day period, and (3) those in nondense areas who completed baseline interviews in the post-45-day period.

population if they had been assigned to the control group, and similarly for the weighted outcomes of program group members. In the 30-month impact analysis, the weight for a youth was constructed to be inversely proportional to the probability that the youth was included in the 30-month follow-up interview sample. The weights were also adjusted for the effects of nonresponse to the follow-up interviews. The estimation of standard errors of the impact estimates accounted for design effects due to unequal weighting of the data and clustering of the post-45-day sample.

Obtaining regression-adjusted impact estimates that account for the study design is more complex. The usual regression model where the outcome measures are regressed on a program status indicator variable (which is 1 for program group members and 0 for control group members) and other control variables can yield biased estimates of program impacts (that is, biased coefficient estimates on the program status indicator variable) because the estimates may be “weighted” incorrectly. Furthermore, estimating weighted regressions using the sample weights described above does not solve the problem (DuMouchel and Duncan 1983). To obtain unbiased impact estimates, separate regression-adjusted estimates must be obtained in each of the 48 weighting cells (many of which contain only a small number of sample members), and the weighted average of these 48 separate estimates must be calculated.

Specifically, unbiased regression-adjusted impacts can be obtained using the following procedure:

1. Define the 48 cells with unique pairs of control and program research group weights and assign each sample member to a weighting cell
2. Estimate regression-adjusted impacts and standard errors within each of the 48 cells
3. Obtain the overall regression-adjusted impacts as a weighted average of the regression-adjusted impacts in each cell, where a cell weight is the proportion of the study population within that cell

4. Use a similar procedure to obtain the overall standard errors of the impact estimates

This procedure is straightforward if there are few cells. For example, if the sampling rates to the control and program research groups differed only by gender (and if there were no clustering of the post-45-day baseline interview sample), then there would be only two cells. Regression-adjusted impacts could then be obtained by estimating separate models for males and females, and by taking a weighted average of the regression-adjusted impacts for males and females.

In the Job Corps study design, however, there are 48 potential cells, and 45 of them contain at least one sample member. Furthermore, there are many cells with few sample members. Having small numbers of sample members in some weighting cells necessitates aggregating across weighting cells, which could introduce some bias if impacts differ across the cells.

We estimated regression-adjusted impacts using four cells defined by gender and residential/nonresidential designation status. This grouping captures the key features of the sample design, and the sample sizes in each cell were large enough to facilitate subgroup analyses.²⁷ In addition, the impacts on key outcomes across the other weighting strata did not appear to differ substantially.²⁸

²⁷The 30-month sample contains 6,225 male residents (2,592 controls), 614 male nonresidents (219 controls), 3,380 female residents (1,150 controls), and 1,568 female nonresidents (515 controls). The population weights were .55, .04, .31, and .10, respectively.

²⁸We estimated separate models for the four cells (that is, a fully interacted model), because the parameter estimates on the control variables differed somewhat across the four cells. The use of F-tests led to the rejection of the hypothesis that the parameter estimates across the four groups were similar for several models that we estimated using different outcome measures.

2. Selecting Control Variables

The following two main criteria were used to select the control variables that we included in the regression models:

1. ***The variables should be "baseline" measures that pertain to the period prior to random assignment.*** Thus, the control variables were constructed using data from the baseline interview, program intake (ETA-652) forms, and special study (Supplemental ETA-652) forms. Potential control variables were those discussed in the report describing the baseline characteristics of youths served by Job Corps (Schochet 1998b), and in the report containing methodological appendixes on sample implementation and baseline interviewing (Schochet 1998a). In general, the control variables were binary. For example, we constructed 0/1 indicator variables for several groups defined by age, race and ethnicity, and months worked in the year prior to random assignment.²⁹
2. ***The variables should have predictive power in regression models for key outcomes.*** For simplicity, the same set of variables was used to estimate impacts for all outcome measures. Thus, we selected a core set of control variables that were statistically significant in most (but not necessarily all) models.

Stepwise regression and other exploratory data-analytic methods were used to select the control variables. These methods were used to select variables that had predictive power in regression models for the following 12 key outcome measures that span the range of outcomes examined in the impact analysis:

1. Average earnings in quarter 10 (that is, months 28 to 30 after random assignment)

²⁹If a control variable was missing for less than 5 percent of cases, we replaced the missing values with mean values for the nonmissing cases by age, gender, and race/ethnicity. If a control variable was missing for more than 5 percent of cases, we constructed a missing indicator variable which was set to 1 for missing cases and 0 for nonmissing cases. In this case, the missing values for the original variable were set to 0 if the data item was a binary variable, but they were set to the mean value for the nonmissing cases if the data item was continuous. These rules were applied separately to data items that referred to all sample members (for example, whether the case ever worked or had a high school diploma), and to those that referred only to certain sample members (for example, the number of arrests for those ever arrested and the number of jobs for those who worked in the prior year).

2. Total earnings during the 30-month period
3. Proportion of weeks worked in quarter 10
4. Average hours employed per week in quarter 10
5. Whether employed in quarter 10
6. Whether a GED was obtained (for those without a high school credential at random assignment)
7. Average hours per week spent in education and training programs during the 30-month period
8. Average months received AFDC/TANF benefits during the 30-month period
9. Average months received food stamp benefits during the 30-month period
10. Whether ever arrested during the 30-month period
11. Whether ever in jail during the 30-month period
12. Whether ever had a child during the 30-month period

Ordinary least squares (OLS) methods were used to estimate models for the continuous outcome measures (for example, average earnings in quarter 10). To estimate models for binary dependent variables (for example, whether the youth was ever arrested or had a child), we used both OLS (linear probability) and logit maximum likelihood methods. These models produced very similar results; we present the OLS results.

Table D.1 displays the list of control variables that were selected. The categories of variables include demographic characteristics, fertility and living arrangements, education and training experiences, employment and earnings, public assistance receipt, arrest experience, drug use, and health.

TABLE D.1

CONTROL VARIABLES INCLUDED IN THE REGRESSION MODELS TO OBTAIN
REGRESSION-ADJUSTED IMPACT ESTIMATES

Demographic Characteristics

Age at Application to Job Corps

16 to 17

18 to 19

20 to 24

Race/Ethnicity

White non-Hispanic

Black non-Hispanic

Hispanic

American Indian, Alaskan Native, Asian, or Pacific Islander

Job Corps Region of Residence

1

2

3

4

5

6

7/8

9

10

PMSA or MSA Residence Status

In PMSA

In MSA

In neither

Lived in One of 57 Areas Sending a Large Number of Nonresidential Females to Job Corps

Job Corps Application Date

11/94 to 2/95

3/95 to 6/95

7/95 to 9/95

10/95 to 12/95

Completed the Baseline Interview More Than 45 Days After Random Assignment

TABLE D.1 (continued)

Fertility and Living Arrangements at the Baseline Interview

Had Own Children

Lived with Spouse or Partner

Education and Training Experiences Prior to Random Assignment

Had High School Diploma (not GED)

Had GED Certificate

Months in Education or Training in the Past Year

0

1 to 6

6 to 12

Missing months in school

Employment and Earnings Prior to Random Assignment

Ever Worked

Employed in the Past Year

Months Employed in the Past Year

0 to 3

3 to 9

9 to 12

Missing months employed

Earnings in the Past Year (in Dollars)

Less than 1,000

1,000 to 5,000

5,000 to 10,000

10,000 or more

Missing earnings in the past year

Currently Employed

TABLE D.1 (continued)

Public Assistance Receipt Prior to Random Assignment

Received AFDC in the Past Year and a Missing Indicator Variable

Received Food Stamps in the Past Year and a Missing Indicator Variable

Lived in Public Housing

Family Was on Welfare for Most of the Time When Youth Was Growing Up

Arrest Experience, Drug Use, and Health Prior to Random Assignment

Ever Arrested

Smoked Marijuana or Hashish in the Past Year

Used Hard Drugs in the Past Year

Ever in Drug Treatment

Had Physical or Emotional Problems That Limited the
Amount of Work That Could Be Done

SOURCE: Baseline interview and ETA-652 data.

NOTE: Separate regressions were estimated for the following four groups: (1) males designated for residential slots, (2) males designated for nonresidential slots, (3) females designated for residential slots, and (4) females designated for nonresidential slots. Thus, control variables signifying gender and residential/nonresidential designation status were not included in the models.

3. Estimation Results

The regression R^2 values for the continuous variables were about .10 for the quarter 10 employment and earnings measures, .20 for the total earnings measure, and .15 for the measure on time spent in education and training. The R^2 values for the welfare receipt measures were nearly .40 for females but only .10 for males. Thus, except for the welfare receipt measures for females, the control variables explained only a small portion of the variance of the outcome measures. These findings suggest that the regression-adjusted approach does not substantially increase the precision of the impact estimates relative to the differences-in-means approach.

Tables D.2 to D.9 display estimated impacts per eligible applicant for the 12 outcome measures using the differences-in-means and regression approaches for the total sample and for key youth subgroups. The table also displays estimated standard errors of the impact estimates, the percentage reduction in the standard errors from using the regression approach, and p-values from t-tests to gauge the statistical significance of the impacts. The results are displayed for the total sample and for the following key youth subgroups: (1) males and females; (2) age at application to Job Corps (16 and 17, 18 and 19, and 20 to 24); and (3) residential and nonresidential designees.

The impact estimates are very similar using the two approaches. In addition, the p-values to test the statistical significance of the impacts are very similar. The reductions in the standard errors using the regression approach are small except for the welfare measures. Consequently, the same policy conclusions can be drawn using the two approaches for the full sample and for key population subgroups (including the small subgroups such as nonresidential designees).

Despite the similarity of the results using the two approaches, it is noteworthy that the impact estimates using the two approaches generally vary *more* than the standard errors. For example, the impacts on average hours employed per week in quarter 10 differ by about 10 percent, whereas the

TABLE D.2

IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES
FOR THE FULL SAMPLE

Outcome Measure	Differences-in-Means Approach				Regression Approach			
	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Percentage Reduction in the Standard Error	
Average Earnings per Week (in 1998 Dollars)								
Quarter 10	12.9	3.75	0.001***	12.6	3.58	0.000***	4.4	
Entire 30-month period	-7.4	2.15	0.001***	-8.1	1.96	0.000***	8.9	
Average Percentage of Weeks Employed in Quarter 10	1.9	0.86	0.030**	1.8	0.85	0.032**	1.2	
Average Hours Employed Per Week in Quarter 10	1.0	0.43	0.019**	0.9	0.42	0.026**	2.0	
Percentage Employed in Quarter 10	2.1	0.90	0.023**	1.9	0.90	0.031**	0.3	
Received a GED Certificate ^a	17.6	0.97	0.000***	17.0	0.97	0.000***	0.2	
Average Hours per Week Ever in Education or Training	5.6	0.15	0.000***	5.7	0.15	0.000***	0.6	
Average Number of Months Received AFDC/TANF Benefits	-0.2	0.15	0.134	-0.2	0.11	0.043**	21.2	
Average Number of Months Received Food Stamp Benefits	-0.4	0.16	0.006***	-0.5	0.13	0.000***	19.3	
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-4.4	0.81	0.000***	-3.9	0.78	0.000***	4.8	
Percentage Served Time in Jail for Convictions	-2.8	0.62	0.000***	-2.4	0.60	0.000***	2.9	
Percentage Had New Children	-0.3	0.81	0.751	-0.3	0.81	0.746	-0.5	
Sample Size	11,787			11,787				

TABLE D.2 (continued)

SOURCE: Baseline, 12-month, and 30-month follow-up interview data for those who completed 30-month interviews.

NOTE: The differences-in-means impact estimates are measured as the difference between the weighted means for program and control group members. Standard errors of these estimates account for design effects due to unequal weighting of the data and clustering caused by the selection of areas slated for in-person interviewing at baseline. The regression-adjusted impact estimates were obtained in two steps. First, separate regressions were estimated for the following four groups: (1) males designated for residential slots, (2) males designated for nonresidential slots, (3) females designated for residential slots, and (4) females designated for nonresidential slots. Each regression model included an indicator variable signifying whether the youth was in the program or control group and other control variables. In the second stage, a weighted average of the regression-adjusted impact estimates for each of the four groups was calculated.

^aFigures pertain to those without a high school credential at random assignment.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

TABLE D.3

IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES
FOR MALES

Outcome Measure	Differences-in-Means Approach				Regression Approach			
	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Standard Error	Percentage Reduction in the Standard Error
Average Earnings per Week (in 1998 Dollars)								
Quarter 10	13.1	5.23	0.012**	12.2	5.07	0.017**		3.0
Entire 30-month period	-9.2	2.96	0.002***	-10.9	2.77	0.000***		6.7
Average Percentage of Weeks Employed in Quarter 10	1.4	1.09	0.204	1.1	1.08	0.295		0.9
Average Hours Employed per Week in Quarter 10	0.9	0.58	0.111	0.7	0.57	0.192		1.3
Percentage Employed in Quarter 10	1.8	1.13	0.112	1.4	1.13	0.215		-0.1
Received a GED Certificate ^a	16.8	1.21	0.000***	16.6	1.21	0.000***		0.1
Average Hours per Week Ever in Education or Training	5.7	0.20	0.000***	5.7	0.20	0.000***		0.7
Average Number of Months Received AFDC/TANF Benefits	-0.2	0.11	0.051*	-0.2	0.10	0.073*		4.0
Average Number of Months Received Food Stamp Benefits	-0.4	0.13	0.001***	-0.4	0.13	0.000***		3.1
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-5.6	1.14	0.000***	-4.8	1.13	0.000***		1.6
Percentage Served Time in Jail for Convictions	-3.6	0.93	0.000***	-3.1	0.93	0.001***		0.2
Percentage Had New Children	-1.4	0.95	0.129	-1.3	0.96	0.168		-1.2
Sample Size	6,839			6,839				

TABLE D.3 (continued)

SOURCE: Baseline, 12-month, and 30-month follow-up interview data for those who completed 30-month interviews.

NOTE: The differences-in-means impact estimates are measured as the difference between the weighted means for program and control group members. Standard errors of these estimates account for design effects due to unequal weighting of the data and clustering caused by the selection of areas slated for in-person interviewing at baseline. The regression-adjusted impact estimates were obtained in two steps. First, separate regressions were estimated by residential designation status. Each regression model included an indicator variable signifying whether the youth was in the program or control group and other control variables. In the second stage, a weighted average of the regression-adjusted impact estimates for each of the two groups was calculated.

^aFigures pertain to those without a high school credential at random assignment.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

TABLE D.4

IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES
FOR FEMALES

Outcome Measure	Differences-in-Means Approach			Regression Approach			
	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Percentage Reduction in the Standard Error
Average Earnings per Week (in 1998 Dollars)							
Quarter 10	12.9	4.93	0.009***	13.2	4.80	0.006***	2.6
Entire 30-month period	-4.5	2.90	0.123	-4.2	2.64	0.111	9.0
Average Percentage of Weeks Employed in Quarter 10	2.6	1.38	0.058*	2.8	1.36	0.039**	1.1
Average Hours Employed per Week in Quarter 10	1.2	0.63	0.062*	1.2	0.63	0.052*	0.5
Percentage Employed in Quarter 10	2.5	1.47	0.094*	2.7	1.46	0.066*	0.5
Received a GED Certificate*	18.8	1.65	0.000***	17.7	1.62	0.000***	1.7
Average Hours per Week Ever in Education or Training	5.6	0.25	0.000***	5.6	0.25	0.000***	0.8
Average Number of Months Received AFDC/TANF Benefits	-0.3	0.30	0.333	-0.3	0.24	0.211	20.8
Average Number of Months Received Food Stamp Benefits	-0.5	0.32	0.095*	-0.5	0.26	0.045**	19.4
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-2.3	0.97	0.019**	-2.5	0.97	0.010**	-0.1
Percentage Served Time in Jail for Convictions	-1.4	0.60	0.021**	-1.4	0.60	0.016**	-1.0
Percentage Had New Children	1.3	1.40	0.358	1.3	1.42	0.370	-1.1
Sample Size	4,948			4,948			

TABLE D.4 (continued)

SOURCE: Baseline, 12-month, and 30-month follow-up interview data for those who completed 30-month interviews.

NOTE: The differences-in-means impact estimates are measured as the difference between the weighted means for program and control group members. Standard errors of these estimates account for design effects due to unequal weighting of the data and clustering caused by the selection of areas slated for in-person interviewing at baseline. The regression-adjusted impact estimates were obtained in two steps. First, separate regressions were estimated by residential designation status. Each regression model included an indicator variable signifying whether the youth was in the program or control group and other control variables. In the second stage, a weighted average of the regression-adjusted impact estimates for each of the two groups was calculated.

^aFigures pertain to those without a high school credential at random assignment.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

TABLE D.5

IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES
FOR 16- AND 17-YEAR-OLDS

Outcome Measure	Differences-in-Means Approach				Regression Approach			
	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Percentage Reduction in the Standard Error	
Average Earnings per Week (in 1998 Dollars)								
Quarter 10	20.4	5.42	0.000***	20.5	5.30	0.000***	2.3	
Entire 30-month period	3.1	3.00	0.295	1.8	2.76	0.503	8.2	
Average Percentage of Weeks Employed in Quarter 10	3.7	1.31	0.005***	3.8	1.33	0.005***	-1.5	
Average Hours Employed per Week in Quarter 10	2.1	0.65	0.001***	1.9	0.65	0.004***	0.5	
Percentage Employed in Quarter 10	3.9	1.43	0.006***	4.1	1.46	0.005***	-2.4	
Received a GED Certificate ^a	16.5	1.32	0.000***	16.1	1.33	0.000***	-0.7	
Average Hours per Week Ever in Education or Training	4.1	0.24	0.000***	4.3	0.24	0.000***	-1.0	
Average Number of Months Received AFDC/TANF Benefits	-0.4	0.20	0.040**	-0.3	0.19	0.164	5.0	
Average Number of Months Received Food Stamp Benefits	-0.5	0.21	0.025**	-0.4	0.20	0.086*	3.3	
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-3.8	1.36	0.006***	-3.8	1.29	0.003***	5.1	
Percentage Served Time in Jail for Convictions	-3.6	1.06	0.001***	-3.4	1.01	0.001***	5.3	
Percentage Had New Children	-0.2	1.24	0.847	-0.3	1.29	0.810	-3.4	
Sample Size	4,863			4,863				

TABLE D.5 (continued)

SOURCE: Baseline, 12-month, and 30-month follow-up interview data for those who completed 30-month interviews.

NOTE: The differences-in-means impact estimates are measured as the difference between the weighted means for program and control group members. Standard errors of these estimates account for design effects due to unequal weighting of the data and clustering caused by the selection of areas slated for in-person interviewing at baseline. The regression-adjusted impact estimates were obtained in two steps. First, separate regressions were estimated for the following four groups: (1) males designated for residential slots, (2) males designated for nonresidential slots, (3) females designated for residential slots, and (4) females designated for nonresidential slots. Each regression model included an indicator variable signifying whether the youth was in the program or control group and other control variables. In the second stage, a weighted average of the regression-adjusted impact estimates for each of the four groups was calculated.

^aFigures pertain to those without a high school credential at random assignment.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

TABLE D.6

IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES
FOR 18- AND 19-YEAR-OLDS

Outcome Measure	Differences-in-Means Approach			Regression Approach			Percentage Reduction in the Standard Error
	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Estimated Impact per Eligible Applicant	Standard Error	P-Value	
Average Earnings per Week (in 1998 Dollars)							
Quarter 10	-1.7	6.89	0.807	-0.3	6.75	0.967	2.1
Entire 30-month period	-19.6	3.84	0.000***	-19.6	3.63	0.000***	5.4
Average Percentage of Weeks Employed in Quarter 10	-1.5	1.52	0.313	-1.4	1.53	0.375	-0.1
Average Hours Employed per Week in Quarter 10	-0.7	0.78	0.386	-0.7	0.77	0.384	0.7
Percentage Employed in Quarter 10	0.0	1.59	0.978	0.2	1.59	0.901	0.1
Received a GED Certificate*	20.3	1.82	0.000***	19.4	1.85	0.000***	-1.4
Average Hours per Week Ever in Education or Training	6.3	0.26	0.000***	6.5	0.26	0.000***	-1.6
Average Number of Months Received AFDC/TANF Benefits	0.0	0.25	0.844	0.0	0.20	0.819	20.3
Average Number of Months Received Food Stamp Benefits	-0.1	0.28	0.653	-0.1	0.23	0.592	18.3
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-5.5	1.40	0.000***	-4.6	1.35	0.001***	3.1
Percentage Served Time in Jail for Convictions	-2.2	1.07	0.040**	-1.6	1.04	0.117	2.5
Percentage Had New Children	-0.5	1.49	0.723	-1.0	1.50	0.503	-1.2
Sample Size	3,724			3,724			

TABLE D.6 (continued)

SOURCE: Baseline, 12-month, and 30-month follow-up interview data for those who completed 30-month interviews.

NOTE: The differences-in-means impact estimates are measured as the difference between the weighted means for program and control group members. Standard errors of these estimates account for design effects due to unequal weighting of the data and clustering caused by the selection of areas slated for in-person interviewing at baseline. The regression-adjusted impact estimates were obtained in two steps. First, separate regressions were estimated for the following four groups: (1) males designated for residential slots, (2) males designated for nonresidential slots, (3) females designated for residential slots, and (4) females designated for nonresidential slots. Each regression model included an indicator variable signifying whether the youth was in the program or control group and other control variables. In the second stage, a weighted average of the regression-adjusted impact estimates for each of the four groups was calculated.

^a Figures pertain to those without a high school credential at random assignment.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

TABLE D.7

IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES
FOR 20- TO 24-YEAR-OLDS

Outcome Measure	Differences-in-Means Approach			Regression Approach			
	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Percentage Reduction in the Standard Error
Average Earnings per Week (in 1998 Dollars)							
Quarter 10	17.7	7.45	0.018**	14.1	7.66	0.066*	-2.8
Entire 30-month period	-9.9	4.44	0.025**	-12.6	4.39	0.004***	1.2
Average Percentage of Weeks Employed in Quarter 10	2.9	1.64	0.080*	2.3	1.73	0.190	-5.5
Average Hours Employed per Week in Quarter 10	1.2	0.84	0.137	0.8	0.89	0.348	-5.7
Percentage Employed in Quarter 10	1.5	1.66	0.354	0.8	1.76	0.659	-6.0
Received a GED Certificate ^a	16.0	2.34	0.000***	17.2	2.57	0.000***	-10.2
Average Hours per Week Ever in Education or Training	7.2	0.31	0.000***	7.3	0.32	0.000***	-4.6
Average Number of Months Received AFDC/TANF Benefits	-0.1	0.33	0.676	-0.2	0.22	0.308	30.8
Average Number of Months Received Food Stamp Benefits	-0.8	0.37	0.026**	-1.0	0.29	0.000***	23.0
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-3.8	1.35	0.005***	-3.2	1.43	0.024**	-6.2
Percentage Served Time in Jail for Convictions	-2.1	0.99	0.033**	-1.1	1.07	0.286	-7.6
Percentage Had New Children	0.1	1.51	0.931	0.8	1.61	0.641	-6.6
Sample Size	3,200			3,200			

TABLE D.7 (continued)

SOURCE: Baseline, 12-month, and 30-month follow-up interview data for those who completed 30-month interviews.

NOTE: The differences-in-means impact estimates are measured as the difference between the weighted means for program and control group members. Standard errors of these estimates account for design effects due to unequal weighting of the data and clustering caused by the selection of areas slated for in-person interviewing at baseline. The regression-adjusted impact estimates were obtained in two steps. First, separate regressions were estimated for the following four groups: (1) males designated for residential slots, (2) males designated for nonresidential slots, (3) females designated for residential slots, and (4) females designated for nonresidential slots. Each regression model included an indicator variable signifying whether the youth was in the program or control group and other control variables. In the second stage, a weighted average of the regression-adjusted impact estimates for each of the four groups was calculated.

^a Figures pertain to those without a high school credential at random assignment.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

TABLE D.8

IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES
FOR RESIDENTIAL DESIGNNEES

Outcome Measure	Differences-in-Means Approach			Regression Approach			
	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Percentage Reduction in the Standard Error
Average Earnings per Week (in 1998 Dollars)							
Quarter 10	13.5	4.14	0.001***	12.7	3.95	0.001***	4.7
Entire 30-month period	-7.9	2.37	0.001***	-8.9	2.16	0.000***	8.8
Average Percentage of Weeks Employed in Quarter 10	1.5	0.94	0.113	1.3	0.93	0.163	1.5
Average Hours Employed per Week in Quarter 10	1.0	0.48	0.039**	0.8	0.47	0.069*	2.4
Percentage Employed in Quarter 10	1.8	0.99	0.072*	1.5	0.98	0.122	0.6
Received a GED Certificate ^a	18.1	1.06	0.000***	17.6	1.06	0.000***	0.1
Average Hours per Week Ever in Education or Training	5.7	0.17	0.000***	5.7	0.17	0.000***	0.7
Average Number of Months Received AFDC/TANF Benefits	-0.3	0.13	0.035**	-0.2	0.12	0.053*	11.8
Average Number of Months Received Food Stamp Benefits	-0.6	0.15	0.000***	-0.5	0.14	0.000***	9.6
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-4.9	0.92	0.000***	-4.5	0.87	0.000***	5.2
Percentage Served Time in Jail for Convictions	-2.9	0.70	0.000***	-2.5	0.68	0.000***	3.7
Percentage Had New Children	0.1	0.88	0.944	0.2	0.88	0.831	-0.6
Sample Size	9,605			9,605			

TABLE D.8 (continued)

SOURCE: Baseline, 12-month, and 30-month follow-up interview data for those who completed 30-month interviews.

NOTE: The differences-in-means impact estimates are measured as the difference between the weighted means for program and control group members. Standard errors of these estimates account for design effects due to unequal weighting of the data and clustering caused by the selection of areas slated for in-person interviewing at baseline. The regression-adjusted impact estimates were obtained in two steps. First, separate regressions were estimated by gender. Each regression model included an indicator variable signifying whether the youth was in the program or control group and other control variables. In the second stage, a weighted average of the regression-adjusted impact estimates for each of gender group was calculated.

^aFigures pertain to those without a high school credential at random assignment.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

TABLE D.9

IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES
FOR NONRESIDENTIAL DESIGNEES

Outcome Measure	Differences-in-Means Approach			Regression Approach			
	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Estimated Impact per Eligible Applicant	Standard Error	P-Value	Percentage Reduction in the Standard Error
Average Earnings per Week (in 1998 Dollars)							
Quarter 10	9.2	8.38	0.274	11.9	8.19	0.145	2.3
Entire 30-month period	-4.5	4.95	0.368	-3.3	4.44	0.454	10.3
Average Percentage of Weeks Employed in Quarter 10	4.2	2.02	0.040**	5.1	2.03	0.013**	-0.6
Average Hours Employed per Week in Quarter 10	1.2	0.96	0.226	1.5	0.96	0.114	0.1
Percentage Employed in Quarter 10	3.7	2.11	0.079*	4.5	2.15	0.037**	-1.8
Received a GED Certificate ^a	14.0	2.39	0.000***	13.4	2.46	0.000***	-3.0
Average Hours per Week Ever in Education or Training	5.5	0.36	0.000***	5.6	0.37	0.000***	-0.7
Average Number of Months Received AFDC/TANF Benefits	0.2	0.50	0.736	-0.2	0.38	0.509	25.3
Average Number of Months Received Food Stamp Benefits	0.4	0.53	0.397	-0.2	0.39	0.612	25.9
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-1.6	1.53	0.287	-0.1	1.48	0.961	3.2
Percentage Served Time in Jail for Convictions	-2.2	1.07	0.035**	-1.6	1.06	0.122	0.5
Percentage Had New Children	-2.2	2.02	0.273	-3.1	2.08	0.141	-3.0
Sample Size	2,182			2,182			

TABLE D.9 (continued)

SOURCE: Baseline, 12-month, and 30-month follow-up interview data for those who completed 30-month interviews.

NOTE: The differences-in-means impact estimates are measured as the difference between the weighted means for program and control group members. Standard errors of these estimates account for design effects due to unequal weighting of the data and clustering caused by the selection of areas slated for in-person interviewing at baseline. The regression-adjusted impact estimates were obtained in two steps. First, separate regressions were estimated by gender. Each regression model included an indicator variable signifying whether the youth was in the program or control group and other control variables. In the second stage, a weighted average of the regression-adjusted impact estimates for each of gender group was calculated.

^aFigures pertain to those without a high school credential at random assignment.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

standard errors differ by only about 2 percent. This finding contributes to our fear that the regression-adjusted approach may yield impact estimates that are slightly biased for the reasons discussed above.

4. Conclusions

On the basis of this analysis, we used the differences-in-means estimates as our benchmark estimates, for four main reasons. First, the gains in precision using the regression approach are small in general. In addition, because sample sizes are large, most impact estimates using the differences-in-means approach are relatively precise.

Second, because of large sample sizes, there are very few differences in the average baseline characteristics of program research and control group members (as discussed in Schochet 1998a), so that controlling for these differences in a regression does not materially affect the estimates.

Third, we can fully account for the complex study design using the differences-in-means approach by using sample weights, so that we are confident that these estimates are unbiased and can be generalized to the study population (that is, are externally valid). As discussed, it is more difficult to account for the complex study design using the regression approach. The finding that the impact estimates using the two approaches typically differ more than the standard errors contributes to our concerns about the bias in the regression-adjusted estimates.

Finally, we can adjust for potential survey nonresponse bias using the differences-in-means approach by adjusting the weights. A similar approach in the regression context would create an even larger number of weighting cells, which would add to the estimation problem. Furthermore, adjusting for potential nonresponse bias using sample selection correction models would be difficult because we have no credible “instrumental” variables that are correlated with response status but uncorrelated with unobservable factors associated with the outcome measures.

We conclude by restating our finding that the two approaches yield very similar conclusions about the short-term impacts of Job Corps for the full sample and for key youth subgroups. This result increases our confidence about the robustness of the short-term impact findings.

APPENDIX E

THE ADJUSTMENT FOR CROSSOVERS

About 1.4 percent of control group members enrolled in Job Corps before their three-year restriction period ended. These “crossovers” were treated as control group members in the analysis to preserve the integrity of the random assignment design. Consequently, impact estimates that do not account for these crossovers could be biased downwards if crossovers benefited from participation in Job Corps.

The 30-month impact report described statistical procedures that were used to estimate impacts per eligible applicant and impacts per program participant that do not account for control group crossovers. Impacts *per eligible applicant* were estimated by comparing the distribution of outcomes for *all* program and control group members. This procedure generates unbiased estimates, because random assignment was performed at the time applicants were determined to be eligible for Job Corps. Impacts *per participant* were estimated by dividing the impacts per eligible applicant by the proportion of program group members who enrolled in Job Corps (73 percent). These estimates are unbiased under the assumption that Job Corps has zero impact on eligible applicants who do not enroll in the program.

The procedure to obtain impact estimates per participant can be extended to accommodate control group crossovers (Angrist et al. 1996). The modified procedure involves dividing the estimated impact per eligible applicant by the difference between the Job Corps enrollment rate (the “show” rate) for the program group (73 percent) and the crossover rate for the control group (1.4 percent).

To illustrate how this works, we divide the population of eligible applicants into four mutually exclusive groups. These groups are defined by whether each youth would or would not enroll in Job Corps if assigned to the program group, and by whether each youth would or would not enroll in Job Corps if assigned to the control group. The four groups are as follows:

1. **Never-takers.** These are youths who would not enroll in Job Corps if they were in the program group and would not enroll in Job Corps if they were in the control group.
2. **Compliers.** These are youths who would enroll in Job Corps if they were in the program group, but would not enroll in Job Corps if they were in the control group.
3. **Defiers.** These are youths who would not enroll if they were assigned to the program group, but would enroll if they were assigned to the control group.
4. **Always-takers.** These are youths who would enroll in Job Corps if they were in the program group and also would enroll in Job Corps if they were in the control group.

Because of random assignment, the study's observed program and control groups each include equal proportions of the four groups. Furthermore, we can decompose the impact per eligible applicant on an outcome measure into a weighted sum of the contrasts between program and control group members in each of the four groups above (that is, $I = p_N I_N + p_C I_C + p_D I_D + p_A I_A$, where I is the impact per eligible applicant, p_N is the proportion of never-takers in the study population, I_N is the difference between the mean outcome of program and control group members in the never-taker group--the impact per never-taker--and similarly for compliers, defiers, and always-takers whose terms are subscripted by C , D , and A , respectively).

In this framework, controlling for crossovers amounts to estimating the impact of Job Corps participation per complier.

The following two-by-two table shows whether never-takers, compliers, defiers, and always-takers would be enrollees or nonenrollees, based on their research status:

If Youth Were Assigned to the Control Group	If Youth Were Assigned to the Program Group	
	Does Not Enroll	Enrolls
Does Not Enroll	Never-taker	Complier
Enrolls	Defier	Always-taker

Importantly, *we do not know who in the study population is in which of the four groups* because youths were assigned only to one research status. We do not know whether control group members who enrolled in Job Corps--the crossovers--were defiers or always-takers because that would depend on whether they would have enrolled in Job Corps if they had instead been assigned to the program group. Furthermore, we do not know which program group members would have been crossovers if they had instead been assigned to the control group. Likewise, we do not know whether a program group member who enrolled in Job Corps was a complier or an always-taker.

As stated, we do not know which program and control group members are in which of the four groups. However, three identifying assumptions, each of which is plausible, enable us to estimate the impact per complier.

First, we assume that impacts per never-taker are zero. This is similar to the assumption that impacts on no-shows are zero that we used to estimate impacts per participant in the absence of crossovers.

Second, we assume that impacts per always-taker are zero. This assumption implies that the mean outcomes of always-takers in the program and control groups were identical because all these youths enrolled in Job Corps. In other words, the outcomes of always-takers would be the same if they enrolled as part of the program group or as part of the control group.

Third, we assume that there are no defiers. This is reasonable because it is highly likely that a youth who would enroll as part of the control group would also enroll if the youth were in the program group. In other words, no youth would enroll in Job Corps if they were told they could not enroll, but would not enroll if they were told they could enroll. As can be seen from the bottom row of the table, this assumption means that all control group crossovers were always-takers; that is, all crossovers would have enrolled in Job Corps if they had been assigned to the program group.

Using these assumptions, we can write the impact per complier as the impact per eligible applicant divided by the proportion of compliers in the population (that is, $I_c = I/p_c$). Finally, using the table above, we find that the proportion of compliers in the population equals the show-rate minus the crossover rate. This result follows from the fact that (1) the show rate equals the sum of the proportion of eligible applicants who were compliers and the proportion who were always-takers, and (2) the proportion who were always-takers equals the control group crossover rate because of the assumption that there were no defiers in the population.

We estimated program impacts using this procedure to account for crossovers, but we did not present these estimates in the 30-month impact report, for two main reasons. First, the impacts per complier were very similar to the impacts per program participant, because the crossover rate was very small. For example, the impacts per participant for the full sample were obtained by dividing the impact per eligible applicant by .73, whereas the impacts per complier were obtained by dividing the impact per eligible applicant by .716 (.73-.014). Second, we were concerned that the assumption of zero impacts on always-takers may not be tenable: control group members who enrolled in Job Corps may have had a different program experience than they would have had if they had been in the program group, because some may have been pressured to leave the program or were stigmatized. Thus, we presented impact estimates per program participant in the impact report, although these estimates are likely to be slightly biased downwards.

REFERENCES

- Angrist, J., G. Imbens, and D. Rubin. "Identification of Casual Effects Using Instrumental Variables." *Journal of the American Statistical Association*, vol. 91, no. 434, 1996.
- DuMouchel, W., and G. Duncan. "Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples." *Journal of the American Statistical Association*, vol. 78, no. 383, September 1983.
- Schochet, P. "National Job Corps Study: Methodological Appendixes on Sample Implementation and Baseline Interviewing." Princeton, NJ: Mathematica Policy Research, Inc., January 1998a.
- Schochet, P. "National Job Corps Study: Characteristics of Youths Served by Job Corps." Princeton, NJ: Mathematica Policy Research, Inc., January 1998b.
- Schochet, P., J. Burghardt, and S. Glazerman. "National Job Corps Study: The Short-Term Impacts of Job Corps on Participants' Employment and Related Outcomes, Final Report." Princeton, NJ: Mathematica Policy Research, Inc., February 2000.
- Scott, A.J., and J.N.K. Rao. "Chi-Squared Tests for Contingency Tables with Proportions Estimated from Survey Data." In *Current Topics in Survey Sampling*, edited by D. Krewski, R. Platek, and J.N.K. Rao. New York: Academics Press, 1981.

