

Contract No.: K-4279-3-00-80-30  
MPR Reference No.: 8140-530

**MATHEMATICA**  
Policy Research, Inc.

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

*June 2001*

*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

Project Officer:

Daniel Ryan

Project Director:

John Burghardt

Principal Investigators:

Terry Johnson  
Charles Metcalf  
Peter Z. Schochet

Submitted by:

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

In conjunction with:

Battelle Human Affairs Research  
Centers (Subcontractor)  
4500 Sand Point Way NE, Suite 100  
Seattle, WA 98105-3949

Decision Information Resources, Inc.  
(Subcontractor)  
2600 Southwest Freeway, Suite 900  
Houston, TX 77098

This report has been produced under Contract Number K-4279-3-00-80-30 with the U.S. Department of Labor, Employment and Training Administration. The contents of the report do not necessarily reflect the views or policies of the Department of Labor, nor does mention of trade names, commercial products, or organizations imply endorsement of these by the U.S. Government.

## **ACKNOWLEDGMENTS**

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

# CONTENTS

	<b>Page</b>
INTRODUCTION .....	ix
APPENDIX A: THE 12-, 30-, AND 48-MONTH INTERVIEWS	
A. INTRODUCTION .....	3
B. SURVEY DESIGN .....	3
1. Design of the Baseline Interview .....	3
2. Design of the 12-Month Interview .....	5
3. Design of the 30-Month Interview .....	6
4. Design of the 48-Month Interview .....	7
C. INTERVIEW RESPONSE ISSUES .....	8
1. The Baseline Interview .....	8
2. The 12-Month Interview .....	9
3. The 30-Month Interview .....	9
4. The 48-Month Interview .....	10
APPENDIX B: THE TREATMENT OF MISSING VALUES AND OUTLIERS	
A. INTRODUCTION .....	23
B. THE PREVALENCE OF MISSING VALUES .....	23
C. THE TREATMENT OF MISSING VALUES AND OUTLIERS .....	34
1. Employment and Earnings .....	34
2. Education and Training .....	40
3. Nonlabor Market Outcomes .....	43

**CONTENTS** *(continued)*

<b>Chapter</b>	<b>Page</b>
APPENDIX C: THE ADJUSTMENT FOR CROSSOVERS	
A. INTRODUCTION .....	47
B. THE ADJUSTMENT FOR EARLY CROSSOVERS .....	47
C. THE ADJUSTMENT FOR LATE CROSSOVERS .....	52
APPENDIX D: THE CALCULATION OF SAMPLE WEIGHTS AND STANDARD ERRORS	
A. INTRODUCTION .....	57
B. CALCULATION OF SAMPLE WEIGHTS .....	57
1. Weights to Account for the Sample Design .....	58
2. Weights to Account for the Survey Design .....	60
3. The Adjustment of Weights to Account for Nonresponse to the 48-Month Interview .....	63
C. CALCULATION OF STANDARD ERRORS .....	72
1. Standard Errors for Impacts per Eligible Applicant .....	73
2. Standard Errors for Impacts per Job Corps Participant .....	78
3. Significance Tests for Impacts on the Distribution of Categorical Variables .....	80
APPENDIX E: THE ESTIMATION OF REGRESSION-ADJUSTED IMPACTS	
A. INTRODUCTION .....	85
1. Impact Estimation Issues .....	86
2. Selecting Control Variables .....	89
3. Estimation Results .....	94
4. Conclusions .....	111
REFERENCES .....	113

## TABLES

Table	Page
A.1	EFFECTIVE RESPONSE RATES TO THE 48-MONTH FOLLOW-UP INTERVIEW, BY RESEARCH STATUS AND KEY SUBGROUP ..... 12
A.2	DISTRIBUTION OF THE NUMBER OF MONTHS BETWEEN 48 MONTHS AFTER RANDOM ASSIGNMENT AND COMPLETION OF THE 48-MONTH INTERVIEW FOR THOSE IN THE IN-PERSON AREAS, BY RESEARCH STATUS ..... 15
A.3	INTERVIEW MODE FOR CASES WHO COMPLETED THE 48-MONTH INTERVIEW, BY RESEARCH STATUS AND GENDER ..... 17
A.4	REASONS FOR NONCOMPLETION OF THE 48-MONTH INTERVIEW, BY RESEARCH STATUS AND GENDER ..... 18
B.1	DATA ITEM RESPONSE FOR KEY OUTCOME MEASURES USED IN THE 48-MONTH IMPACT ANALYSIS, BY RESEARCH STATUS AND GENDER ..... 24
D.1	PROBABILITIES THAT ELIGIBLE APPLICANTS WERE SELECTED TO THE CONTROL AND PROGRAM RESEARCH GROUPS, BY SAMPLING STRATA ..... 59
D.2	COMPARISON OF THE CHARACTERISTICS OF RESPONDENTS AND THE FULL SAMPLE OF RESPONDENTS AND NONRESPONDENTS TO THE 48-MONTH INTERVIEW, BY RESEARCH STATUS ..... 66
E.1	CONTROL VARIABLES INCLUDED IN THE REGRESSION MODELS TO OBTAIN REGRESSION-ADJUSTED IMPACT ESTIMATES ..... 91
E.2	IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES, FOR THE FULL SAMPLE ..... 95
E.3	IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES, FOR MALES ..... 97
E.4	IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES, FOR FEMALES ..... 99
E.5	IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES, FOR 16- AND 17-YEAR-OLDS ..... 101

**TABLES** *(continued)*

<b>Table</b>		<b>Page</b>
E.6	IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES, FOR 18- AND 19-YEAR-OLDS . . . . .	103
E.7	IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES, FOR 20- TO 24-YEAR-OLDS . . . . .	105
E.8	IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES, FOR RESIDENTIAL DESIGNEES . . . .	107
E.9	IMPACTS ON KEY OUTCOMES USING THE DIFFERENCES-IN-MEANS AND REGRESSION APPROACHES, FOR NONRESIDENTIAL DESIGNEES . . . . .	109

## INTRODUCTION

In a series of appendixes, this report discusses methodological issues related to the 48-month impact analysis for the National Job Corps Study. The appendixes are intended to complement the 48-month impact report (Schochet et al. 2001), which presents impacts of Job Corps on key participant outcomes during the 48 months after random assignment.

This report contains the following five appendixes:

1. **“The 12-, 30-, and 48-Month Interviews.”** The outcome measures for the 48-month impact analysis were constructed using follow-up interview data collected 12, 30, and 48 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 48-month impact analysis.
3. **“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 their three-year restriction period and afterwards.
4. **“The Calculation of Sample Weights and Standard Errors.”** This appendix discusses the calculation of sample weights used in the 48-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.
5. **“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 impact report.

**APPENDIX A**  
**THE 12-, 30-, AND 48-MONTH INTERVIEWS**

## **A. INTRODUCTION**

We obtained estimates over the 48 months after random assignment 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, 30, and 48 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 its survey design must be understood if the survey design for the follow-up interview is to be understood.

## **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.<sup>1</sup> About two-thirds of randomized youths in the study population lived in areas selected for in-person interviewing when they applied to Job Corps.<sup>2</sup>

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

---

<sup>1</sup>In 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.

<sup>2</sup>The 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.

<sup>3</sup>We selected the 45-day cutoff after analyzing the cumulative telephone response rates by time  
(continued...)

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. We attempted interviews 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 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

---

<sup>3</sup>(...continued)

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.

attended Job Corps, we removed these program readmits from the study sample.<sup>4</sup> 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.

We attempted a 30-month interview with all sample members who completed either the baseline or the 12-month interview, except for 54 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

---

<sup>4</sup>Because 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.

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.

#### **4. Design of the 48-Month Interview**

We conducted the 48-month interview between December 1998 and May 2000. Initially, a \$10 incentive fee was offered to all those released for interviews, but it was increased to \$25 in June 1999 to help boost the response rate.

We attempted a 48-month interview with those who completed any previous interview, with two exceptions. First, we excluded 37 youths who were confirmed to have died since their last interview. Second, to reduce data collection costs, we released only about 93 percent of program group members (8,268 of 8,907) who were eligible for 48-month interviews. These program group members were randomly selected using systematic sampling techniques, where program group members were sorted by residential status, gender, random assignment date, whether the baseline interview was completed within 45 days after random assignment, and age. In total, 13,850 youths (8,268 program and 5,582 control) were released for 48-month interviews. Respondents were asked about their experiences since their previous interview.

We attempted 48-month interviews by telephone first, and attempted interviews in person for youths in all areas who could not be reached by telephone. In-person interviewing started in late April 1999 and concluded in May 2000.

## **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. First, we summarize results from the baseline interview and the 12- and 30-month follow-up interviews (which were discussed in detail in Schochet 1998a and Schochet 2000). Second, we provide a detailed discussion of results for the 48-month interview.

### **1. The Baseline Interview**

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 have 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 (Schochet 1998a). 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. 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).<sup>5,6</sup> Most interview respondents completed the 12-month interview soon after their 12-month release date (Schochet 2000).

The effective response rate to the 12-month interview differed only slightly across key youth subgroups (Schochet 2000). 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.

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 (Schochet 2000). This is an expected result, because the in-person areas were randomly selected.

## **3. The 30-Month Interview**

The sample of those who completed 30-month interviews was the primary analysis sample used in the 30-month (short-term) impact report. 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

---

<sup>5</sup>As 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.

<sup>6</sup>The response rates exclude the program readmits and youths who died.

8,257 of the 10,405 interviews attempted, which resulted in an effective response rate of 79.4 percent (80.7 percent program, 77.4 percent control).<sup>7</sup> The effective response rate to the 30-month interview was fairly high across all key youth subgroups, and most interview respondents completed the interview soon after it was due to be completed (Schochet 2000).

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 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.

#### **4. The 48-Month Interview**

The sample of those who completed 48-month interviews was the primary analysis sample used in the 48-month impact report. Thus, obtaining sufficiently high response rates to the 48-month interview was crucial for obtaining credible estimates of the impacts of Job Corps on key participant outcomes.

We completed 48-month interviews with 11,313 of the 13,850 youths released for 48-month interviews.<sup>8</sup> In the in-person areas only, we completed 7,940 of the 9,937 interviews attempted. Thus, the effective response rate to the 48-month interview was 79.9 percent (81.5 percent program, 77.8 percent control).<sup>9</sup>

---

<sup>7</sup>The response rates exclude the program readmits and those who died.

<sup>8</sup>As noted, 639 randomly selected program group members who were eligible for 48-month interviews were not released for 48-month interviews to reduce data collection costs. Thus, 14,489 youths were eligible for 48-month interviews, although only 13,850 were released.

<sup>9</sup>The response rates exclude the program readmits and those who died.

About 88 percent of the 48-month sample also completed 30-month interviews, and 95 percent completed 12-month interviews. More than 85 percent completed both 12- and 30-month interviews, and only 2 percent completed neither. As with the 30-month sample, baseline interview data are available for everyone in the 48-month sample, because all youths completed either the full baseline interview or an abbreviated baseline interview as part of the 12-month interview.

The response rates differed across some key subgroups, although the differences are small (Table A.1). The response rate was higher for females than for males (85 percent, compared to 76 percent), and the response rate was about six percentage points higher for those who lived in less populated areas than for those who lived in more populated ones. 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 age, race/ethnicity, or region. Interestingly, the pattern of findings is very similar to that of the 30-month interview.

Because of these subgroup differences in response rates, we adjusted sample weights for the 48-month interview sample to help reduce the potential bias in the impact estimates due to interview nonresponse (see Appendix D). We used these adjusted weights to calculate all impact estimates.

Most interview respondents completed the 48-month interview soon after the 48-month point (Table A.2). We completed the average 48-month interview in month 49.8, and more than 78 percent within 3 months after the 48-month interview release date (that is, before month 51). Less

TABLE A.1

EFFECTIVE RESPONSE RATES TO THE 48-MONTH FOLLOW-UP INTERVIEW,  
BY RESEARCH STATUS AND KEY SUBGROUP

Subgroup	Effective Response Rate		
	Program Group	Control Group	Combined Sample
Full Sample	81.5	77.8	79.9
<b>Demographic Characteristics</b>			
Gender			
Male	78.2	73.7	76.2
Female	85.6	84.6	85.2
Age at Application			
16 to 17	81.4	79.2	80.4
18 to 19	81.9	77.3	80.0
20 to 21	81.0	76.8	79.2
22 to 24	81.1	75.6	78.9
Race/Ethnicity			
White, non-Hispanic	82.4	81.4	82.0
Black, non-Hispanic	83.7	80.5	82.4
Hispanic	80.1	76.4	78.5
Other	80.9	79.2	80.2
Region			
1	79.7	77.1	78.6
2	78.8	67.6	73.8
3	81.1	76.6	79.1
4	83.1	81.2	82.3
5	80.9	77.5	79.5
6	81.9	78.3	80.4
7/8	84.4	84.3	84.4
9	80.1	73.8	77.3
10	76.7	78.5	77.5
Size of City of Residence			
Less than 2,500	84.3	84.2	84.2
2,500 to 10,000	86.4	81.8	84.5
10,000 to 50,000	81.2	79.2	80.4
50,000 to 250,000	80.4	77.0	79.0
250,000 or more	81.0	76.3	79.0
PMSA or MSA Residence Status			
In PMSA	79.8	73.3	77.1
In MSA	82.4	80.4	81.5
In neither	84.4	85.1	84.7

TABLE A.1 (continued)

Subgroup	Effective Response Rate		
	Program Group	Control Group	Combined Sample
Density of Area of Residence			
Superdense	80.8	75.4	78.6
Dense	80.7	78.8	79.9
Nondense	83.8	81.8	83.0
Lived in Areas with a Large Concentration of Nonresidential Females			
Yes	81.4	78.7	80.2
No	81.5	77.0	79.7
Legal U.S. Resident			
Yes	81.4	77.9	80.0
No	84.0	68.6	78.0
Job Corps Application Date			
11/94 to 2/95	81.7	77.9	80.2
3/95 to 6/95	83.4	80.5	82.1
7/95 to 9/95	81.8	77.2	79.8
10/95 to 12/95	78.1	74.7	76.7
<b>Fertility and Family Status</b>			
Fertility			
Had dependents	84.1	85.5	84.7
Had no dependents	80.9	76.1	78.9
Family Status			
Family head	83.1	80.7	82.1
Family member	81.9	78.9	80.7
Unrelated individuals	79.0	73.2	76.6
<b>Education</b>			
Completed the 12th grade	84.4	79.0	82.2
Did not complete the 12th grade	80.6	77.6	79.3
<b>Welfare Dependence</b>			
Public Assistance			
Received AFDC	82.9	81.0	82.1
Received other assistance	80.4	79.9	80.2
Did not receive	81.0	75.7	78.7
<b>Health</b>			
Had Any Health Conditions That Were Being Treated			
Yes	88.3	83.3	86.3
No	81.8	78.2	80.3

TABLE A.1 (continued)

Subgroup	Effective Response Rate		
	Program Group	Control Group	Combined Sample
<b>Crime</b>			
Arrests			
Arrested in past three years	80.6	73.8	77.6
Not arrested in past three years	81.6	78.6	80.3
Convictions			
Ever convicted or adjudged delinquent	78.2	72.7	75.8
Never convicted or adjudged delinquent	81.6	78.2	80.2
<b>Anticipated Program Enrollment Information</b>			
Residential Designation Status			
Resident	81.1	76.6	79.2
Nonresident	82.9	82.1	82.6
CCC/Contract Center Designation <sup>a</sup>			
CCC	82.1	78.1	80.4
Contract center	81.3	78.6	80.2
Performance Level of Designated Center <sup>a</sup>			
High or medium-high	81.2	77.5	79.7
Medium-low or low	81.6	79.4	80.6
Size of Designated Center <sup>a</sup>			
Large or medium-large	80.7	77.1	79.2
Medium-small or small	81.8	79.3	80.8
<b>Sample Size</b>	<b>5,725</b>	<b>4,212</b>	<b>9,937</b>

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 in the in-person areas were excluded from the calculations: (1) 97 cases (43 control group and 54 program group members) who were confirmed to have died since their previous interview, (2) 63 cases (31 control and 32 program) who were determined to have enrolled in Job Corps prior to random assignment, and (3) 443 randomly selected program group members who were eligible for 48-month interviews but, in an effort to reduce data collection costs, were not released for 48-month interviews.

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

TABLE A.2

DISTRIBUTION OF THE NUMBER OF MONTHS BETWEEN 48 MONTHS AFTER  
RANDOM ASSIGNMENT AND COMPLETION OF THE 48-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 <sup>a</sup>	11.6	13.4	12.4
0 to .5	28.2	29.2	28.6
.5 to 1	14.4	14.7	14.5
1 to 2	13.7	13.2	13.5
2 to 3	9.7	9.2	9.5
3 to 4	6.6	6.6	6.6
4 to 5	4.9	4.5	4.7
5 to 6	3.6	3.4	3.5
6 to 12	6.4	5.3	5.9
12 or More	0.9	0.6	0.8
Average Number of Months	1.8	1.7	1.8
<b>Number of Respondents to the 48-Month Interview</b>	<b>4,661</b>	<b>3,273</b>	<b>7,934</b>

SOURCE: 48-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.

<sup>a</sup>Youths in the in-person areas who did not complete the 30-month interview within 45 months after random assignment but who were located before 48 months after random assignment were administered the 48-month interview.

than 7 percent of interviews were completed more than six months after the release date (that is, after month 54). The distributions of completion times were similar for program and control group members. The fact that most interviews were conducted quickly and that most 48-month respondents also completed 12- and 30-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 85 percent of interviews were completed by telephone in MPR's phone center (Table A.3). About 15 percent were conducted in the field (8.5 percent in person, 5.5 percent when the field interviewer had the youth call the MPR phone center, and 1 percent when the field interviewer called the youth).<sup>10</sup> About 1 percent were completed while the respondent was at a Job Corps center. A higher percentage of males than females completed interviews in the field (about 18 percent, compared to 11 percent) and a smaller percentage of males completed interviews by telephone (about 82 percent, compared to 89 percent). The figures are similar for program and control group members.

As expected, the proportion of interviews completed in the field was higher for the 48-month interview (15 percent) than for the 30-month interview (13 percent) and the 12-month interview (6.5 percent), because it became increasingly difficult to locate youths by phone.

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 74 percent of program group nonrespondents and 73 percent of control group

---

<sup>10</sup>We conducted (1) 45 interviews with youths who were living at a Job Corps center, (2) 12 interviews with youths who were living in a school or college, (2) 522 interviews with youths in jail (297 program and 223 control), (3) 29 interviews with youths living in halfway houses or residential treatment centers, (4) 79 interviews with youths in the military, (5) 25 interviews with youths in a group home, and (6) 16 interviews with homeless youths. The rest were conducted at a private residence.

TABLE A.3

INTERVIEW MODE FOR CASES WHO COMPLETED THE 48-MONTH INTERVIEW,  
BY RESEARCH STATUS AND GENDER  
(Percentages)

Interview Mode	Program Group			Control Group		
	Males	Females	Total	Male	Females	Total
Telephone Center	80.7	88.3	84.1	82.2	89.5	84.9
In the Field	19.3	11.7	15.8	17.8	10.4	15.0
Interviewer called youth	1.8	1.0	1.5	1.2	1.2	1.2
Interviewer had youth use a cell phone to call the phone center	10.0	7.5	8.9	9.0	7.0	8.3
In person	7.4	3.2	5.5	7.6	2.2	5.6
Interview Conducted While Respondent Was at a Job Corps Center <sup>a</sup>	0.6	0.8	0.7	1.4	1.2	1.3
<b>Number of Respondents to the 48-Month Interview</b>	<b>3,741</b>	<b>3,087</b>	<b>6,828</b>	<b>2,787</b>	<b>1,698</b>	<b>4,485</b>

SOURCE: 48-month follow-up interview data.

<sup>a</sup>Interviews conducted at Job Corps are counted as having been conducted by telephone or in the field (that is, this category is not exclusive of the other categories).

TABLE A.4  
 REASONS FOR NONCOMPLETION OF THE 48-MONTH INTERVIEW,  
 BY RESEARCH STATUS AND GENDER  
 (Percentages)

Reasons for Noncompletion	Program Group			Control Group		
	Males	Females	Total	Male	Females	Total
Unable to Locate	75.6	71.6	74.3	73.7	70.4	73.0
Refusal	16.5	23.6	18.8	17.6	24.3	19.1
Incarcerated and Unavailable	3.2	1.1	2.5	4.9	1.2	4.1
In Military and Unavailable	2.0	0.2	1.4	1.4	0.8	1.3
Break-Off or Partial Interview	2.2	2.0	2.1	1.3	1.7	1.4
Other	0.5	1.5	0.9	1.1	1.7	1.2
<b>Number of Nonrespondents to the 48-Month Interview</b>	<b>960</b>	<b>457</b>	<b>1,417</b>	<b>837</b>	<b>243</b>	<b>1,080</b>

SOURCE: 48-month follow-up interview data.

nonrespondents. The refusal rate was about 19 percent for both research groups, but it was higher for females than males (24 percent, compared to 17 percent). Not surprisingly, the refusal rate at 48 months (19 percent) was higher than it was at 30 months (15 percent) and at 12 months (6.5 percent). Among male nonrespondents, about 3 percent of program group members and 5 percent of control group members did not complete the interview because they were in jail and unavailable, and about 2 percent were in the military and unavailable. Finally, an additional 2 percent of nonrespondents broke off the interview or completed only part of it.

## **APPENDIX B**

### **THE TREATMENT OF MISSING VALUES AND OUTLIERS**

## **A. INTRODUCTION**

We constructed three categories of outcome measures for the 48-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, child care, and mobility. The 48-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 48-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 48-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; (9) child care; and (10) marital status and living arrangements.

TABLE B.1

DATA ITEM RESPONSE FOR KEY OUTCOME MEASURES  
USED IN THE 48-MONTH IMPACT ANALYSIS,  
BY RESEARCH STATUS AND GENDER  
(Percentages)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
<b>Job Corps Experiences</b>						
Enrolled in a Job Corps Center						
All months	98.4	99.4	98.8	NA	NA	NA
Quarter 1	97.2	98.7	97.9	NA	NA	NA
Quarter 5	98.6	99.1	98.8	NA	NA	NA
Quarter 16	100.0	100.0	100.0	NA	NA	NA
Months Between Random Assignment and Center Enrollment <sup>a</sup>	95.0	96.8	95.8	NA	NA	NA
Months Enrolled <sup>a</sup>	92.0	94.2	92.9	NA	NA	NA
Months Between Date Left Job Corps and the 48-Month Interview <sup>a</sup>	94.9	96.3	95.5	NA	NA	NA
Participated in Academic Classes or Vocational Training <sup>a</sup>	97.1	98.6	97.8	NA	NA	NA
Total Hours in Academic Classes and Vocational Training <sup>a</sup>	82.2	85.6	83.7	NA	NA	NA
Took Academic Classes <sup>a</sup>	97.2	98.6	97.9	NA	NA	NA
Total Hours in Academic Classes <sup>a</sup>	84.3	87.9	85.9	NA	NA	NA
Took Vocational Training <sup>a</sup>	97.2	98.6	97.8	NA	NA	NA
Total Hours in Vocational Training <sup>a</sup>	84.5	87.9	86.1	NA	NA	NA
Participation in Other Job Corps <sup>b</sup> Activities						
World of Work	92.6	95.4	93.8	NA	NA	NA
Progress/Performance Evaluation Panels	92.8	95.8	94.1	NA	NA	NA
Health Classes	92.8	95.5	94.0	NA	NA	NA
Parenting Skills Classes	93.2	96.3	94.6	NA	NA	NA
Social Skills Training	92.3	94.7	93.3	NA	NA	NA
Cultural Awareness Classes	92.3	94.9	93.5	NA	NA	NA
Alcohol and Other Drugs of Abuse Program	93.3	96.4	94.6	NA	NA	NA
<b>Education and Training in Job Corps and Elsewhere</b>						
Enrolled in a Program, by Period						
Ever during the 48 months	96.2	97.8	96.9	95.5	97.8	96.4
Quarter 1	96.0	97.4	96.6	94.2	96.7	95.2
Quarter 5	96.8	98.3	97.5	97.9	98.8	98.2
Quarter 16	98.2	98.8	98.5	98.2	98.3	98.2
Number of Programs Attended	94.7	96.3	95.4	91.4	95.6	93.1
Percentage of Weeks in Programs	83.7	87.9	85.6	86.3	90.1	87.8

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Hours per Week in Programs						
All months	82.7	87.2	84.8	85.8	89.7	87.3
Quarter 1	90.8	94.1	92.3	94.4	95.9	95.0
Quarter 5	93.9	95.6	94.7	96.1	96.9	96.4
Quarter 16	97.8	98.4	98.1	97.7	97.9	97.8
Attended Programs Other than Job Corps, by Type						
Any	96.8	98.3	97.5	96.5	98.4	97.2
High school <sup>c</sup>	94.0	95.3	94.6	93.6	95.8	94.4
ABE or ESL <sup>c</sup>	94.1	95.2	94.6	93.4	95.4	94.1
GED <sup>c</sup>	94.6	96.1	95.2	94.0	96.1	94.7
Vocational/technical school	95.2	96.7	95.8	94.7	97.3	95.7
Two-year college	95.0	96.5	95.7	94.3	96.9	95.3
Four-year college	94.7	96.3	95.5	94.2	96.6	95.1
Percentage of Weeks in Programs Other than Job Corps						
	87.3	90.4	88.7	85.8	89.7	87.3
Hours per Week in Programs Other than Job Corps, by Type						
Any	87.3	90.4	88.7	85.8	89.7	87.3
High school	92.0	93.7	92.7	89.9	93.8	91.2
GED	92.0	93.7	92.7	90.8	92.5	91.4
Vocational/technical school	93.6	95.4	94.4	92.9	96.1	94.1
Two-year college	95.1	96.4	95.7	94.2	96.3	95.0
Took Academic Classes	48.5	45.6	47.2	48.0	46.6	47.4
Weeks in Academic Classes	39.2	37.8	38.5	42.2	41.9	42.1
Hours per Week in Academic Classes	39.2	37.9	38.6	42.2	42.0	42.2
Took Vocational Training	49.7	46.4	48.2	50.1	47.8	49.2
Percentage of Weeks in Vocational Training						
	42.5	40.8	41.7	46.9	45.4	46.4
Hours per Week in Vocational Training						
	42.7	40.8	41.9	46.9	45.4	46.4
Degrees, Diplomas, and Certificates Received						
GED certificate <sup>b</sup>	98.8	99.1	98.9	99.1	99.0	99.0
High school diploma <sup>b</sup>	98.4	99.1	98.7	98.9	98.6	98.8
Vocational/technical certificate	99.3	99.4	99.3	99.2	99.1	99.2
College degree (two-year or four-year)	99.4	99.6	99.5	99.6	99.4	99.6
Highest Grade Completed at 48 Months	99.7	99.8	99.8	99.7	99.8	99.7
<b>Employment and Earnings</b>						
Employed, by Period						
Quarter 1	94.3	96.3	95.2	94.4	97.2	95.5
Quarter 5	97.5	98.5	97.9	97.6	98.3	97.8
Quarter 16	98.1	98.4	98.3	98.2	98.7	98.4
Year 1	95.9	97.5	96.6	95.7	98.0	96.6
Year 2	98.2	98.6	98.4	98.0	98.8	98.3
Year 3	98.6	98.6	98.6	98.3	98.4	98.4

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Year 4	98.4	98.7	98.5	98.4	98.6	98.5
Ever during the 48 months	99.1	99.2	99.2	99.1	99.1	99.1
Number of Jobs	96.8	98.1	97.3	96.5	98.3	97.2
Percentage of Weeks Employed, by Period						
Quarter 1	91.8	94.7	93.1	92.9	95.5	93.8
Quarter 5	95.3	96.6	95.9	95.6	96.6	96.0
Quarter 16	97.0	97.7	97.3	96.9	97.4	97.1
Year 1	90.9	94.0	92.3	91.4	94.7	92.6
Year 2	93.1	95.2	94.1	93.3	95.2	94.0
Year 3	92.6	94.6	93.5	92.2	94.5	93.1
Year 4	94.4	96.1	95.2	94.3	95.5	94.7
All months	81.6	87.8	84.4	82.0	87.4	84.1
Hours per Week Employed, by Period						
Quarter 1	91.2	94.4	92.6	91.9	94.9	93.0
Quarter 5	94.0	95.6	94.7	94.2	95.9	94.8
Quarter 16	94.7	96.3	95.4	95.0	96.2	95.5
Year 1	90.4	93.7	91.9	90.6	94.1	91.9
Year 2	91.7	94.2	92.8	91.6	94.3	92.6
Year 3	90.4	93.2	91.7	90.0	93.3	91.3
Year 4	92.4	94.9	93.6	92.6	94.8	93.4
All months	79.9	86.4	82.8	80.5	86.3	82.7
Earnings per Week, by Period						
Quarter 1	91.2	94.4	92.6	91.9	94.9	93.0
Quarter 5	94.0	95.6	94.7	94.2	95.9	94.8
Quarter 16	94.6	96.3	95.4	95.0	96.1	95.4
Year 1	90.4	93.7	91.9	90.6	94.1	91.9
Year 2	91.7	94.2	92.8	91.6	94.3	92.6
Year 3	90.4	93.2	91.7	90.0	93.3	91.3
Year 4	92.3	94.8	93.5	92.5	94.7	93.3
All months	79.9	86.4	82.8	80.5	86.3	82.7
Characteristics of the Most Recent Job in Quarter 16 for Those Employed						
Number of months on job	98.8	99.1	98.9	98.9	99.2	99.0
Usual hours worked per week	99.6	99.7	99.7	99.7	99.5	99.7
Hourly wage	99.6	99.7	99.7	99.7	99.5	99.7
Weekly earnings	99.6	99.7	99.7	99.7	99.5	99.7
Occupation	99.3	99.5	99.4	99.2	99.3	99.3
Type of employer	96.0	95.8	95.9	96.1	93.6	95.2
Fringe benefits available						
Health insurance	98.1	98.5	98.3	97.1	98.4	97.6
Paid sick leave	97.3	97.9	97.5	97.0	96.8	96.9
Paid vacation	98.2	98.4	98.3	97.9	97.6	97.8
Retirement or pension benefits	94.9	95.1	95.0	95.3	93.4	94.6
Employed or in an Education or Training Program, by Period						
Quarter 1	94.6	96.8	95.6	93.3	96.2	94.4
Quarter 5	97.2	98.5	97.8	97.5	98.2	97.7

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Quarter 16	98.1	98.4	98.2	98.0	98.2	98.1
Year 1	96.3	97.9	97.0	95.6	97.9	96.5
Year 2	98.3	98.9	98.6	98.2	98.8	98.4
Year 3	98.4	98.8	98.6	98.5	98.3	98.4
Year 4	98.6	98.7	98.6	98.3	98.4	98.4
Ever during the 48 months	99.7	99.7	99.7	99.3	99.6	99.4
Percentage of Weeks in Any Activity	75.2	82.3	78.4	76.8	82.6	79.0
Hours per Week in Any Activity						
Quarter 1	85.1	89.8	87.2	87.5	91.6	89.1
Quarter 5	88.8	91.9	90.2	90.8	93.3	91.8
Quarter 16	93.1	94.9	93.9	93.0	94.3	93.5
Year 1	84.1	89.0	86.3	86.0	90.5	87.7
Year 2	85.4	89.7	87.3	88.0	91.1	89.2
Year 3	85.9	89.1	87.3	86.3	89.6	87.5
Year 4	90.1	92.5	91.2	89.7	92.1	90.6
All months	70.1	78.0	73.7	73.1	79.7	75.6
<b>Receipt of Public Assistance</b>						
Received AFDC/TANF, Food Stamps, SSI/SSA, or GA Benefits, by Period						
All months	93.4	97.7	95.3	93.2	97.8	94.9
Year 1	94.0	97.1	95.4	93.4	97.1	94.8
Year 2	95.2	97.8	96.4	94.3	97.8	95.7
Year 3	96.4	97.7	97.0	96.6	97.9	97.1
Year 4	97.6	98.1	97.8	97.9	98.2	98.0
Number of Months Received Benefits	88.3	91.5	89.7	87.6	90.8	88.8
Amount of Benefits Received	69.2	69.0	69.1	67.7	69.0	68.2
Received AFDC/TANF Benefits, by Period						
All months	93.6	97.6	95.4	94.0	97.5	95.3
Year 1	94.3	97.4	95.7	94.5	97.3	95.6
Year 2	96.6	98.4	97.5	96.3	97.9	96.9
Year 3	98.1	98.8	98.4	98.4	98.1	98.3
Year 4	98.8	98.8	98.8	99.1	98.5	98.9
Number of Months Received AFDC/TANF Benefits	91.7	94.1	92.8	92.1	93.3	92.5
Amount of AFDC/TANF Benefits Received	81.1	81.7	81.4	81.3	82.2	81.6
Received Food Stamp Benefits, by Period						
All months	95.6	98.1	96.7	95.5	97.9	96.4
Year 1	96.2	97.9	97.0	95.8	98.0	96.7
Year 2	98.2	99.0	98.6	97.7	98.9	98.1
Year 3	98.2	98.6	98.4	98.3	98.6	98.5
Year 4	98.9	98.7	98.8	98.8	98.5	98.7
Number of Months Received Food Stamp Benefits	93.5	94.4	93.9	92.5	93.9	93.1

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Amount of Food Stamp Benefits						
Received	77.5	78.2	77.8	76.7	78.7	77.4
Received SSI/SSA Benefits	95.0	97.4	96.1	95.3	97.9	96.3
Number of Months Received SSI/SSA						
Benefits	94.6	97.0	95.7	94.9	97.4	95.9
Amount of SSI/SSA Benefits Received	92.9	95.8	94.2	93.6	96.3	94.6
Received GA Benefits	93.6	95.7	94.5	93.6	95.9	94.5
Number of Months Received GA						
Benefits	93.5	95.3	94.3	93.1	95.6	94.0
Amount of GA Benefits Received	93.2	95.0	94.0	92.8	95.2	93.7
Covered by Public Health Insurance						
At 12 months	93.1	97.6	95.1	93.3	97.4	94.9
At 30 months	95.0	99.0	96.8	95.7	98.1	96.6
At 48 months	96.3	99.1	97.6	96.2	98.9	97.2
Received WIC Benefits (for females only)	NA	98.5	98.5	NA	98.7	98.7
Number of Months Received WIC						
Benefits (for females only)	NA	95.7	95.7	NA	96.6	96.6
Lived in Public Housing						
At 12 months	98.1	98.2	98.2	96.8	98.2	97.3
At 30 months	98.2	99.0	98.5	98.1	98.5	98.2
At 48 months	98.5	98.8	98.7	98.9	99.0	99.0
Received UI Benefits	96.4	98.2	97.2	96.6	98.5	97.3
Number of Weeks Received UI Benefits	96.1	97.9	96.9	96.0	97.9	96.7
Amount of UI Benefits Received	96.1	97.9	96.9	95.7	97.7	96.5
Received Child Support	99.5	99.5	99.5	99.6	99.5	99.6
Amount of Child Support Received	99.3	97.2	98.4	99.5	98.7	99.2
Received Income from Friends	99.1	99.4	99.2	99.5	99.4	99.5
Amount of Income Received from						
Friends	96.5	95.8	96.1	97.3	95.9	96.7
Received Other Income	99.1	99.4	99.2	99.4	99.2	99.3
Amount of Other Income Received	98.0	98.3	98.1	98.8	98.5	98.7
<b>Involvement with the Criminal Justice System</b>						
Arrested or Charged with a Delinquency or Criminal Complaint, by Period						
Year 1	99.2	99.7	99.4	99.3	99.9	99.5
Year 2	99.2	99.7	99.4	99.3	99.9	99.5
Year 3	99.1	99.7	99.4	99.4	99.9	99.6
Year 4	97.5	97.9	97.7	96.6	97.6	96.9
All months	99.8	99.9	99.8	99.7	99.9	99.8
Number of Arrests	98.6	99.5	99.0	98.9	99.9	99.2
Months Until First Arrested	97.5	98.0	97.7	97.3	97.8	97.5

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Most Serious Charge for Which Arrested	98.6	99.5	99.0	98.9	99.9	99.2
Arrested for:						
Murder	98.6	99.5	99.0	98.9	99.9	99.2
Assault	98.7	99.6	99.1	99.0	99.9	99.3
Robbery	98.7	99.5	99.1	99.0	99.9	99.3
Burglary	98.7	99.5	99.1	99.0	99.9	99.4
Larceny, vehicle theft, or other property crimes	99.0	99.7	99.3	99.3	99.9	99.5
Drug law violations	98.9	99.6	99.2	99.3	99.9	99.5
Other personal crimes	98.9	99.6	99.2	99.0	99.9	99.4
Other miscellaneous crimes	99.4	99.7	99.6	99.4	99.9	99.6
Convicted, Pled Guilty, or Adjudged						
Delinquent	99.2	99.4	99.3	98.9	99.6	99.1
Made a Deal or Plea-Bargained	98.0	99.1	98.5	97.5	98.8	98.0
Most Serious Charge for Which Convicted	98.3	99.1	98.7	98.1	99.4	98.6
Convicted of:						
Murder	97.2	98.9	98.0	97.0	99.2	97.8
Assault	97.3	98.9	98.1	97.1	99.2	97.9
Robbery	97.2	98.9	98.0	97.0	99.2	97.8
Burglary	97.4	98.9	98.1	97.1	99.2	97.9
Larceny, vehicle theft, or other property crimes	97.6	99.0	98.2	97.2	99.2	97.9
Drug law violations	97.3	98.9	98.0	97.3	99.2	98.0
Other personal crimes	97.3	99.0	98.1	97.1	99.2	97.9
Other miscellaneous crimes	98.1	99.0	98.5	97.7	99.4	98.3
Served Time in Jail for Convictions	99.2	99.4	99.3	98.9	99.6	99.1
Weeks Spent in Jail	97.5	99.2	98.3	96.6	99.5	97.7
Put on Probation or Parole	98.8	99.3	99.0	98.3	99.6	98.8
<b>Tobacco, Alcohol, and Illegal Drug Use</b>						
Smoked Cigarettes						
At 12 months	99.7	99.8	99.8	99.8	99.9	99.8
At 30 months	99.7	100.0	99.8	100.0	99.8	99.9
At 48 months	99.7	99.8	99.7	99.9	99.8	99.8
Consumed Alcoholic Beverages						
At 12 months	99.7	99.8	99.7	99.8	99.9	99.9
At 30 months	99.7	100.0	99.8	99.9	99.8	99.9
At 48 months	99.7	99.8	99.7	99.8	99.6	99.8
Used Marijuana, Hashish, or Hard Drugs						
At 12 months	99.5	99.8	99.6	99.7	99.8	99.7
At 30 months	99.5	99.9	99.7	99.8	99.7	99.7
At 48 months	99.5	99.7	99.6	99.6	99.5	99.5
Used Marijuana or Hashish						
At 12 months	99.6	99.8	99.7	99.7	99.9	99.8
At 30 months	99.6	99.9	99.8	99.9	99.8	99.9
At 48 months	99.6	99.7	99.6	99.7	99.6	99.7

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Used Hard Drugs						
At 12 months	99.5	99.8	99.7	99.7	99.8	99.7
At 30 months	99.7	99.9	99.8	99.8	99.7	99.8
At 48 months	99.7	99.9	99.8	99.9	99.8	99.8
Snorted Cocaine Powder						
At 12 months	99.7	99.8	99.8	99.8	99.9	99.8
At 30 months	99.7	100.0	99.8	99.8	99.8	99.8
At 48 months	99.7	99.8	99.7	99.8	99.6	99.7
Smoked Crack Cocaine or Freebased						
At 12 months	99.7	99.8	99.8	99.8	99.8	99.8
At 30 months	99.7	99.9	99.8	99.9	99.8	99.8
At 48 months	99.6	99.7	99.7	99.8	99.6	99.7
Used Speed, Uppers, or Methamphetamines						
At 12 months	99.6	99.8	99.7	99.8	99.8	99.8
At 30 months	99.7	99.9	99.8	99.9	99.8	99.8
At 48 months	99.6	99.7	99.7	99.9	99.7	99.8
Used Hallucinogenic Drugs						
At 12 months	99.6	99.8	99.7	99.8	99.9	99.9
At 30 months	99.7	99.9	99.8	99.9	99.8	99.8
At 48 months	99.6	99.7	99.6	99.9	99.7	99.8
Used Heroin, Opium, Methadone, or Downers						
At 12 months	99.7	99.8	99.7	99.8	99.9	99.9
At 30 months	99.7	99.9	99.8	99.9	99.8	99.9
At 48 months	99.6	99.7	99.7	99.8	99.5	99.7
Used Other Drugs						
At 12 months	99.7	99.8	99.7	99.8	99.9	99.8
At 30 months	99.6	99.9	99.8	99.9	99.7	99.8
At 48 months	99.6	99.7	99.7	99.8	99.7	99.8
Shot or Injected Drugs with a Needle or Syringe						
At 12 months	99.7	99.8	99.7	99.8	99.9	99.9
At 30 months	99.7	99.9	99.8	99.9	99.7	99.8
At 48 months	99.6	99.7	99.7	99.9	99.7	99.8
In Alcohol or Drug Treatment	99.6	99.9	99.8	99.8	99.9	99.8
Weeks in Alcohol or Drug Treatment	99.4	99.8	99.6	99.6	99.8	99.7
<b>Health</b>						
Health Status						
At 12 months	99.6	99.8	99.7	99.7	99.8	99.7
At 30 months	99.6	99.9	99.7	99.9	99.7	99.8
At 48 months	99.6	99.6	99.6	99.7	99.7	99.7

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
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	99.8	99.7
At 30 months	99.7	99.8	99.7	99.8	99.6	99.7
At 48 months	99.5	99.7	99.6	99.6	99.6	99.6
<b>Fertility, Marriage, and Living Arrangements</b>						
Had New Children	99.3	99.5	99.4	99.5	99.4	99.4
Number of New Children	99.2	99.4	99.3	99.5	99.4	99.4
Had Children out of Wedlock	99.1	99.4	99.2	99.4	99.3	99.4
Pregnant at 48 Months (for females)	NA	99.3	99.3	NA	99.1	99.1
Lived with All Children <sup>d</sup>	97.2	98.9	98.2	97.6	98.9	98.2
Time Spent with Noncustodial Children <sup>e</sup> Provided Support for Noncustodial Children <sup>e</sup>	94.7	82.6	92.5	95.7	87.8	94.4
Any (such as food, toys, and money)	92.7	80.9	90.5	93.7	84.7	92.3
Money	92.9	82.6	91.1	93.7	87.0	92.7
Household Membership	98.5	98.2	98.3	98.1	98.5	98.2
Whether Youth Is the Household Head	99.3	99.6	99.5	99.6	99.7	99.6
Number in Household	98.9	99.1	99.0	98.8	99.3	99.0
Marital Status at 48 Months	99.7	99.8	99.7	99.7	99.8	99.8
<b>Child Care</b>						
Ever Used Child Care	98.4	99.2	98.8	98.5	99.4	98.8
Ever Used Child Care by Relatives	98.3	99.2	98.7	98.5	99.0	98.7
Ever Used Child Care by Nonrelatives	97.7	98.9	98.2	98.0	98.6	98.2
Ever Used Day Care	97.8	98.8	98.3	98.1	98.8	98.4
Child Care Hours Per Week	95.1	95.4	95.2	95.7	95.3	95.5
Relative Child Care Hours per Week	95.4	96.4	95.9	95.8	96.1	95.9
Nonrelative Child Care Hours per Week	97.6	98.4	98.0	98.0	97.9	98.0
Day Care Hours per Week	97.5	97.9	97.7	97.9	97.6	97.8
<b>Mobility</b>						
Distance in Miles Between Zip Codes of Residence at Application to Job Corps and at the 48-Month Interview	96.6	98.3	97.3	96.1	97.2	96.5

TABLE B.1 (continued)

Outcome Measure	Program Group			Control Group		
	Males	Females	Total	Males	Females	Total
Lived in Same State at Application to Job Corps and at the 48-Month Interview	96.6	98.3	97.3	96.1	97.2	96.5
<b>Sample Size</b>	<b>3,741</b>	<b>3,087</b>	<b>6,828</b>	<b>2,787</b>	<b>1,698</b>	<b>4,485</b>

SOURCE: Baseline and 30-month, and 48-month follow-up interview data for those who completed 48-month interviews.

NOTE: All figures are unweighted.

<sup>a</sup>Data pertain to program group members who enrolled in Job Corps.

<sup>b</sup>Data pertain to program group members who enrolled in Job Corps and had a 12- or 30-month interview.

<sup>c</sup>Data pertain to those without a high school credential at random assignment.

<sup>d</sup>Data pertain to those with children.

<sup>e</sup>Data 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 48-month period was missing for about 17 percent of cases.<sup>11</sup>

Measures of the amount of benefits that were received from the main public assistance programs (AFDC/TANF and food stamps) were missing for about 20 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, for two reasons. First, there was 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 48-month sample who completed 30-month interviews before then. Consequently, the academic education and vocational training outcome measures were constructed only for those in the 48-month sample who (1) completed 30-month interviews after the error was corrected, and (2) did not complete 30-month interviews.<sup>12</sup>

---

<sup>11</sup>Because of concerns about recall error, we set variables pertaining to the first year after random assignment to missing for all 253 cases who completed a baseline and 48-month interview but not a 12- or 30-month interview.

<sup>12</sup>The 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.

Second, as discussed in the next section, the 48-month interview did not collect detailed information about enrollment in Job Corps. Thus, information on academic and vocational training experiences in Job Corps were missing for program group members who were enrolled in Job Corps during the period covered by the 48-month interview. Furthermore, these measures are missing for the small number of control group members who enrolled in Job Corps, because detailed survey information on Job Corps enrollment was never collected for these youths.

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

## **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 48-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 48-month (208-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 “AC” signified that we were unsure whether the youth was working on job 1 or on job 3, 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 and the most recent job in quarter 16.

#### **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 for the program and control groups 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. We constructed this measure 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 *who we knew worked*, but for whom we did not know for how long, because job start or 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 year 4 after random assignment, the average percentage of weeks employed using the adjusted measure was 60.2 percent for the program group and 57.2 percent for the control group (an impact of 3 percentage points). Using the unadjusted measure, the average percentage of weeks employed was 59.7 percent for the program group and 56.6 percent for the control group (an impact of 3.1 percentage points). We present the impact estimates using the adjusted measures in the impact report.

### **c. Hours Employed per Week**

To calculate measures of hours employed, we constructed for each youth an hours timeline that covered the 208-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.<sup>13</sup> 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

---

<sup>13</sup>The “usual hours” worked variable was missing for about 1 percent of jobs.

squares (OLS) procedures.<sup>14</sup> 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.

We then adjusted the measures of hours employed using the zero-correction procedure to impute missing values for employed youths. We used these adjusted measures in the impact analysis.

#### **d. Earnings**

We constructed the earnings measures using a weekly earnings timeline for each youth. A timeline entry was calculated by (1) multiplying, for each job the 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.<sup>15,16</sup>

We hand-checked cases that reported hourly wages less than \$2.50 (about 2.5 percent of jobs) and greater than \$15 (also about 2.5 percent of jobs). We looked at verbatim job descriptions and

---

<sup>14</sup>The regression  $R^2$  values were about .12.

<sup>15</sup>About 2 percent of jobs had missing wage information.

<sup>16</sup>The  $R^2$  values from the wage regressions were about .22.

other job characteristics to determine whether outlier values were valid. About 90 percent of cases were determined to be valid.

To check the robustness of study findings, we used several methods to treat hourly wages that we considered to be invalid. 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.

We calculated earnings over a given period by summing across entries in the earnings timeline, where each entry was converted into 1995 dollars with the GDP price deflator. 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.

We then adjusted the earnings measures to impute missing values for workers using the zero-correction imputation procedure. In the 48-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 year 4 was \$15.9 (\$211.4 for the program group and \$195.4 for the control group) using the adjusted earnings measure. The impact was \$16.5 using the unadjusted earnings measure, and

as expected, earnings levels were slightly smaller for both research groups (\$208.7 for the program group and \$192.2 for the control group).

**e. Characteristics of the Most Recent Job in Quarters 10 and 16**

In the 48-month impact report, we present differences in the average characteristics of jobs held by program and control group members during quarters 10 and 16, 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 and 16th quarters after random assignment. We identified the most recent job in quarter 10 by searching for the most recent positive job code in the employment timeline between weeks 118 and 130, and identified the most recent job in quarter 16 by searching for the most recent positive job code in the employment timeline between weeks 196 and 208. For ties, we selected the job that the youth had held the longest.

The outcomes describing the characteristics of the most recent job in quarter 10 were conditional on having been employed in quarter 10, and similarly for the most recent job in quarter 16. 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, and converted hourly wages into 1995 dollars.

**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 that signified 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.

Unlike the 12- and 30-month interviews, the 48-month interview did not contain a section about participation in Job Corps (because only a small number of sample members were enrolled in Job Corps during the period covered by the 48-month interview). Thus, we used Job Corps enrollment and termination dates from the Student Pay and Allotment Management Information System (SPAMIS) to extend the Job Corps timelines for the period covered by the 48-month interview.<sup>17</sup> SPAMIS data were used for about 5 percent of program group members who were enrolled in Job Corps between their previous interview and the 48-month interview. Only about 9 percent of all weeks spent in Job Corps were captured by these spells. SPAMIS data were also used to construct Job Corps timelines for control group members who ever enrolled in Job Corps, because none of the follow-up interviews collected direct information on Job Corps enrollment for these youths.

We also used the education and training timelines, along with information about usual hours per week spent in programs, to construct weekly hours timelines.<sup>18</sup> We used regression procedures to impute the small number of missing values for the variable on usual hours per week spent in

---

<sup>17</sup>Some program group members reported that they attended Job Corps in the section of the 48-month follow-up interview on participation in education and training programs. However, Job Corps enrollment rates were substantially smaller using the 48-month survey data than SPAMIS data. This was *not* the case when program group members were *directly* asked about Job Corps participation during the 12- and 30-month follow-up interviews.

<sup>18</sup>We assumed that youths in Job Corps spent 40 hours per week in education and training.

programs.<sup>19</sup> Weekly hours in the timelines were topcoded at 48 hours. We constructed period-specific measures of hours spent in education and training programs using the hours timelines.

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 program start and end dates were missing. 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 48-month impact analysis.

We also created a weekly timeline that signified whether or not the youth was in academic classes during each week of the follow-up period, and another that signified whether or not the youth was in vocational training. We applied the procedures described above to these timelines to construct measures of time spent in academic classes and vocational training.<sup>20</sup> Because SPAMIS does not contain information on time spent in academic classes or vocational training, we used a regression procedure to impute the amount of instruction received in Job Corps during those periods in which SPAMIS data were used to construct the Job Corps timelines.

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 48-month impact report, we constructed several measures of highest grade completed, because of inconsistencies in responses across interviews.

---

<sup>19</sup>The 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.

<sup>20</sup>The academic and vocational training hours timeline entries were each topcoded at 48 hours.

### 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, GA, 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 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 regression procedures to impute missing benefit amounts for AFDC/TANF and food stamp spells.<sup>21</sup> The control variables used in the models included gender, age, household composition, fertility history, region of residence, and employment and earnings measures.<sup>22</sup> 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 procedure to impute missing *continuous* variables that were

---

<sup>21</sup>The regression  $R^2$  values were about .40 for the AFDC/TANF benefit amount models and about .20 for the food stamp benefit amount models. About 3 percent of AFDC/TANF spells and 3 percent of food stamp spells had missing benefit amounts.

<sup>22</sup>We imputed the small number of missing benefit amounts for SSI, GA, and UI spells using mean benefit amounts for program recipients with nonmissing values.

conditional on other variables. For example, we imputed missing values for the time spent in jail for those who we know were incarcerated. Similarly, we imputed missing values for the time spent in drug or alcohol treatment for those who we know were treated.

**APPENDIX C**  
**THE ADJUSTMENT FOR CROSSOVERS**

## **A. INTRODUCTION**

About 1.4 percent of all control group members (and 1.2 percent of control group members in the 48-month sample) enrolled in Job Corps before their three-year restriction period ended. We refer to these youths as “early crossovers.” In addition, 3.2 percent of control group members enrolled in Job Corps between three and four years after random assignment (that is, after their restriction period ended). We refer to these youths as “late crossovers.” To preserve the integrity of the random assignment design, we treated crossovers as control group members in the analysis. Thus, impact estimates that do not account for these crossovers could be biased if crossovers benefited from participation in Job Corps.

The 48-month impact report describes in detail statistical procedures that we used to estimate impacts per eligible applicant and impacts per program participant that do not account for control group crossovers. We estimated impacts *per eligible applicant* 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. We estimated impacts *per participant* that do not adjust for crossovers 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 impact report, however, only briefly discussed our approach for estimating impacts for crossovers. This appendix describes these procedures in more detail.

## **B. THE ADJUSTMENT FOR EARLY CROSSOVERS**

A small number of control group members enrolled in Job Corps before their three-year restriction period ended. As described in the report on study implementation (Burghardt et al. 1999),

the Job Corps national office allowed most of these youths to remain at centers, but held outreach and admissions and center staff accountable for these errors. The average duration of stay in Job Corps for these youths (7.6 months) was very similar to the average duration of stay for program group enrollees (8 months). Thus, impact estimates on employment and earnings in the postprogram period that do not adjust for these crossovers could be slightly biased downwards if these crossovers benefited from participation in Job Corps.

The procedure to obtain impact estimates per participant in the absence of crossovers can be extended to accommodate early crossovers in the control group (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 (1.2 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 as an early crossover 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 as an early crossover 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 early 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 Program Group	
If Youth Were Assigned to the Control 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 we used to estimate impacts per participant in the absence of crossovers, that impacts on no-shows are zero.

*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. This assumption is reasonable, because, as noted, the average duration of stay was similar for the early crossovers and program group enrollees, and both groups were enrolled in Job Corps at roughly the same time (soon after random assignment).

*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 as part of the program group. In other words, no youths 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 early 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,  $I_c = I/p_c$ ). Using the table above, the proportion of compliers in the population equals the show-rate minus the early 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.

Importantly, the impacts per complier were very similar to the impacts per program participant that do not adjust for the early crossovers, because the early crossover rate was very small. For example, we obtained the impacts per participant for the full sample by dividing the impact per eligible applicant by .73, whereas we obtained the impacts per complier by dividing the impact per eligible applicant by .718 (.73-.012).

Finally, in the impact report, we present the mean of each outcome measure for program group *compliers* (although for clarity, we refer to them as mean outcomes for program group *participants*). We cannot directly observe these mean outcomes, because we do not know which program group members were compliers. However, we can estimate them by noting that the mean value for an outcome measure for the *full* program group ( $T$ ) can be written as a weighted average of the mean outcome for program group members in each of the four groups discussed above (that is,  $T = p_N T_N + p_C T_C + p_D T_D + p_A T_A$ ). Under the assumption that there are no defiers (that is,  $p_D = 0$ ), the mean for always-takers in the program group ( $T_A$ ) equals the mean for the early crossovers in the control group (which is observed, and which we denote by  $C_{CR}$ ), and the mean for never-takers in the program group ( $T_N$ ) equals the mean for no-shows in the program group (which is also observed, and which we denote by  $T_{NS}$ ). Thus, the mean outcome for program group compliers can be estimated using the following expression:

$$(1) \quad \bar{T}_C = \frac{\bar{T} + (1-p_S)\bar{T}_{NS} + p_{CR}\bar{C}_{CR}}{(p_S+p_{CR})}$$

where  $p_S$  is the show rate for the program group and  $p_{CR}$  is the control group early crossover rate.

### C. THE ADJUSTMENT FOR LATE CROSSOVERS

Control group members were allowed to enroll in Job Corps after their three-year restriction period ended. About 3.2 percent of control group members enrolled in the program between their third and fourth years after random assignment. The enrollment rate was 4.6 percent for those 16 and 17 at application to Job Corps, 2.7 percent for those 18 and 19, and 1.1 percent for those 20 to 24. About 55 percent of these late crossovers were enrolled in Job Corps during the last quarter of the four-year period.

The approach to accommodate the *early* crossovers cannot be used to accommodate the *late* crossovers. As discussed, the adjustment procedure for *early* crossovers assumes that the average outcomes of early crossovers in the control group were the same as the outcomes of those in the program group who would have been early crossovers had they instead been assigned to the control group (whom we label “would-be” early crossovers). This assumption (that impacts per always-taker are zero) is reasonable, because most early crossovers in the control group enrolled in Job Corps soon after random assignment and thus were in Job Corps at roughly the same time as the would-be early crossovers in the program group. Thus, average earnings during the postprogram period were probably similar for the two groups.

The *late* crossovers, however, enrolled in Job Corps more than three years after random assignment, whereas nearly all program group participants enrolled within one year. Thus, we cannot assume that the average outcomes of late crossovers in the control group were similar to those of would-be late crossovers in the program group. In other words, the assumption that impacts per always-taker are zero is not tenable in this context. Instead, average earnings late in the observation period were probably much lower for the late control group crossovers than for their program group counterparts, because more than half these control group members were enrolled in Job Corps during this period, and those who had left Job Corps had been out for only a short period. Consequently,

impact estimates on postprogram employment and earnings that do not adjust for these late control group crossovers would probably be biased slightly upwards.

Our procedure to adjust for the late control group crossovers was to “assume” that these crossovers never enrolled in Job Corps, and to impute their employment and education outcomes covering the last five quarters of the 48-month period. We conducted the imputation procedure in two stages. In the first stage, we identified noncrossovers in the control group whose average demographic characteristics and employment and education experiences during the first two years after random assignment were similar to those of the late crossovers.<sup>23</sup> Second, we imputed the employment and education outcomes of late crossovers using the average outcomes of noncrossovers in the matched sample (by age and gender).<sup>24</sup>

---

<sup>23</sup>We used propensity score procedures to select the matched sample. The probability that a control group member was a late crossover was regressed on a set of explanatory variables, and a predicted probability (propensity score) was calculated for each control group member. We then selected the matched sample of noncrossovers as those with the closest propensity scores to those of the crossovers.

<sup>24</sup>We did not impute other outcomes (such as crime and family formation measures) for the late crossovers.

## **APPENDIX D**

### **THE CALCULATION OF SAMPLE WEIGHTS AND STANDARD ERRORS**

## **A. INTRODUCTION**

This technical appendix describes the calculation of sample weights that were used in the 48-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-, 30-, or 48-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 D.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.<sup>25</sup> 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 D.1 were adjusted for the following sample members:

- C Four youths in the program research group who were also randomly assigned to the program nonresearch group.<sup>26</sup> The selection probabilities for each of these youths is  $2p$ , where  $p$  is the relevant sampling probability from Table D.1 for each youth.

---

<sup>25</sup>Sampling rates were higher in these 57 areas to meet sample size targets for nonresidential students.

<sup>26</sup>This 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 D.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
<b>Control Group</b>		
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
<b>Program Research Group</b>		
Residential designees	10.7	11.1
Nonresidential designees	15.4	17.0
<b>Number in Sample Universe</b>	<b>47,288</b>	<b>33,595</b>

C Twenty-seven 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-, 30-, and 48-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.<sup>27</sup> 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. To maximize the precision of the impact estimates, we selected different proportions of superdense, dense, and nondense areas for in-person interviewing, subject to the cost of conducting interviews in each type of area and the limitations of 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.

We calculated the probability that a youth was selected to the baseline interview sample 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
- $= 29/75$  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

---

<sup>27</sup>Control 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.

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-, 30-, and 48-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 sample were identical to the selection probabilities to the 12-month interview sample. The selection probabilities to the 48-month interview sample were also identical to those to the 12-month sample for control group members. However, for program group members, the 48-month selection probabilities were slightly smaller than the 12-month selection probabilities, because to reduce data collection costs, we randomly selected for 48-month interviewing 93 percent of program group members who were eligible for 48-month interviews.<sup>28</sup>

---

<sup>28</sup>This subsampling, however, affected selection probabilities for all program group members *equally* because of random sampling.

The primary weights used in the 48-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 48-month interview 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 48-Month Interview**

The main analysis sample for the 48-month impact analysis included the 11,313 youths (6,828 program group and 4,485 control group members) who completed 48-month interviews. The effective response rate (that is, the response rate in the in-person areas) to the 48-month interview was 79.9 percent (81.5 percent for the program group and 77.8 percent for the control group). Because about 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 respondents 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 48-month interview on estimated impacts and discuss our approach for adjusting for these effects.<sup>29</sup>

**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 48-month nonrespondents who did not complete the baseline interview.

We performed the analysis using *only* the 9,937 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 have 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 7,940 respondents to the 48-month interview (3,276 control group and 4,664 program group members) and 1,997 nonrespondents (936 control group and 1,061 program group members). We excluded from the analysis the 443 program group members in the in-person areas who were eligible for 48-month interviews but, in an effort to reduce data collection costs, were not released for interviewing.

---

<sup>29</sup>We also adjusted for the effects of nonresponse to the 12- and 30-month interviews using the same procedure as described next for the 48-month sample (Schochet 2000). The sample of those who completed the 12- and 30-month interviews were used in the impact analysis to test the robustness of our findings using the 48-month sample.

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.<sup>30</sup> 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. We also conducted similar tests comparing the characteristics of *respondents* in the program and control groups.

There are some differences in the characteristics of respondents to the 48-month interview and the full sample of respondents and nonrespondents (Table D.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.

---

<sup>30</sup>The 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 D.2

COMPARISON OF THE CHARACTERISTICS OF RESPONDENTS AND THE FULL SAMPLE  
OF RESPONDENTS AND NONRESPONDENTS TO THE 48-MONTH INTERVIEW,  
BY RESEARCH STATUS  
(Percentages)

Characteristic <sup>a</sup>	Control Group		Program Group	
	Respondents <sup>b</sup>	Respondents and Nonrespondents	Respondents <sup>b</sup>	Respondents and Nonrespondents
<b>Demographic Characteristics</b>				
Male	54.4***	57.7	55.8***	58.0
Age at Application				
16 to 17	40.5	39.7	39.9	39.9
18 to 19	31.8	32.1	32.0	31.8
20 to 21	16.6	16.8	16.3	16.4
22 to 24	11.1	11.4	11.8	11.9
(Average age)	18.9**	19.0	18.9	19.0
Race/Ethnicity				
White, non-Hispanic	22.4**	21.8	22.5*	22.5
Black, non-Hispanic	51.5	51.2	52.8	52.1
Hispanic	18.6	19.5	17.8	18.3
Other	7.5	7.5	6.9	7.1
Region				
1	5.5**	5.5	5.3*	5.5
2	8.1	9.2	8.5	8.8
3	14.1	14.3	13.9	14.0
4	22.3	21.4	22.9	22.4
5	9.6	9.7	9.8	9.9
6	13.5	13.4	14.0	13.9
7/8	12.0	11.2	12.3	11.9
9	9.9	10.4	8.9	9.0
10	5.1	5.0	4.4	4.7
Size of City of Residence				
Less than 2,500	5.9***	5.4	5.5**	5.3
2,500 to 10,000	7.4	7.1	8.3	7.8
10,000 to 50,000	15.6	15.2	15.9	16.0
50,000 to 250,000	17.9	18.2	18.0	18.2
250,000 or more	53.2	54.2	52.2	52.6
PMSA or MSA Residence Status*				
In PMSA	41.5***	44.1	44.2***	45.2
In MSA	44.3	43.0	42.1	41.6
In neither	14.1	12.8	13.6	13.2

TABLE D.2 (continued)

Characteristic <sup>a</sup>	Control Group		Program Group	
	Respondents <sup>b</sup>	Respondents and Nonrespondents	Respondents <sup>b</sup>	Respondents and Nonrespondents
Density of Area of Residence*				
Superdense	48.0***	49.6	51.0**	51.4
Dense	27.1	26.8	25.1	25.3
Nondense	24.9	23.6	24.0	23.3
Lived in 57 Areas with a Large Concentration of Nonresidential Females**	40.3	40.0	37.2	37.3
Legal U.S. Resident*	98.9	98.8	98.5	98.6
Job Corps Application Date				
11/94 to 2/95	21.8**	21.7	24.4***	24.3
3/95 to 6/95	31.0	30.0	29.1	28.5
7/95 to 9/95	28.0	28.3	27.5	27.3
10/95 to 12/95	19.2	20.0	19.0	19.9
<b>Fertility and Family Status</b>				
Had Dependents***	18.5***	17.0	15.6**	15.1
Family Status				
Family head	14.7***	14.3	14.2**	13.9
Family member	62.2	61.1	61.3	60.9
Unrelated person	23.1	24.6	24.5	25.2
Average Family Size	3.2***	3.2	3.2	3.2
<b>Education</b>				
Completed the 12th Grade	22.1	21.8	21.9***	21.1
<b>Welfare Dependence</b>				
Public Assistance Receipt				
Received AFDC	29.2***	28.1	28.7	28.3
Received other assistance	15.0	14.5	15.1	15.3
Did not receive	55.8	57.3	56.2	56.4
<b>Health</b>				
Had Any Health Conditions That Were Being Treated	3.4	3.2	3.5**	3.2

TABLE D.2 (continued)

Characteristic <sup>a</sup>	Control Group		Program Group	
	Respondents <sup>b</sup>	Respondents and Nonrespondents	Respondents <sup>b</sup>	Respondents and Nonrespondents
<b>Crime</b>				
Arrested in Past Three Years	11.1**	11.7	11.3	11.4
Ever Convicted or Adjudged Delinquent	5.3**	5.8	5.4	5.6
<b>Completion Status to Previous Interviews</b>				
Baseline Interview Completion Status				
Completed within 45 days	91.6***	88.3	91.1***	89.2
Completed between 46 and 270 days	5.7	5.9	6.2	6.5
Did not complete	2.8	5.8	2.8	4.3
Completed the 12-Month Interview	94.5***	88.5	94.5***	91.2
Completed the 30-Month Interview	88.5***	77.9	88.1***	80.4
<b>Anticipated Program Enrollment Information</b>				
Designated for a Nonresidential Slot***	20.6***	19.7	15.1	14.8
Designated for a CCC <sup>c</sup>	12.4	12.4	13.1	13.0
Designated for a High- or Medium-High-Performing Center <sup>c</sup>	45.7	46.3	46.8	47.0
Designated for a Large or Medium-Large Center <sup>c</sup>	36.3*	37.0	37.2	37.5
<b>Sample Size</b>	<b>3,276</b>	<b>4,212</b>	<b>4,664</b>	<b>5,725</b>

SOURCE: 48-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 D.2 (continued)

---

3. The following cases in the in-person areas are excluded from the calculations: (1) 97 cases (43 control group and 54 program group members) who died between random assignment and the 48-month interview date, (2) 63 cases (31 control and 32 program) who were determined to have enrolled in Job Corps prior to random assignment, and (3) 443 randomly selected program group members who were eligible for 48-month interviews but who were not released for 48-month interviews to reduce data collection costs.

<sup>a</sup>Significance levels pertain to tests of differences between respondents in the program and control groups.

<sup>b</sup>Significance levels pertain to tests of differences between respondents and nonrespondents in the respective research group.

<sup>c</sup>Figures are obtained using data on OA counselor projections about the centers that youths were likely to attend.

\*Difference is significant at the .10 level, two-tailed test.

\*\*Difference is significant at the .05 level, two-tailed test.

\*\*\*Difference is significant at the .01 level, two-tailed test.

The characteristics of program and control group respondents are more similar (Table D.2). Only 3 of the 25 univariate test statistics are statistically significant at the 5 percent level (which is slightly larger than the 1.25 that is expected by chance for 25 independent tests), and the joint test statistic from the multivariate model is statistically insignificant. Thus, although there are some differences in the average baseline characteristics of respondents and nonrespondents in each research group, it does not appear that there are large differences in the average baseline characteristics of program and control group respondents.

**c. The Adjustment of the Weights**

Because of the differences between the characteristics of respondents and nonrespondents, we adjusted the 48-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 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.<sup>31</sup>

To construct the adjusted weights, we estimated models where the probability that a youth in the in-person areas completed the 48-month interview was regressed on a set of control variables.

---

<sup>31</sup>Sample 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 estimated the models using logit maximum likelihood techniques and estimated separate models for program and control group members.

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) region; (4) whether the youth was a family member or family head; (5) whether the youth lived in a superdense, dense, or nondense area at application; (6) the size of city of residence; (7) high school completion status; (8) whether the youth ever had any serious illnesses or injuries; and (9) application date to Job Corps. The models using control group members included 0/1 indicator variables signifying (1) gender; (2) region; (3) whether the youth needed a bilingual program in Job Corps; (4) whether the youth lived in an PMSA, MSA, or neither; (5) the size of city of residence; (6) family size; (7) whether the youth was a family member or family head; (8) whether the youth was arrested in the three years prior to program application; and (9) application date to Job Corps.<sup>32</sup>
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 48-month interview response rate was calculated for each of the six propensity score groups.*** The response rates ranged from about .71 to .89 for the program group, and .58 to .90 for the control group. The variation in the response rates

---

<sup>32</sup>We did not include indicator variables signifying completion status to the baseline interview in the final models, because the response rate to the 48-month interview was much higher for those who completed full baseline interviews than for those who did not (82 percent, compared to 61 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.

suggests that the control variables had some predictive power in explaining whether or not a youth was an interview respondent.

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).<sup>33</sup>

Using these adjusted weights, we found 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 48-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 48-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 that adjust for the control group crossovers. Finally, we

---

<sup>33</sup>The 48-month sample contains youths who completed 48-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 them 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).

discuss how we conducted chi-squared 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:

$$(1) \bar{y} = \hat{\theta}_1 + (1-\hat{\theta})[\hat{e}_s \bar{y}_{2s} + \hat{e}_d \bar{y}_{2d} + \hat{e}_n \bar{y}_{2n}]$$

where:

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

$\bar{y}_1$  = the weighted mean (using the sample design weights) of those in the 48-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.

$\hat{e}_s$ ,  $\hat{e}_d$ ,  $\hat{e}_n$   
= the proportion of the post-45-day population in superdense, dense, and nondense areas, respectively

$\hat{\theta}$  = 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,  $\hat{\theta}$ , 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:

$$(2) \text{ var}(\bar{I}) = \hat{\sigma}^2 \text{ var}(\bar{I}_1) + (1-\hat{\delta})^2 [\hat{e}_s^2 \text{ var}(\bar{I}_{2s}) + \hat{e}_d^2 \text{ var}(\bar{I}_{2d}) + \hat{e}_n^2 \text{ var}(\bar{I}_{2n})],$$

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:

$$(3) \text{ var}(\bar{y}_1) = (1-g)deffw_1 \frac{\hat{\sigma}_1^2}{n_1},$$

where:

- $\hat{\sigma}_1^2$  = 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  $\sigma_i^2$  is calculated in the usual way, and this estimate is inserted in place of  $\sigma_i^2$  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:

$$(4) \text{ var}(\bar{I}_2) = \left[ \sigma_{2w}^2 \left[ \frac{(1 \& g_c)}{n_{2c} a} \% \frac{(1 \& g_p)}{n_{2p} a} \right] \% \frac{(1 \& f) \sigma_{2b}^2}{a} \right] deff_{2w},$$

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  $\hat{\sigma}_{2b}^2$  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  $\hat{\sigma}_{2w}^2$  represents the variance of the measure within areas.

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

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

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 48-month sample contains only 156 post-45-day sample members (92 program and 64 control group members) who lived in the 18 selected dense areas and 163 post-45-day sample members (92 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 the *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.

We can then calculate an estimate of the total variance of the impact estimate, that is, of the expression in equation (2), using the estimated variances for the within-45-day and post-45-day samples. We estimated design effects 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 the entire 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 48-month impact report, we present estimated impacts per eligible applicant, as well as per Job Corps participant that adjust for the control group crossovers. We obtained the impact per participant on an outcome measure by dividing the estimated impact per eligible applicant by the *difference* between the proportion of program group members who enrolled in Job Corps and the proportion of control group members who enrolled in Job Corps during their three-year restriction period.<sup>34</sup> In mathematical terms, the estimated impact per participant ( $I_p$ ) can be expressed as follows:

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

where  $I$  is the estimated impact per eligible applicant,  $S$  is the Job Corps participation (show) rate among the program group, and  $C$  is the early crossover rate among the control group.

The variance of  $I_p$  must account for *both* the variance of  $I$  and the variance of  $(S-C)$ , because both 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:

$$(7) \quad \text{var}(I_p) = \text{var}[I \& I_{p0}(S \& C)] / (S_0 \& C_0)^2,$$

---

<sup>34</sup>For clarity, we refer to these impacts as impacts per participant for the remainder of this section, although it is technically correct to refer to them as impacts per complier.

where  $I_{p0}$  is the “true” but unknown impact on participants,  $S_0$  is the true but unknown show rate, and  $C_0$  is the true but unknown early crossover rate. Using the definition of the variance of the sum of two random variables, equation (7) yields the following expression:

$$(8) \quad var(I_p) = \frac{var(I) + I_{p0}^2[var(S) + var(C)] + 2I_{p0}[cov(I,S) + cov(I,C)]}{(S_0 + C_0)^2}$$

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,  $S$ , and replace  $C_0$  by the estimated early crossover rate,  $C$ .
3. Calculate  $var(S)$  using program group members,  $var(C)$  using control 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  $S$ ,  $cov(I,S) = cov(\bar{y} - \bar{z}, S) = cov(\bar{y}, S)$ , 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,  $cov(\bar{y}, S)$ , can be estimated using the program group as follows:

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

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

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

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

5. The covariance of  $I$  and  $C$ ,  $cov(I,C) = cov(\bar{z}, C)$ , was estimated using control group members and the same procedure as described in step 4 for estimating  $cov(I,S)$ .

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 errors in the show and early crossover rates were very small as a result of the large sample sizes. Thus, the estimated show and crossover rates could almost be treated as constants.

### 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 48-month impact report, we also presented impacts on categorical variables (for example, the type of living arrangement at the 48-month interview or categories of total earnings over the 48-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:

$$(9) \quad \frac{\chi^2_{SR}}{\bar{d}} = \frac{\chi^2_w}{\bar{d}},$$

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

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

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

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 E**

**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 D, 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 48-month analysis sample is large (6,828 program group and 4,485 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 48-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 D, 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).<sup>35</sup>

As discussed in Appendix D, 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

---

<sup>35</sup>There 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 48-month impact analysis, the weight for a youth was constructed to be inversely proportional to the probability that the youth was included in the 48-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 their 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.<sup>36</sup> In addition, the impacts on key outcomes across the other weighting strata did not appear to differ substantially.<sup>37</sup>

---

<sup>36</sup>The 48-month sample contains 5,954 male residents (2,581 controls), 574 male nonresidents (206 controls), 3,283 female residents (1,172 controls), and 1,502 female nonresidents (526 controls). The population weights were .55, .04, .31, and .10, respectively.

<sup>37</sup>We 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.<sup>38</sup>
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 year 4 after random assignment
2. Total earnings during the 48-month period

---

<sup>38</sup>If 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).

3. Proportion of weeks worked in year 4
4. Average hours employed per week in year 4
5. Whether employed in quarter 16
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 48-month period
8. Average months received AFDC/TANF benefits during the 48-month period
9. Average months received food stamp benefits during the 48-month period
10. Whether ever arrested during the 48-month period
11. Whether ever in jail during the 48-month period
12. Whether ever had a child during the 48-month period

Ordinary least squares (OLS) methods were used to estimate models for the continuous outcome measures (for example, average earnings in year 4). 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 E.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 E.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 E.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 E.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 year 4 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 about .35 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 E.2 to E.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 the proportion of weeks worked in year 4 differ by about 5 percent, whereas the standard

TABLE E.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 1995 Dollars)							
Year 4	15.9	3.77	0.000***	16.0	3.64	0.000***	3.3
Entire 48-month period	2.0	2.17	0.346	1.7	1.96	0.376	9.5
Average Percentage of Weeks Employed in Year 4	3.0	0.75	0.000***	2.8	0.74	0.000***	2.0
Average Hours Employed Per Week in Year 4	1.4	0.39	0.000***	1.4	0.38	0.000***	2.1
Percentage Employed in Quarter 16	2.4	0.89	0.007***	2.2	0.88	0.012**	0.6
Received a GED Certificate <sup>a</sup>	15.0	1.04	0.000***	14.7	1.04	0.000***	-0.2
Average Hours per Week Ever in Education or Training	3.5	0.11	0.000***	3.5	0.11	0.000***	0.8
Average Number of Months Received AFDC/TANF Benefits	-0.4	0.21	0.068*	-0.4	0.16	0.027**	22.2
Average Number of Months Received Food Stamp Benefits	-0.5	0.23	0.026**	-0.5	0.18	0.003***	21.9
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-3.7	0.87	0.000***	-3.3	0.82	0.000***	5.7
Percentage Served Time in Jail for Convictions	-2.1	0.71	0.003***	-1.7	0.68	0.011**	4.0
Percentage Had New Children	1.2	0.93	0.184	0.9	0.93	0.327	0.0
<b>Sample Size</b>	<b>11,313</b>			<b>11,313</b>			

TABLE E.2 (continued)

---

SOURCE: Baseline and 12-, 30-, and 48-month follow-up interview data for those who completed 48-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.

<sup>a</sup>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 E.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	Percentage Reduction in the Standard Error
Average Earnings per Week (in 1995 Dollars)							
Year 4	17.7	5.18	0.001***	18.3	5.09	0.000***	1.7
Entire 48-month period	3.0	2.99	0.314	2.4	2.77	0.393	7.2
Average Percentage of Weeks Employed in Year 4	2.6	0.96	0.007***	2.5	0.94	0.009***	1.5
Average Hours Employed per Week in Year 4	1.2	0.52	0.022**	1.2	0.51	0.019**	1.1
Percentage Employed in Quarter 16	1.9	1.12	0.087*	1.6	1.11	0.159	1.0
Received a GED Certificate <sup>a</sup>	13.6	1.31	0.000***	13.4	1.32	0.000***	-0.7
Average Hours per Week Ever in Education or Training	3.5	0.14	0.000***	3.5	0.14	0.000***	1.1
Average Number of Months Received AFDC/TANF Benefits	-0.4	0.13	0.004***	-0.3	0.13	0.009***	3.0
Average Number of Months Received Food Stamp Benefits	-0.6	0.17	0.001***	-0.6	0.16	0.000***	2.6
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-5.1	1.20	0.000***	-4.5	1.17	0.000***	2.5
Percentage Served Time in Jail for Convictions	-3.0	1.05	0.004***	-2.6	1.04	0.013**	1.1
Percentage Had New Children	0.3	1.14	0.771	0.1	1.16	0.946	-1.8
<b>Sample Size</b>	<b>6,528</b>			<b>6,528</b>			

TABLE E.3 (continued)

---

SOURCE: Baseline and 12-, 30-, and 48-month follow-up interview data for those who completed 48-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.

<sup>a</sup>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 E.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 1995 Dollars)							
Year 4	14.1	5.08	0.006***	12.8	5.03	0.011**	1.0
Entire 48-month period	1.1	2.86	0.696	0.8	2.64	0.754	7.5
Average Percentage of Weeks Employed in Year 4	3.7	1.20	0.002***	3.4	1.18	0.004***	1.8
Average Hours Employed per Week in Year 4	1.8	0.57	0.002***	1.6	0.57	0.005***	0.6
Percentage Employed in Quarter 16	3.1	1.43	0.029**	3.2	1.43	0.028**	0.0
Received a GED Certificate <sup>a</sup>	17.5	1.72	0.000***	16.7	1.70	0.000***	1.5
Average Hours per Week Ever in Education or Training	3.4	0.19	0.000***	3.5	0.19	0.000***	0.3
Average Number of Months Received AFDC/TANF Benefits	-0.5	0.42	0.247	-0.4	0.35	0.270	18.1
Average Number of Months Received Food Stamp Benefits	-0.6	0.47	0.222	-0.4	0.38	0.249	17.9
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-1.5	1.09	0.174	-1.7	1.09	0.121	0.3
Percentage Served Time in Jail for Convictions	-0.6	0.71	0.389	-0.5	0.72	0.492	-1.3
Percentage Had New Children	2.3	1.51	0.134	2.1	1.53	0.168	-1.2
<b>Sample Size</b>	<b>4,785</b>			<b>4,785</b>			

TABLE E.4 (continued)

---

SOURCE: Baseline and 12-, 30-, and 48-month follow-up interview data for those who completed 48-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.

<sup>a</sup>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 E.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			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 1995 Dollars)							
Year 4	13.3	5.57	0.017**	12.5	5.43	0.021**	2.5
Entire 48-month period	6.7	3.02	0.027**	6.7	2.75	0.015**	9.1
Average Percentage of Weeks Employed in Year 4	2.7	1.14	0.020**	2.3	1.15	0.045**	-0.6
Average Hours Employed per Week in Year 4	1.2	0.60	0.042**	1.1	0.59	0.069*	0.8
Percentage Employed in Quarter 16	1.9	1.41	0.188	2.1	1.43	0.151	-1.4
Received a GED Certificate <sup>a</sup>	13.6	1.43	0.000***	13.3	1.45	0.000***	-1.2
Average Hours per Week Ever in Education or Training	2.6	0.17	0.000***	2.7	0.18	0.000***	-1.0
Average Number of Months Received AFDC/TANF Benefits	-0.6	0.29	0.052*	-0.4	0.27	0.126	7.2
Average Number of Months Received Food Stamp Benefits	-0.6	0.31	0.046**	-0.5	0.28	0.091*	7.9
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-3.4	1.43	0.019**	-3.4	1.36	0.013**	5.2
Percentage Served Time in Jail for Convictions	-3.5	1.22	0.004***	-3.0	1.15	0.009***	6.0
Percentage Had New Children	0.7	1.44	0.604	0.8	1.46	0.568	-1.4
<b>Sample Size</b>	<b>4,649</b>			<b>4,649</b>			

TABLE E.5 (continued)

---

SOURCE: Baseline and 12-, 30-, and 48-month follow-up interview data for those who completed 48-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.

<sup>a</sup>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 E.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			
	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 1995 Dollars)							
Year 4	3.9	6.78	0.566	4.2	6.73	0.538	0.7
Entire 48-month period	-9.2	3.82	0.015**	-10.0	3.61	0.006***	5.5
Average Percentage of Weeks Employed in Year 4	1.5	1.35	0.253	1.6	1.35	0.229	0.0
Average Hours Employed per Week in Year 4	0.2	0.70	0.755	0.4	0.70	0.572	0.3
Percentage Employed in Quarter 16	1.4	1.57	0.358	0.6	1.58	0.698	-0.8
Received a GED Certificate <sup>a</sup>	16.4	1.93	0.000***	16.4	1.98	0.000***	-2.3
Average Hours per Week Ever in Education or Training	3.6	0.20	0.000***	3.7	0.20	0.000***	-1.4
Average Number of Months Received AFDC/TANF Benefits	-0.2	0.37	0.658	-0.2	0.29	0.505	21.5
Average Number of Months Received Food Stamp Benefits	0.0	0.43	0.936	-0.1	0.33	0.784	21.9
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-4.7	1.51	0.002***	-4.6	1.48	0.002***	2.5
Percentage Served Time in Jail for Convictions	-1.0	1.21	0.412	-0.9	1.20	0.473	1.1
Percentage Had New Children	0.8	1.68	0.621	0.0	1.71	0.999	-1.8
<b>Sample Size</b>	<b>3,577</b>			<b>3,577</b>			

TABLE E.6 (continued)

---

SOURCE: Baseline and 12-, 30-, and 48-month follow-up interview data for those who completed 48-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.

<sup>a</sup>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 E.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 1995 Dollars)							
Year 4	33.5	7.50	0.000***	33.6	7.77	0.000***	-3.5
Entire 48-month period	7.3	4.55	0.110	6.4	4.47	0.154	1.8
Average Percentage of Weeks Employed in Year 4	5.0	1.45	0.001***	4.6	1.51	0.002***	-4.4
Average Hours Employed per Week in Year 4	3.0	0.75	0.000***	2.7	0.78	0.000***	-4.3
Percentage Employed in Quarter 16	4.3	1.63	0.009***	3.5	1.72	0.040**	-5.1
Received a GED Certificate <sup>a</sup>	17.3	2.45	0.000***	17.9	2.72	0.000***	-10.8
Average Hours per Week Ever in Education or Training	4.6	0.23	0.000***	4.7	0.24	0.000***	-2.8
Average Number of Months Received AFDC/TANF Benefits	-0.4	0.45	0.388	-0.3	0.31	0.267	30.7
Average Number of Months Received Food Stamp Benefits	-1.1	0.53	0.037**	-1.1	0.40	0.007***	24.4
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-3.0	1.46	0.039**	-1.5	1.52	0.336	-4.1
Percentage Served Time in Jail for Convictions	-1.2	1.13	0.267	0.0	1.18	0.968	-5.3
Percentage Had New Children	2.6	1.74	0.141	2.8	1.86	0.133	-6.7
<b>Sample Size</b>	<b>3,087</b>			<b>3,087</b>			

TABLE E.7 (continued)

---

SOURCE: Baseline and 12-, 30-, and 48-month follow-up interview data for those who completed 48-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.

<sup>a</sup>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 E.8

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

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 1995 Dollars)							
Year 4	15.7	4.14	0.000***	15.6	3.99	0.000***	3.8
Entire 48-month period	1.9	2.39	0.437	1.5	2.15	0.484	9.8
Average Percentage of Weeks Employed in Year 4	2.9	0.83	0.000***	2.7	0.81	0.001***	2.3
Average Hours Employed per Week in Year 4	1.4	0.43	0.001***	1.3	0.42	0.001***	2.7
Percentage Employed in Quarter 16	2.4	0.97	0.014**	2.1	0.97	0.026**	1.0
Received a GED Certificate <sup>a</sup>	15.2	1.13	0.000***	15.0	1.13	0.000***	-0.3
Average Hours per Week Ever in Education or Training	3.5	0.12	0.000***	3.5	0.12	0.000***	0.9
Average Number of Months Received AFDC/TANF Benefits	-0.4	0.19	0.060*	-0.3	0.17	0.066*	13.4
Average Number of Months Received Food Stamp Benefits	-0.6	0.22	0.013**	-0.5	0.19	0.006***	13.1
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-4.1	0.98	0.000***	-3.8	0.92	0.000***	5.9
Percentage Served Time in Jail for Convictions	-2.5	0.80	0.002***	-2.1	0.77	0.007***	4.8
Percentage Had New Children	1.6	1.01	0.111	1.3	1.01	0.205	-0.2
<b>Sample Size</b>	<b>9,237</b>			<b>9,237</b>			

TABLE E.8 (continued)

---

SOURCE: Baseline and 12-, 30-, and 48-month follow-up interview data for those who completed 48-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.

<sup>a</sup>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 E.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 1995 Dollars)							
Year 4	17.0	8.76	0.053*	18.9	8.82	0.033**	-0.7
Entire 48-month period	3.1	5.05	0.542	3.2	4.70	0.499	6.9
Average Percentage of Weeks Employed in Year 4	3.3	1.77	0.060*	3.7	1.78	0.037**	-0.7
Average Hours Employed per Week in Year 4	1.3	0.88	0.137	1.5	0.89	0.090*	-1.6
Percentage Employed in Quarter 16	2.4	2.06	0.243	2.7	2.10	0.206	-1.9
Received a GED Certificate <sup>a</sup>	14.1	2.57	0.000***	13.2	2.65	0.000***	-3.0
Average Hours per Week Ever in Education or Training	3.4	0.28	0.000***	3.5	0.28	0.000***	-1.0
Average Number of Months Received AFDC/TANF Benefits	-0.4	0.72	0.594	-0.7	0.54	0.208	25.1
Average Number of Months Received Food Stamp Benefits	0.0	0.77	0.974	-0.6	0.56	0.294	26.9
Percentage Arrested or Charged with a Delinquency or Criminal Complaint	-1.7	1.72	0.316	-0.5	1.67	0.772	3.0
Percentage Served Time in Jail for Convictions	0.1	1.24	0.934	0.3	1.24	0.790	0.1
Percentage Had New Children	-0.9	2.23	0.695	-1.4	2.28	0.537	-2.0
<b>Sample Size</b>	<b>2,076</b>			<b>2,076</b>			

TABLE E.9 (continued)

---

SOURCE: Baseline and 12-, 30-, and 48-month follow-up interview data for those who completed 48-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.

<sup>a</sup>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.

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 fairly precise.

Second, because of the 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 impacts of Job Corps for the full sample and for key youth subgroups. This result increases our confidence about the robustness of the impact findings.

## 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.
- Schochet, P. "National Job Corps Study: Methodological Appendixes on the Short-Term Impact Analysis." Princeton, NJ: Mathematica Policy Research, Inc., February 2000.
- Schochet, P., J. Burghardt, and S. Glazerman. "National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes." Princeton, NJ: Mathematica Policy Research, Inc., June 2001.
- 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.