An Evaluation of the National Longitudinal Survey on Youth
Author(s): Thomas MaCurdy, Thomas Mroz and R. Mark Gritz
Source: *The Journal of Human Resources*, Vol. 33, No. 2, Special Issue: Attrition in Longitudinal Surveys (Spring, 1998), pp. 345-436
Published by: University of Wisconsin Press
Stable URL: https://www.jstor.org/stable/146435
Accessed: 29-03-2019 10:46 UTC

An Evaluation of the National Longitudinal Survey on Youth

Thomas MaCurdy
Thomas Mroz
R. Mark Gritz

I. Introduction

The National Longitudinal Survey of Youth (NLSY) has become one of the most widely used data sources for investigating many of the economic and demographic circumstances faced by young adults during the 1980s. The usefulness of the NLSY for empirical analyses relies on the presumption that these data are representative of the population of U.S. youths throughout the 1980s. As the NLSY approaches its third decade, researchers may be concerned about the ongoing representativeness of the NLSY due to the possibility of nonrandom attrition. Also, in contrast to most other data sources, sample members attriting from the NLSY are often recruited back into the sample at a later date, and this "returning" group may also be nonrandom, possibly either offsetting or exacerbating the effects of attrition.

There has been surprisingly little attrition from the NLSY, compared with other longitudinal data sets that have been plagued with such problems. Yet concerns about the nonrandomness of attrition in the NLSY may still be justified, since there has been little research to date on this issue.

To investigate the nature and potential consequences of attrition in the NLSY, this paper studies the reliability of these data by focusing on three issues:

- *Attrition patterns:* Does the selection of particular samples from the NLSY on the basis of attrition status alter the distributions of earnings and other labor-market variables in a way that changes our picture of youths' employment experiences? If so, does weighting offset the contaminating influences?

Thomas MaCurdy is a professor of economics and Senior Fellow at the Hoover Institution, Stanford University. Thomas Mroz is a professor of economics at the University of North Carolina, Chapel Hill. R. Mark Gritz is a research scientist at Battelle Memorial Institute in Seattle. *The National Longitudinal Survey on Youth was funded by the U.S. Department of Labor Bureau of Labor Statistics. Opinions stated in this document do not necessarily represent the official position or policy of the U.S. Department of Labor. The authors gratefully acknowledge many useful comments from Ed Johnson and comments and expert research assistance from Bret Dickey, Hoon Soh, Ya-Chen Shih, and Chris Timmins.*

- *Characteristics of attriters and returnees:* What are the labor-market qualities of those who miss surveys in the NLSY? Also, what are the earnings and work experiences of those who return to the sample?
- *Comparability:* How does the depiction of youths' labor market outcomes in the NLSY differ from that in the Current Population Survey (CPS)?

Our analysis considers these sets of questions in order.

We begin in Section II by describing the structure of the NLSY sample. We then summarize the patterns of attrition and returning found in the NLSY, including breakdowns by year and demographic group. In conjunction, we assess how choice of sample composition and weights alters estimates of the distributions of wages, earnings, and employment.

In Section III, we explore how attrition might corrupt the picture of labor market experiences portrayed in the NLSY by developing a characterization of two groups of people: those who miss interviews in the NLSY, and those who return to the sample after missing interviews. We characterize attriters and returnees by their position in the wage, earnings and employment distributions. Such a characterization is vital to understanding how to use the NLSY and how to interpret its findings. For example, if the youth who attrit tend to be the low earners in the population, then the earnings distributions based on the NLSY will suggest greater growth over time than actually occurs because the lower tail of the distribution will thin due to the systematic loss of its members. Alternatively, if the attriters from the sample are regularly the high-earning members of the population, then distributions based on the NLSY will understate the growth of earnings and the levels of earnings in the later years of the sample. Yet another possibility is that the very high and the very low earners make up the attrition population, in which case the dispersion of earnings measured in the NLSY artificially shrinks over time. Any of these sources of potential contaminants can be mitigated if the "right" attriters return to the sample, and, alternatively, biases can be sharply enhanced if the "wrong" people return.

In Section IV we carry out an extensive comparison of the NLSY with the CPS to explore the point-in-time representativeness of the NLSY. Our study goes substantially beyond what is known to date about the differences between the NLSY and the CPS, examining all 13 years available in the NLSY. It investigates the evolution of hourly wages and earnings in detail for demographic groups distinguished by sex, race-ethnic origin, age, and educational attainment.

Finally, Section V summarizes and interprets our findings.

II. Attrition and Sample Composition of the NLSY

This section summarizes the structure of the NLSY and describes how attrition alters its makeup during the first 13 years of the survey. The NLSY permits analyses of a number of different sample compositions, including some that exclude nonrespondents according to a variety of rules. This discussion explores the consequences of using these alternative samples to estimate wage, earnings, and hours-of-work distributions. Further, it examines how weighting influences inferences about these distributions and how the choice of weights accounts for initial sample design and attrition.

A. Description of the NLSY

The NLSY is a multistage, stratified, clustered probability sample designed to represent the entire population of youth residing in the United States on January 1, 1979. Three independent probability samples make up the NLSY, each comprised of youth born January 1, 1957 through December 31, 1964: (1) a cross-sectional sample designed to be nationally representative of the noninstitutionalized civilian population in the United States; (2) a supplemental sample of the noninstitutionalized civilian population comprised of Hispanics, Blacks, and economically disadvantaged White (namely, non-Hispanic, non-Black) youth; and (3) a military sample representing men and women who were on active duty status on September 30, 1978. Households served as the interview unit for the civilian samples, and the military sample was drawn from rosters of active duty military personnel. The civilian samples were selected after the completion of initial screening interviews of approximately 75,000 dwelling units. Only about 90 percent of the youths originally designated to be a part of the NLSY completed the first-round interviews and were included in the NLSY sample. All findings in this paper abstracts from the ten percent rate of initial interview nonresponse, which is its own form of attrition.

For the purposes of our discussion, it is convenient to consider the NLSY as made up of five distinct components: (1) the cross-sectional sample comprised of 6,111 youths; (2) the supplemental sample of Hispanics with 1,480 youths; (3) the supplemental sample of Blacks containing 2,172 youths; (4) the supplemental sample of economically disadvantaged Whites with 1,643 youths; and (5) the military sample of 1,280 personnel. We restrict our analysis to the 9,763 young men and women included in the first three components of the sample. Two factors lead us to exclude the economically disadvantaged White supplemental sample. First, this particular supplemental sample was discontinued from the NLSY in 1991. Second, this action was taken due to serious suspicions about the representativeness of this supplemental sample; the primary criteria used to screen households into the sample was based solely on household income in 1978—not necessarily parent's income—relative to the poverty level. We do not consider the military sample either because the vast majority of its members were not interviewed after 1983.[1]

Table 1 presents simple unweighted summary statistics for the men and women included in the cross-sectional sample (which we also call the random sample), and the supplemental samples of Hispanics and Blacks. The table reports sample sizes and statistics calculated separately for the three race-ethnic categories Whites, Blacks, and Hispanics. The variables summarized include average hourly earnings (wages), annual reported earnings, annual imputed earnings, and annual hours of work, all measured for the calendar year preceding the year of the interview in nominal dollars. (Henceforth in this section, we express all variables in 1990 dollars.) Annual reported earnings in the NLSY corresponds to a CPS-type measure of annual earnings. The variable used for the annual hours of work measure is a key variable created by the Center for Human Resource Research (CHRR) from the work history data. Average hourly earnings equals annual reported earnings divided by annual hours of work.

A unique feature of the NLSY is the reference period used to collect the work-

1. A total of 201 military respondents were retained from the original military sample of 1,280.

Table 1

Percentage of Observations with Missing Information and Summary Statistics for Earnings and Hours Variables (Nominal Values, Unweighted)

		Descriptive Statistics											
		Average Hourly Earnings			Reported Annual Earnings			Imputed Annual Earnings (from weekly work history)			Annual Hours of Work		
Sample Composition (size)	Percent Not Interviewed	Percent Missing[a]	Percent Zero[b]	Mean (sd)[c]	Percent Missing[a]	Percent Zero[b]	Mean (sd)[c]	Percent Imputed[a]	Percent Zero[b]	Mean (sd)[c]	Percent Missing[a]	Percent Zero[b]	Mean (sd)[c]
Men													
Entire sample (4,837)	7.1	8.8	10.9	7.24 (9.16)	4.9	10.9	12,356 (14,420)	3.0	16.9	17,733 (335,963)	5.9	10.8	1,703 (863)
Random sample (3,003)	6.9	8.4	8.8	7.60 (9.86)	4.1	8.8	13,332 (16,077)	2.9	14.4	18,259 (285,520)	5.9	8.4	1,754 (866)
Whites (2,439)	6.9	8.2	8.0	7.73 (9.56)	3.7	8.0	13,816 (16,837)	2.9	13.5	17,574 (200,710)	6.0	7.5	1,785 (868)
Blacks (1,451)	6.8	9.7	16.4	6.37 (8.11)	6.5	16.4	9,843 (10,339)	2.9	23.3	16,765 (464,529)	5.7	17.0	1,545 (864)
Hispanics (947)	8.2	9.1	10.2	7.11 (9.31)	5.2	10.2	11,892 (11,318)	3.8	15.8	19,513 (402,974)	6.0	9.9	1,706 (816)
Women													
Entire sample (4,926)	5.5	8.2	20.3	6.20 (29.18)	5.3	20.3	8,855 (11,666)	2.0	25.8	16,496 (431,652)	5.1	19.8	1,421 (788)
Random sample (3,108)	5.5	7.9	17.2	6.43 (35.42)	4.6	17.2	9,137 (12,036)	2.1	22.4	17,191 (450,383)	5.1	16.5	1,443 (781)
Whites (2,477)	5.3	7.6	14.9	6.58 (38.77)	4.0	14.9	9,351 (12,114)	2.1	19.8	18,606 (493,019)	5.2	14.1	1,461 (777)
Blacks (1,472)	4.6	8.9	26.9	5.59 (7.64)	6.9	26.9	8,113 (12,674)	1.5	33.1	10,301 (133,135)	4.8	27.0	1,367 (811)
Hispanics (977)	7.4	8.6	24.0	5.92 (6.40)	5.9	24.0	8,432 (8,094)	2.5	29.7	19,307 (514,628)	5.2	23.3	1,380 (778)

a. For variables other than imputed earnings, this column shows the percentage of values associated with negative codings reflecting "don't know," refusal to answer, and valid and invalid skips. For imputed earnings, this column shows the percentage of observations for which retrospective information is available to infer imputed earnings for years with missed interviews.
b. This column shows the percentage of observations with non-missing information (due to non-interview or missing for other reasons) that equal zero.
c. The mean and standard deviation are calculated for samples incorporating only those observations with positive values of the specified variable.

history data. This period extends back to the date of the last interview completed by a respondent and can span two or more calendar years. In sharp contrast, the reference period for the annual reported measures spans only the previous calendar year. Because the work-history data includes both hours and wage information on a weekly basis, one can construct alternative measures of annual earnings. Moreover, not only can one calculate a second measure of annual earnings in the calendar year preceding an interview, one can also impute this measure in the calendar years with missed interviews for respondents who eventually return to the sample. We construct such ''imputed earnings'' for all respondents in all calendar years for which information is available.[2]

The calculation of the descriptive statistics reported in Table 1 does not use weights. The first column presents the percentage of noninterviews for the samples relevant for each row, summed over the first 13 rounds of the NLSY. According to this column, considering observations for all years grouped together, well less than ten percent of the sample is not interviewed, men are more likely to miss interviews, and Hispanic men and women are more likely to miss interviews compared to White and Black respondents. The next four groups of columns present statistics for average hourly earnings, annual reported earnings, annual imputed earnings, and annual hours of work. The first column of each group presents the percentage of observations with missing information.[3] For imputed earnings, the first column in the group indicates the extent to which we are able to impute annual earnings during the calendar years preceding a missed interview for those respondents that return to the sample. For example, in the case of Hispanic men, for 3.8 percent of all calendar years we have imputed earnings but not reported earnings, because of missed interviews. This translates into 46 percent of all calendar years preceding a missing interview by this subgroup (namely, 3.8./8.2). The second column of each group reports the percentage of nonmissing observations with a value of zero. The third column presents means and standard deviations calculated over observations with positive values.[4] Since these statistics do not account for weighting or the sample composition of the NLSY, we do not discuss them here.

B. Definitions of Variables

For a convenient reference, we list the notation and the variables used throughout our analysis. The index t indicates single calendar years, (t = 79, . . . , 91), making

2. We impute our earnings measure by first calculating an earnings measure for each week in which the respondent was working at a job, and then summing weekly earnings over the relevant weeks in each calendar year. Weekly earnings are inferred by multiplying usual hours worked per week and usual hourly wage rate for each job held by a respondent during the specific week and summing across all jobs held during this week. As a matter of convention, if for any reason either usual hours worked or usual wage is missing for a job, the job does not contribute to weekly earnings.

3. Missing information occurs because of item nonresponse (refusal), a ''don't know'' response, a valid skip, or an invalid skip.

4. The relatively large standard deviations result from extreme values of truncated variables or misreported/misrecorded information; these large standard deviations point out the importance of using statistics that are not unduly influenced by these extreme values. Extreme values for reported earnings result from the replacement of values above $100,001 with the average earnings of respondents who are U.S. residents and who reported values above this threshold in the latter years of the sample. The extreme values for imputed earnings stem from misreported/misrecorded values for hourly wage rates in the work-history data. Hours of work are also truncated in the data at 96 hours per week. In our subsequent analysis, we top code all earnings data to $100,001, which is the level used in the earlier years of the NLSY.

up the sample period covered by the NLSY. The index α denotes the ages (α = 14, . . . , 33) of individuals observed over this period. T represents an index for the year ranges or time brackets (T = 79–83, 84–87, 88–91), and A indexes the age brackets (A = 14–17, 18–20, 21–25, 26–30, 31+).

The primary variables considered in the analysis are:

ω_{it} = average hourly earnings, the wage rate, of individual i in year t (in 1990 dollars).
y_{it} = annual earnings of individual i in year t (in 1990 dollars).
h_{it} = annual hours worked by individual i in year t.
E_{it} = 0–1 indicator variable that takes a value of one if individual i is interviewed and is employed in year t.[5]
δ_{it} = 0–1 indicator variable that takes a value of one if individual i attrits (namely is not interviewed) in year t.
δ^*_{it} = 0–1 indicator variable that equals one if any of the $\delta_{ij} = 1$ for $j \leq t - 1$, which signifies that individual i attrited in a past year and returned to become a member of the sample again before the last available interview date in 1991.
S_{it} = 0–1 indicator variable that takes a value of one if individual i is interviewed and is in school more than five months during year t.
$D_{\alpha it}$ = 0–1 dummy variables indicating the age α of individual i in year t.
$D_{A\cdot\cdot}$ = vector of 0–1 dummy variables indicating the age bracket A for which the dummy variable $D_{\alpha it} = 1$.

When referring to an arbitrary or generic statistic calculated for some function of the above variables, we designate:

$\mu_{\alpha t}$ = statistic computed for age group α in year t.
μ_{AT} = statistic computed for age bracket A over years included in the range T.

Examples of such statistics include:

$P\#\#_{**}(X_{it})$ = ## percentile of the variable X_{it} calculated for sample of individuals i included in the group "**".

The notation "**" in this definition refers to any combination of age and time brackets.

C. Patterns of Attrition

Tables 2a and 2b summarize the cumulative effects of attrition on the sample composition of the NLSY, with Table 2a reporting results for men and 2b listing findings for women. The first set of rows shows the fraction of original respondents who are

5. We define employment to be that condition when individuals report both positive hours of work and positive annual earnings. In a more extensive analysis, we considered this definition along with three other specifications of employment rates: (1) positive annual hours of work; (2) positive annual reported earnings; (3) positive hours and positive total labor income. We found little impact on employment rates using these alternative definitions with one notable exception; specifically, defining employment as merely having positive annual reported earnings practically doubles the employment rate of 14 to 17 year old men and women. See MaCurdy, Mroz, and Gritz (1994) for further discussion of this issue.

Table 2a
Total Attrition and Fraction of Sample Who are Returnees for Men

Race-Ethnic Groups	Year 80	81	82	83	84	85	86	87	88	89	90	91
Percentage of Original Sample Not Interviewed by Year												
All	4.3	3.6	4.5	4.1	5.3	6.8	9.2	11.2	10.7	10.0	11.5	11.0
Whites	3.6	3.8	3.8	3.4	5.6	6.7	9.0	10.6	10.5	10.3	11.6	10.5
Blacks	4.2	2.8	4.8	4.1	4.6	7.0	9.1	10.6	9.5	9.1	10.6	11.7
Hispanics	6.3	4.5	6.2	5.5	5.5	7.1	9.8	13.7	13.3	10.7	12.8	11.3
Percentage of Interviewed Population Who are Returnees												
All	—	2.7	3.7	5.4	6.5	7.7	8.5	10.0	13.3	15.9	16.8	19.0
Whites	—	2.0	3.3	4.5	5.0	6.3	6.9	8.3	11.1	13.2	14.0	16.3
Blacks	—	2.8	3.2	5.5	6.6	7.4	8.6	10.9	14.5	17.1	18.6	20.3
Hispanics	—	4.3	5.5	7.7	10.1	11.7	12.3	13.2	17.3	20.8	20.9	23.8

not interviewed in each year, and the second set reports the fraction of interviewees who are returnees (former attriters) in each year. The tables show that:

- About 10 percent of the original respondents are regularly missing in the latest years of the NLSY. (This figure is somewhat higher for men, and lower for women.).
- Around 20 percent of the male interviewees are returnees in the latest years, while less than 15 percent of the women are returnees.

Tables 3a and 3b provide more detail on the patterns of initial attrition and returning to the sample. The columns in this table represent the year of *first* attrition. The top set of rows shows the percentage of individuals who miss an interview for

Table 2b
Total Attrition and Fraction of Sample Who are Returnees for Women

Race-Ethnic Groups	Year 80	81	82	83	84	85	86	87	88	89	90	91
Percentage of Original Sample Not Interviewed by Year												
All	3.8	3.3	3.9	3.2	4.1	5.2	7.0	8.1	9.0	7.3	8.9	8.0
Whites	3.8	3.2	3.7	3.3	4.5	5.3	7.0	7.7	8.1	7.1	8.3	7.3
Blacks	3.3	2.7	2.9	2.6	3.1	3.6	5.0	6.0	7.5	6.3	8.4	8.4
Hispanics	4.3	4.3	5.6	4.1	4.8	7.3	10.3	12.1	13.4	9.3	11.0	9.1
Percentage of Interviewed Population Who are Returnees												
All	—	2.0	2.8	4.4	5.0	5.8	6.2	7.4	9.1	11.8	12.0	13.7
Whites	—	2.1	2.7	4.0	4.4	5.2	5.8	6.9	8.5	10.3	10.7	12.3
Blacks	—	1.6	2.2	3.8	4.5	5.1	5.4	6.6	7.9	10.6	10.1	11.6
Hispanics	—	2.4	3.9	6.4	7.2	8.4	8.7	10.4	12.5	17.4	18.2	20.6

Table 3a
Attrition Patterns for Men

Sample Composition	80	81	82	83	84	85	86	87	88	89	90	91
	Year											
First-Time Attrition Rate (Percentage of Remaining Population Who Attrit for First Time)												
Race-Ethnic Groups												
All	4.3	2.0	2.0	1.2	2.4	2.9	3.3	3.9	3.1	2.2	2.7	2.1
Whites	3.6	2.2	1.3	0.8	2.8	2.5	3.1	3.3	2.8	2.3	2.3	1.5
Blacks	4.2	1.4	2.4	1.7	1.7	3.3	3.5	4.1	2.9	2.7	3.4	3.3
Hispanics	6.3	2.5	3.0	1.5	2.5	3.5	3.6	5.3	4.2	1.3	2.5	2.0
Cohort												
14–15	2.4	0.9	1.6	1.2	2.1	2.0	3.1	3.0	3.3	2.0	2.5	2.6
16–17	3.9	1.9	1.9	1.2	2.0	2.9	3.6	4.5	2.6	2.3	2.7	1.9
18–19	5.3	2.1	1.9	1.5	2.8	3.5	2.6	4.3	3.1	2.9	3.2	2.0
20–21	5.9	3.1	2.6	1.2	2.6	3.4	3.8	3.8	3.4	1.7	2.2	1.9

Percentage of Attriters Who Never Return												
Cohort												
14–15	11.1	10.0	38.9	0.0	13.6	14.3	18.8	20.0	12.5	47.4	30.4	—
16–17	3.7	24.0	8.0	13.3	23.1	30.6	22.7	20.8	20.7	20.0	27.6	—
18–19	7.6	20.8	22.7	17.6	19.4	28.9	14.8	14.0	36.7	40.7	34.5	—
20–21	16.1	35.7	8.7	30.0	27.3	25.0	33.3	13.8	36.0	41.7	26.7	—
Percentage of Attriters Who Miss Multiple Interviews & Return at Least Once												
Cohort												
14–15	55.6	60.0	27.8	53.8	50.0	52.4	43.8	30.0	21.9	26.3	—	—
16–17	63.0	48.0	56.0	46.7	61.5	52.8	31.8	32.1	10.3	24.0	—	—
18–19	56.1	45.8	54.5	76.5	54.8	50.0	55.6	27.9	16.7	22.2	—	—
20–21	53.6	50.0	60.9	30.0	45.5	42.9	23.3	37.9	20.0	25.0	—	—
Number and Length of Spells for Multiple-Yr Attriters												
Average number attrition spells[a]	2.1	2.0	2.2	1.9	1.9	1.6	1.7	1.7	1.4	1.3	—	—
Average missing years:												
1980–83	2.4	2.0	1.4	1.0	—	—	—	—	—	—	—	—
1984–87	1.5	1.7	1.8	2.1	2.5	2.3	1.6	1.0	—	—	—	—
1988–91	1.6	1.7	1.7	1.5	1.8	1.4	1.4	1.8	2.4	2.0	—	—

a. This row reports the average number of attrition spells experienced by individuals who initially attrit in the specified year, who attrit for more than one year, and who return at sometime after this initial attrition.

Table 3b
Attrition Patterns for Women

Sample Composition	Year											
	80	81	82	83	84	85	86	87	88	89	90	91
First-Time Attrition Rate (Percentage of Remaining Population Who Attrit for First Time)												
Race-Ethnic Groups												
All	3.8	1.5	1.4	1.0	1.5	2.0	2.4	2.4	2.8	1.1	1.9	1.1
Whites	3.8	1.4	1.1	0.9	1.7	1.7	2.3	1.9	2.2	0.9	1.8	0.7
Blacks	3.3	1.1	0.9	1.3	1.2	1.2	1.6	2.4	3.0	1.6	1.6	1.7
Hispanics	4.3	2.4	3.0	1.0	1.6	3.8	3.6	3.8	3.9	1.1	2.7	1.0
Cohort												
14–15	2.8	0.7	0.8	0.7	1.3	1.2	2.7	3.3	3.4	0.9	1.8	1.7
16–17	2.7	1.3	1.4	1.1	1.3	1.8	2.4	1.5	2.8	0.9	2.3	0.8
18–19	4.5	2.0	1.8	0.8	1.8	2.0	2.2	2.3	2.4	1.6	1.7	0.9
20–21	5.2	1.8	1.6	1.5	1.6	2.7	2.2	2.8	2.5	1.0	1.7	1.1

Percentage of Attriters Who Never Return												
Cohort												
14–15	16.7	0.0	0.0	42.9	23.1	41.7	7.7	29.0	32.3	50.0	25.0	—
16–17	21.6	17.6	11.1	33.3	23.5	21.7	30.0	5.3	20.6	40.0	29.6	—
18–19	12.1	20.8	9.1	22.2	23.8	26.1	28.0	12.0	26.9	35.3	38.9	—
20–21	12.3	10.5	18.8	20.0	25.0	19.2	19.0	15.4	21.7	44.4	26.7	—
Percentage of Attriters Who Miss Multiple Interviews & Return at Least Once												
Cohort												
14–15	60.0	57.1	62.5	42.9	46.2	50.0	69.2	25.8	12.9	12.5	—	—
16–17	51.4	70.6	38.9	53.3	58.8	56.5	43.3	47.4	23.5	30.0	—	—
18–19	55.2	50.0	63.6	55.6	57.1	47.8	56.0	40.0	34.6	29.4	—	—
20–21	64.9	68.4	37.5	46.7	43.8	53.8	61.9	34.6	26.1	33.3	—	—
Number and Length of Spells for Multiple-Yr Attriters												
Average number attrition spells[a]	1.9	1.9	2.1	1.9	1.9	1.6	1.6	1.6	1.4	1.6	—	—
Average missing years:												
1980–83	2.5	2.0	1.4	1.0	—	—	—	—	—	—	—	—
1984–87	1.4	1.6	2.3	1.7	2.5	2.3	1.7	1.0	—	—	—	—
1988–91	1.2	1.5	2.1	1.4	1.7	1.5	1.6	1.9	2.6	2.0	—	—

a. This row reports the average number of attrition spells experienced by individuals who initially attrit in the specified year, who attrit for more than one year, and who return at sometime after this initial attrition.

the first time in the year designated in the column, with percentages computed using those who have continuously remained in the sample since the baseline interview. (Thus, these percentages are estimates for the hazard rates for first leaving the NLSY.) Results are given for the entire sample, the three race-ethnic groups, and four age cohorts based on age at the time of the baseline interview. The second set of rows gives the fraction of these initial attriters in each year who never return to the sample by 1991, by age cohort. The third set of rows reports the fraction of these attriters who miss more than one interview but who return in at least one year after initial attrition, again by age cohort. The fourth set of rows gives the average number of spells experienced by multiple-year attriters (identified in the previous set of rows), along with the average number of years that they are missing during the year brackets T = 1980–83, 1984–87, 1988–91.

We see that initial attrition rates never exceed six percent for men or women, and these rates are typically in the one percent to three percent range. Among the race-ethnic groups, Hispanics experience the highest rates of initial attrition, especially in the early years of the survey. Among the cohorts, attrition rates tend to rise with age, once again particularly early in the survey. The second set of rows (percentage of attriters who never return) show that of the persons who attrit in the early years, around 10 to 25 percent are never seen again by 1991. This range rises to as much as 50 percent in the later years reflecting, in part, the shorter length of time available for finding these individuals and recruiting them back into the sample. The results presented in the third set of rows show that around 50 percent of those missing at least one interview and who are not lost forever (namely, return at some time before 1991) end up missing two or more interviews. The figures reported for the number of spells experienced by multiple-year attriters indicate that intermittent periods of absence are common. Finally, the findings for the average numbers of missing years during the various time horizons, listed in the bottom rows of the tables, further show that periods of absence for the multi-year attriters involve about the same number of missed interviews in the early years in the NLSY as in its later interviews.

D. Choices of Sample Composition and Weighting

The structure of the NLSY permits a number of different sample compositions and alternative approaches to weighting the sample. The ones that we examine here vary in three dimensions: which sample components are included, how attrition is treated, and which weights are used (if any).

There are two prevalent combinations of the sample components that are widely used in analyses of the NLSY data. One combines all interviewed observations from the cross-sectional and supplemental probability samples of the noninstitutionalized civilian population, excluding the supplemental sample of economically-disadvantaged Whites. We refer to this sample composition as the "entire sample." The other configuration uses all interviewed observations from only the cross-sectional probability sample, which we refer to as the "random sample."

A popular way to deal with attrition in many longitudinal studies is to restrict the empirical analysis to individuals who are interviewed in every year, excluding anyone who ever misses an interview. We designate such a sample composition as a ''continuous'' sample because it includes only individuals interviewed in every survey year from 1979 to 1991.

Researchers often rely on sample weights to account for initial sample design and attrition. The NLSY provides two sets of weights. Base-year weights to account for the original design and the differential probabilities of completing the base-year interviews. The NLSY also updates base-year weights every year to account for nonresponse at each interview using a post-stratification adjustment procedure described in Frankel, McWilliams and Spencer (1983).[6] We refer to the base-year weights as ''*79 weights*'' and the updated weights as ''current weights.''

Before examining the consequences of these particular samples, it is vital to recognize that the base-year and current-year weights supplied by the NLSY are inappropriate for the sample compositions considered here, as well as for those typically analyzed elsewhere in the literature. Specifically, the design of the NLSY weights make them appropriate for analyzing the combined cross-sectional and supplemental samples *including* the economically-disadvantaged White supplemental sample. However, we exclude this latter supplemental sample, as do most researchers.[7] Thus, the weights supplied in the NSLY are wrong for analyzing any of the samples considered in this paper. To construct the appropriate base-year and current-year weights for the samples considered here, we construct two sets of both the 79 (base-year) and current weights.[8] One set of our 79 weights applies to the combined cross-sectional, Black supplemental, and Hispanic supplemental samples (namely, the entire sample); and the other set applies to the cross-sectional sample (namely, the random sample).[9] We also calculate analogous sets of current-year weights that adjust for attrition following the approach of Frankel, McWilliams and Spencer (1983). In analyses of the NLSY similar to the ones presented here, we learned that using these properly constructed weights matters in the sense that they generate different statistics for earnings and hours than those produced by the weights currently available in the NLSY. All of the findings presented here use these revised weights.

E. Consequences of Sample Composition and Weighting

How sensitive are depictions of labor-market data to alternative sample compositions of the NLSY? Tables 4 and 5 address this question.

1. Distributional Statistics for Wages, Earnings, and Hours

Tables 4a and 4b investigate the extent to which the use of alternative sample compositions and weighting schemes influence the distribution of wages and annual re-

6. The adjustment to the sampling weights take into account differential interview nonresponse rates that are explainable by respondents' age, sex, race-ethnicity, and whether the household provided complete income information in the initial household screening interview.

7. The economically disadvantaged White subsample is excluded from many analyses for two reasons: first, the definition of ''economically disadvantaged'' is problematical due to inconsistencies in evaluating family income for those respondents who had already left their parent's household at the time of the initial interview; and, second, this subsample was discontinued in 1991.

8. We thank Randy Olsen of CHRR for providing us the data needed to construct the appropriate weights.

9. The two sets of base year weights are constructed using the five step procedure described in Frankel, McWilliams and Spencer (1983), which was used to develop the original base year weights for the NLSY.

Table 4a
Summary Statistics for Wages and Annual Earnings for Men for Various Sample Compositions[a]

		Average Hourly Earnings				Annual Earnings			
			Percentiles				Percentiles		
Sample Composition	Years	Trimmed Mean[b]	10 Percent	50 Percent	90 Percent	Trimmed Mean[b]	10 Percent	50 Percent	90 Percent
Entire sample, current weights	79–83	6.09	2.12	5.54	11.98	8,542	1,078	6,470	21,568
	84–87	8.18	3.09	7.52	15.42	15,464	2,516	14,382	32,198
	88–91	10.56	4.42	9.85	19.05	22,672	7,677	21,608	41,983
Differences from Entire Sample, Current Weights[c]									
Entire sample, 79 weights	79–83	−0.02	−0.01	−0.01	0.00	−58	−22	−64	0
	84–87	−0.05	−0.02	−0.06	−0.13	−159	0	−72	−239
	88–91	−0.03	−0.01	−0.03	−0.12	−96	−88	−113	−349
Random sample, current weights	79–83	0.03	0.08	0.09	0.12	89	71	302	373
	84–87	0.06	0.13	0.16	0.15	185	360	575	596
	88–91	0.11	0.21	0.28	0.22	252	334	527	199

Random sample, unweighted	79–83	−0.30	−0.04	−0.16	−0.23	−729	−66	−511	−575
	84–87	−0.42	−0.08	−0.22	−0.34	−1,121	0	−576	−749
	88–91	−0.53	0.00	−0.09	−0.23	−1,527	0	−496	−719
Continuous entire, current	79–83	−0.01	0.03	−0.01	−0.16	−23	−28	0	0
weights	84–87	0.01	0.05	0.03	0.05	28	108	194	17
	88–91	0.02	0.12	0.12	−0.13	6	57	463	−240
Continuous random, current	79–83	0.04	0.10	0.06	0.03	61	6	127	233
weights	84–87	0.06	0.17	0.16	0.17	185	377	575	598
	88–91	0.12	0.32	0.33	0.18	293	755	1,392	178
Continuous random, unweighted	79–83	−0.30	−0.02	−0.20	−0.37	−745	−103	−576	−743
	84–87	−0.42	−0.05	−0.22	−0.36	−1,114	0	−576	−749
	88–91	−0.52	0.07	0.01	−0.31	−1,476	57	392	−876

a. Wages refer to average hourly earnings expressed in 1990 $.

b. Trimming removes the lower and the upper five percent of the observations.

c. These differences show the value of the statistic computed using the sample composition designated below in the first column minus the corresponding statistic computed using the entire sample, current weights. Thus, a positive value means that the entire-sample-current-weights statistic is lower.

Table 4b

Summary Statistics for Wages and Annual Earnings for Women for Various Sample Compositions[a]

		Average Hourly Earnings				Annual Earnings			
			Percentiles				Percentiles		
Sample Composition	Years	Trimmed Mean[b]	10 Percent	50 Percent	90 Percent	Trimmed Mean[b]	10 Percent	50 Percent	90 Percent
Entire sample, current weights	79–83	4.99	1.81	4.77	8.94	6,083	719	4,741	14,533
	84–87	6.66	2.36	6.31	12.17	10,676	1,381	10,136	22,658
	88–91	8.47	3.12	7.94	15.46	15,097	2,740	14,363	29,748
Differences from Entire Sample, Current Weights[c]									
Entire sample, 79 weights	79–83	−0.01	−0.01	−0.01	−0.01	−26	−6	−11	−98
	84–87	−0.02	−0.01	−0.01	−0.04	−47	−17	−72	−15
	88–91	0.00	0.00	0.00	0.00	3	0	0	0
Random sample, current weights	79–83	0.00	0.03	0.04	0.01	19	1	17	365
	84–87	0.03	0.07	0.08	0.09	69	77	219	353
	88–91	0.05	0.07	0.12	0.23	81	139	31	252

Random sample, unweighted	79–83	−0.17	−0.01	−0.1	−0.16	−396	−37	−240	−155
	84–87	−0.23	0.01	−0.04	−0.14	−452	0	−72	−141
	88–91	−0.27	0.01	−0.09	−0.14	−504	22	−134	−248
Continuous entire, current	79–83	0.00	−0.01	−0.03	−0.01	0	−30	−48	−98
weights	84–87	0.01	0.04	0.00	0.00	13	6	149	0
	88–91	0.01	0.04	0.03	0.02	55	139	31	252
Continuous random, current	79–83	0.00	0.02	0.00	0.01	16	1	17	218
weights	84–87	0.04	0.13	0.11	0.09	71	129	334	353
	88–91	0.05	0.13	0.17	0.23	146	260	393	252
Continuous random, unweighted	79–83	−0.17	−0.02	−0.13	−0.15	−396	−42	−347	−155
	84–87	−0.22	0.07	−0.02	−0.16	−453	34	0	−141
	88–91	−0.26	0.05	−0.05	−0.15	−485	139	0	−235

a. Wages refer to average hourly earnings expressed in 1990 $.
b. Trimming removes the lower and the upper five percent of the observations.
c. These differences show the value of the statistic computed using the sample composition designated below in the first column minus the corresponding statistic computed using the entire sample, current weights. Thus, a positive value means that the entire-sample-current-weights statistic is lower.

Table 5a
Summary Statistics for Hours of Work for Men for Various Sample Compositions

Sample Composition	Years	Percent Working[a]	Percent Zero Earnings[b]	Hours Percentiles for Workers[c] 10 Percent	50 Percent	90 Percent
Entire sample current weights	79–83	76.2	2.06	312	1,320	2,355
	84–87	88.95	2.79	600	2,080	2,749
	88–91	91.10	4.09	1,180	2,080	2,962
Differences from Entire Sample, Current Weights[d]						
Entire Sample 79 Weights	79–83	−0.22	−0.01	0	−10	−3
	84–87	−0.08	−0.02	0	0	−5
	88–91	−0.04	−0.02	−4	0	−2
Random Sample Current Weights	79–83	1.43	−0.20	10	25	28
	84–87	1.28	−0.01	22	0	30
	88–91	0.94	−0.2	60	40	24
Random Sample Unweighted	79–83	−0.39	−0.07	−12	−50	−15
	84–87	0.57	−0.06	−5	0	−8
	88–91	0.29	−0.02	−1	0	−2
Continuous Entire Current Weights	79–83	0.46	0.01	0	−20	−13
	84–87	1.07	−0.29	0	0	−9
	88–91	0.63	−0.17	40	34	−2
Continuous Random Current Weights	79–83	1.88	−0.19	8	0	18
	84–87	2.30	−0.31	24	0	14
	88–91	1.44	−0.31	100	60	17
Continuous Random Unweighted	79–83	0.12	−0.08	−15	−70	−15
	84–87	1.55	−0.33	−8	0	−11
	88–91	0.85	−0.15	26	28	−11

a. Percentage of interviewed individuals reporting positive annual hours of work.
b. Percentage of individuals with positive hours of work who reported zero annual earnings.
c. The calculations of these percentiles restricts the sample to observations where hours are greater than zero and earnings are greater than zero.
d. These differences show the value of the statistic computed using the sample composition designated below in the first column minus the corresponding statistic computed using the entire sample, current weights. Thus, a positive value means that the entire-sample-current-weights statistic is lower.

ported earnings. Table 4a reports descriptive statistics for men, and Table 4b gives results for women. The top set of rows present trimmed means (calculated by removing the upper and lower 5 percent of observations) and the 10th, 50th, and 90th percentiles of average hourly earnings and annual reported earnings (in 1990 dollars), calculated for the entire sample with current weights, with breakdowns for the years 1979–83, 1984–87, and 1988–91. The six sets of rows below this first group show the differences obtained using other sample compositions and weighting schemes. An entry in these rows denotes the difference between the statistic calculated with the sample composition and weighting designated in the first column of the row and the estimated value of the corresponding statistic calculated using the entire sample with current-year weights. For example, the estimate of the trimmed mean of average

Table 5b
Summary Statistics for Hours of Work for Women for Various Sample Compositions

Sample Composition	Years	Percent Working[a]	Percent Zero Earnings[b]	Hours Percentiles for Workers[c] 10 Percent	50 Percent	90 Percent
Entire sample current weights	79–83	70.63	2.28	236	1,081	2,080
	84–87	81.99	2.85	380	1,730	2,280
	88–91	81.07	6.19	528	1,965	2,470
Differences from Entire Sample, Current Weights[d]						
Entire Sample 79 Weights	79–83	−0.20	−0.02	−2	−1	0
	84–87	0.01	−0.01	0	−6	0
	88–91	0.02	0.00	0	0	2
Random Sample Current Weights	79–83	2.01	−0.29	7	27	0
	84–87	1.57	−0.18	14	20	6
	88–91	0.71	−0.05	11	5	6
Random Sample Unweighted	79–83	−0.16	−0.08	2	−27	0
	84–87	−0.12	−0.10	0	−10	−10
	88–91	−0.09	0.01	0	5	−10
Continuous Entire Current Weights	79–83	0.52	−0.04	−4	−1	0
	84–87	0.77	−0.02	0	2	5
	88–91	0.88	−0.32	32	11	16
Continuous Random Current Weights	79–83	2.54	−0.30	4	21	0
	84–87	2.30	−0.20	12	20	8
	88–91	1.52	−0.36	44	15	22
Continuous Random Unweighted	79–83	0.52	−0.12	2	−31	0
	84–87	0.67	−0.15	4	−7	0
	88–91	0.73	−0.29	32	11	5

a. Percentage of interviewed individuals reporting positive annual hours of work.
b. Percentage of individuals with positive hours of work who reported zero annual earnings.
c. The calculations of these percentiles restricts the sample to observations where hours are greater than zero and earnings are greater than zero.
d. These differences show the value of the statistic computed using the sample composition designated below in the first column minus the corresponding statistic computed using the entire sample, current weights. Thus, a positive value means that the entire-sample-current-weights statistic is lower.

hourly earnings using current-year weights for the entire sample of men interviewed in 79–83 equals to $6.09 according to Table 4a. The difference between this estimate and the trimmed mean for the entire sample with 79 (base-year) weights is −0.02, which implies that the 79-weight trimmed mean is $6.07.

Table 5a and 5b show similar statistics for employment and the distributions of annual hours of work. The descriptive statistics listed in each column include: the percentage of interviewed individuals reporting positive hours of work; the percentage of individuals reporting positive hours of work but no annual earnings; and the 10th, 50th, and 90th percentiles of the annual-hours-of-work distribution for the individuals classified as employed (namely, those reporting both positive hours of work and positive annual earnings). Analogous to Tables 4a and 4b the top set of rows contains the five descriptive statistics calculated for the entire sample using current weights, with results listed for the years 1979–83, 1984–87, and 1988–91. The six sets of rows below this first group show the differences obtained by using other sample compositions.

2. *Effects of Sample Compositions on Distributional Statistics*

Considered together, Tables 4a, 4b, 5a, and 5b clearly convey the impression that varying sample compositions and weighting does not substantially alter the distributions of wages, earnings, or hours. Regarding weighting, two conclusions emerge. First, we see:

- The choice between base-year or current-year weights makes little difference in calculating summary statistics for wages, earnings, and hours of work.

The rows in the lower portions of Tables 4a, 4b and 5a, 5b reveal that the choice between base-year and current-year weights has an imperceptible effect on all distributions. Several percentiles of the distributions of wages, earnings, and hours are literally identical, especially for women. The largest percentage difference implied by any estimate is only 1.2 percent—associated with the 10th percentile of 84–87 annual earnings in Table 4b—which corresponds to a $17 absolute difference in annual earnings.

Second, these tables support the conclusion:

- The use of weights uniformly raises the wage, earnings, and hours-of-work distributions.

This can be seen by comparing the results for the current-weighted and the unweighted random samples in the third and fourth set of rows in Tables 4a, 4b, 5a, and 5b. These comparisons show that the differences in percentiles calculated using weights are always positive, and the unweighted results are almost always negative. The upward shift in the distributions due to weighting is somewhat larger for men than for women, with more than 10 percent increases in some percentiles common in the case of annual earnings of men. In Table 4a, for example, introducing current-year weights into the calculation increases the median of annual earnings by $813 in the 1979–83 period (i.e., from $511 below the value in the top row of the table to $302 above that value), which translates into a 13 percent differential (namely, $813 is 13 percent of $6,470). In 1984–87, the 10th percentile difference is $360, which implies a 14 percent differential. The use of weights also uniformly raises the distribution of average hourly earnings, employment rates, and the distribution of annual hours of work, although the effects are smaller. A comparison of the current-weighted and unweighted continuous random samples, shown in the last two sets of rows, conveys a similar picture, although it is important to note that the weights we constructed were not designed to be used with the continuous sample.

Concerning differences in Tables 4a, 4b, 5a, and 5b related to sample compositions, two conclusions emerge: First, we see:

- Distributions of wages, earnings, and hours of work computed using only the random sample of the NLSY are uniformly higher than those based on the entire sample, but only marginally.

These tables reveal some differences between the random and the entire samples in employment rates and in the lower tails of the earnings distributions. Tables 5a and 5b shows that employment rates are 0.7–2.0 percentage points higher for men and women in the random sample than in the entire sample using current weights. This discrepancy, however, is relatively small given that all rates exceed 70 percentage points. In addition, comparing all the wage, earnings, and hours statistics for the current-weighted random and entire samples in Tables 4a, 4b, 5a, and 5b, we see only one relatively large difference located at the 10th percentile of earnings for men in 1984–87 yielding a differential of 14 percent (namely, the difference is $360 and the 10th percentile for the entire sample is $2,516). With this exception, none of the remaining differences imply differentials even as large as five percentage points, and the vast majority are quite small.

A careful inspection of the findings in Tables 4a, 4b, 5a, and 5b further reveals a second conclusion regarding the influence of sample composition:

- Keeping attriters in the sample introduces a variety of subtle differences in the distributions of earnings and hours of work.

When comparing most sample compositions with their continuous counterparts, one sees that the 10th percentiles are typically lower for those samples containing returnees, especially in the later years of the sample. In addition, keeping the attriters in the samples also slightly decreases the employment rate by about one percentage point. The existence of such an effect suggests that returnees contribute disproportionately to the lower tail of the earnings and hours of work distributions.

In addition to the small number of comparisons presented in this discussion, we did comparisons using all possible combinations of the samples and weights from the previous subsection,[10] and we find that the general patterns are quite robust. We also find that similar patterns surface in investigations of other earnings measures available in the NLSY, including imputed earnings.[11]

F. Does Sample Composition and Weighting Matter?

The above evidence supports three broad conclusions: (1) weighting has substantial effects on the estimated distributions of earnings and hours of work in the NLSY; (2) it does not matter whether baseline or current weights are used in the estimation of these distributions—no apparent gain is evident from using updated weights that adjust for a changing demographic makeup of the NLSY due to attrition; and (3)

10. See MaCurdy, Mroz, and Gritz (1994) for further evidence on this subject.

11. In particular, we checked many of these conclusions using (1) annual imputed earnings and (2) total labor income, defined as the sum of annual reported earnings and business income. For a detailed description and an analysis of the sources of the differences among reported earnings and these alternative measures of earnings, see Cameron, Gritz, and MaCurdy (1989).

including the supplemental Black and Hispanic samples or eliminating attriters has only marginal influences on the distributions of earnings and employment data in the NLSY.

The significance of using weights in the NLSY suggests that it is important to include five new sets in future public releases of this survey. The weights supplied at the moment with the NLSY are inapplicable for use with most sample compositions analyzed in the literature due to their assumption that researchers incorporate the economically-disadvantaged White supplemental sample in their analyses, which is rarely the case. Many longitudinal analyses rely on continuously interviewed samples, and the NLSY weights are inappropriate for use with these samples as well.

The first two sets of weights that we recommend for inclusion in the NLSY are those we constructed for this analysis: one set appropriate for analyzing only the cross-sectional component of the sample; and another set suitable for using our entire sample (namely, the combined cross-sectional, Black supplemental, and Hispanic supplemental samples). The other three sets should be designed for use with three "continuous" samples associated with the entire sample, the random sample, and the entire sample augmented by the economically disadvantaged White supplemental component. The construction of these latter sets of weights should be done in a way to match the sample characteristics of the continuously interviewed individuals (as of each interview) to the attributes of the "youth population." We expect that using such weights would reduce some of the differences that our analysis found between the continuous and full samples.

III. The Nature of Attrition in the NLSY

This section examines data from within the NLSY to assess how attrition might be influencing our understanding of youths' employment and earnings experiences. We determine how attriters differ in their labor-market activities from those who remain in the NLSY sample, summarizing their experience with variables on labor force participation (namely, employment), average hourly earnings, and annual earnings. Identifying whether nonrespondents come disproportionately from particular segments of labor-market distributions offers a basis for drawing inferences about how the absence of these attriters might alter the picture of employment and earnings conveyed by the youth remaining in the sample.

To achieve this objective, it is convenient to separate the analysis into two activities: first, developing a picture of those who miss interviews in the NLSY; and, second, constructing a characterization of those who return to the sample after missing interviews.

A. Who Attrits from the NLSY?

One popular approach to determining the representativeness of attriters involves characterizing how their activities prior to attrition differed from those of their contemporaries who remain in the sample. For this analysis, we focus on employment and earnings activities.[12] We formulate statistical models to identify the relationships linking future attrition to current labor-market experiences. We summarize these experiences using four variables: (1) employment rates; (2) school attendance; (3) annual earnings; and (4) wage rates or average hourly earnings. One model examines employment/nonemployment rates and school attendance; a second examines annual earnings and wages.

1. Empirical Model Describing the Employment-Schooling Status of Attriters

Characterizing the role of nonemployment (unemployment or otherwise not working) and school participation on attrition rates requires an empirical specification that measures the degree to which attriters differentially come from the nonemployed, in-school, and employed segments of the population. We estimate the following probit model for attrition,

$$(3.1)\quad P[\delta_{i(t+1)}|X_{it}] = \Phi(X_{it}\beta) \equiv \Phi(E_{it}[1 - S_{it}]D_{Ait}\ \beta_{AE} + S_{it}D_{Ait}\beta_{AS} + K_{it}\beta_K),$$

where the variables in the equation are defined as in Section 2:

$\delta_{i(t+1)}$ = indicator variable signaling individual i attrits in year $t + 1$;
X_{it} = a generic notation for the covariates incorporated in the analysis;
E_{it} = indicator variable signaling employment in year t (as defined in Section IIB);
D_{Ait} = vector of dummy variables for the various age brackets;
S_{it} = indicator variable denoting whether an individual is in school in year t;
K_{it} = vector containing full complement of age and year dummies (one for each year from 1979–91 and for each age from 14–33); and
Φ = cumulative normal distribution function.

The parameters β_{AE} and β_{AS} in (3.1) gauge the degree to which attriters come from the ranks of the employed and in-school populations relative to the groups that are nonemployed. The values of β_{AE} measure the differentials in the one-year-ahead attrition rate associated with those who are employed but not in school, with positive (negative) values indicating that prospective attriters are more (less) likely to come from this group. The values of β_{AS} measure the extent to which attriters were individuals still in school, with positive (negative) values signaling a higher (lower) likelihood of attrition for the in-school group. In the formulation of (3.1), we make no distinction between the employed and nonemployed among those who are in school—we classify all these individuals as in-school. The interaction terms introduced through inclusion of the vector D_{Ai} permit the relationships between labor-

12. We could, of course, consider a wide variety of other factors influencing and affected by attrition, such as marital status, children, prior geographical mobility, etc. Throughout this study, however, we concentrate on the role of labor-market variables in attrition. We do this in part to keep the analysis manageable, but also in recognition that these variables are among the most carefully measured data in the NLSY, and that information on these variables is a major reason why the U.S. Bureau of Labor Statistics sponsors the NLSY.

market statuses and attrition to vary freely across age brackets. We include the education interaction terms, incorporating the dummy variable S_{it}, for only the youngest two age groups, corresponding to the age ranges associated with attendance of high school (ages 14–17) and the first years of college (ages 18–20). We do not examine schooling effects after age 20 or employment effects before age 18.

We estimate the model by maximum likelihood separately for the following eight samples (defined in Section II): (i) the nationally representative sample (all); (ii) high-school dropouts (11-); (iii) high-school graduates (12); (iv) some-college group (13–15); (v) college graduates (16+); (vi) Whites; (vii) Blacks; and (viii) Hispanics. We estimate each specification using 1979 weights, with distinct specifications estimated for men and women.[13] The age bracket dummies $D_{A..}$ in these specifications segment observations into the five age ranges: 14–17, 18–20, 21–25, 26–30, and 31–34.

Tables 6a and 6b report the estimation results of the above probit model for men and women, respectively. The column headings indicate which of the eight samples serve as the source for estimates, and the rows list estimated effects for the age brackets: 14–17, 18–20, 21–25, and 26–30.[14] To help interpret the probit estimates, we use the fitted parameters of our model to predict differences in attrition rates based on labor-market status, holding age, year, and gender constant. As a reference point, we evaluate coefficients and variables at the average of the age effects for the age bracket under consideration and at the average of the time effects over the period 1979–91. This produces a representative attrition rate for the appropriate demographic group. With $\beta^{\dagger} = K_{it}\hat{\beta}_K$ being the fitted value from this evaluation, the difference in attrition rates for those who are in school, compared to being not employed, is

$$(3.2.a)\quad \Phi[\beta^{\dagger} + D_{A..}\hat{\beta}_{AS}] - \Phi[\beta^{\dagger}];$$

and the difference in these rates for those who are employed is

$$(3.2.b)\quad \Phi[\beta^{\dagger} + D_{A..}\hat{\beta}_{AE}] - \Phi[\beta^{\dagger}].$$

These are the numbers reported in Tables 6a and 6b. The top two rows present attrition rate differentials for those who are in school relative to those not employed based on (3.2.a), and the next three rows report differentials for those who are employed using (3.2.b). The analysis reports no estimates for the youngest bracket (14–17) for labor-market outcomes associated with employment due to the limited information conveyed by this status for members of this age range.

To measure the statistical significance of the differences in rates, we report the significance of the corresponding coefficients $\hat{\beta}_{AE}$ and $\hat{\beta}_{AS}$. The "*" superscript indicates that an estimate is statistically significantly different from zero at a 5 percent level of significance and a "**" superscript indicates a 1 percent level of significance. The bottom row in each group reports the conclusions from testing two

13. We use the 1979 weights in all of our estimations in this section for two reasons. First, the analysis in Section 2 demonstrated that use of these weights yields virtually identical summary statistics to those calculated using current weights. Second, in contrast to the analysis of Section 2, the notion of current weights is an ambiguous concept in the multi-period estimations conducted here; current weights for period *t* variables are formally inappropriate weights for period t-1 variables.

14. Estimates are not presented for the 31–34 age bracket because of small sample sizes.

Table 6a

Differences in Attrition Rates for Men Who Are in School or Employed for Various Education, Race-Ethnicity, and Age Groups[a]

		Education						
Status/Age Group	All	11−	12	13–15	16+	Whites	Blacks	Hispanic
In-school								
14–17	0.09	—	—	—	—	0.11	−0.01	0.10
18–20	−0.08*	—	—	—	—	−0.07	−0.10	−0.10
Employed								
18–20	−0.03	−0.07	−0.03	—	—	−0.02	−0.06	−0.03
21–25	−0.10**	−0.05	−0.11**	−0.09*	−0.09	−0.12**	−0.05	−0.08*
26–30	−0.12**	−0.12**	−0.04	−0.13**	−0.25**	−0.15**	−0.02	−0.07
Tests for no differences across ages:[b]								
18–30 = 0								
21–30 = 0	*RR*	*RR*	*RR*	−*R*	−*R*	*RR*	*AA*	*AR*

a. A "*" superscript on an estimate indicates that the asymptotic t-statistic associated with it exceeds 1.96 in value (namely, P value for this estimate > .05), and a "**" superscript indicates that the value of the t-statistic is greater than 2.57 (namely, P value > .01).

b. These rows present findings for tests of the null hypotheses that all age-group coefficients included in the ranges 18–30 and 21–30 respectively are jointly equal to zero. An A indicates acceptance of the null at 98 percent confidence, whereas an R indicates rejection of the null. Thus, RR in this row signifies rejection of both the coefficient restrictions 18–30 = 0 and 21–30 = 0. A dash means that the hypothesis test for 18–30 involves too few observations to be informative beyond the 21–30 test.

Table 6b
Differences in Attrition Rates for Women Who Are in School or Employed for Various Education, Race-Ethnicity, and Age Groups[a]

Status/Age Group	All	Education: 11−	12	13–15	16+	Whites	Blacks	Hispanic
In-school								
14–17	−0.00	—	—	—	—	−0.01	0.03	−0.01
18–20	−0.06	—	—	—	—	−0.07	−0.13*	0.06
Employed								
18–20	−0.12**	−0.04	−0.08	—	—	−0.14**	−0.11*	−0.07
21–25	−0.03	−0.03	−0.01	−0.04	−0.08	−0.02	−0.06*	−0.04
26–30	−0.08**	−0.03	−0.03	−0.09*	−0.11**	−0.09**	−0.03	−0.05
Tests for no differences across ages:[b]								
18–30 = 0								
21–30 = 0	*RR*	*AA*	*AA*	−*A*	−*R*	*RR*	*AA*	*AA*

a. A "*" superscript on an estimate indicates that the asymptotic t-statistic associated with it exceeds 1.96 in value (namely, P value for this estimate > .05), and a "**" superscript indicates that the value of the t-statistic is greater than 2.57 (namely, P value > .01).

b. These rows present findings for tests of the null hypotheses that all age-group coefficients included in the ranges 18–30 and 21–30 respectively are jointly equal to zero. An *A* indicates acceptance of the null at 98 percent confidence, whereas an *R* indicates rejection of the null. Thus, *RR* in this row signifies rejection of both the coefficient restrictions 18–30 = 0 and 21–30 = 0. A dash means that the hypothesis test for 18–30 involves too few observations to be informative beyond the 21–30 test.

hypotheses concerning the joint significance of coefficients associated with the effects of attrition: the first null hypothesis restricts the coefficients β_{AE} to be jointly equal to zero for age brackets 18–20, 21–25, and 26–30; and the second null constrains the coefficients β_{AE} to be jointly zero for just the two older brackets, 21–25 and 26–30. An A (R) in the first position of the lower row indicates acceptance (rejection) of the first null at 2 percent level of significance, whereas an A (R) in the second position signifies acceptance (rejection) of the second null.

Rather than discussing the findings in Tables 6a and 6b at this point, we will summarize what these results have to say about how attriters differ in terms of their work and schooling statuses after we introduce an empirical framework capable of characterizing how attritters also differ in other factors determining their labor earnings. This allows us to present an integrated picture of how the labor-market experiences of attriters compare to those of other sample members.

2. *Empirical Model Describing the Wages and Earnings of Attriters*

We also estimate an empirical model to assess how the wage and earnings distributions of prospective attriters compare to their nonattriter contemporaries. The conditional means and standard deviations produced by standard regression models offer only cursory information to infer the shapes and locations of distributions. For this reason, we estimate several quantile regressions.

As the basic specification for these quantile regressions, our equation for wages is:

$$\ln \omega_{it} = \delta_{i(t+1)} D_{Ait} \gamma_{A\delta} + K_{it} \gamma_K + \varepsilon_{it}, \tag{3.3}$$

where ω is average hourly earnings as defined in Section IIB, the γ's are coefficients, and ε_{it} is an error term. To analyze earnings, we substitute $\ln y_{it}$ for $\ln \omega_{it}$ in (3.3), where y is annual reported earnings as defined in Section IIB.

We estimate formulations of (3.3) corresponding to the 10th, 50th, and the 90th percentiles for both wages and earnings. Knowledge of how these three quantities differ between attriters and nonattriters characterizes where attriters fall in the wage and earnings distributions. The values of the coefficients $\gamma_{A\delta}$ on the age-bracket-interaction terms in (3.3) measure the amount by which the various quantiles for attriters deviate from the equivalent for nonattriters; a negative (positive) $\gamma_{A\delta}$ indicates a lower (higher) quantile value for the attrition component of the sample.

Tables 7a and 7b report the results from using standard quantile methods to estimate the above specifications separately for men and women, respectively. The first three sets of rows in Tables 7a and 7b report the coefficient estimates for the median, the 10th and the 90th percentiles of the hourly wage distribution for attriters by age bracket. The next three sets of rows list parallel results for the distribution of annual earnings. The coefficients $\gamma_{A\delta}$ presented in these rows can be interpreted as percentage differences in the 10th, 50th, and 90th percentiles of the wage and earnings distributions attributable to prospective attriters.

These tables do not report estimates for 10th and 90th percentiles when the number of attritions in a cell falls below 50 observations, nor do they present estimates of medians when sample sizes are below 20 observations per cell. The number of attri-

Table 7a
Differences in Annual and Hourly Earnings of Men Who Attrit in the Future for Various Education, Race-Ethnicity, and Age Groups[a]

Earnings Measure	Age Group	All	Education 11−	12	13–15	16+	Whites	Blacks	Hispanic
Median Hourly wages	18–20	0.05	0.19	−0.01	—	—	0.06	0.27*	−0.10
	21–25	−0.01	−0.05	0.00	0.05	0.18	0.01	0.04	−0.06
	26–30	−0.00	0.08	0.03	0.06	0.08	0.04	−0.01	−0.09
Test:[b] 18–30 = 0, 21–30 = 0		*AA*	*AA*	*AA*	−*A*	−*A*	*AA*	*AA*	*AA*
10th percentile hourly wages	18–20	−0.04	—	—	—	—	−0.03	0.23	−0.29
	21–25	−0.12	0.09	−0.15	0.17	—	−0.07	0.12	−0.21
	26–30	−0.03	−0.20	−0.02	0.01	—	−0.05	0.11	−0.40
Test:[b] 18–30 = 0, 21–30 = 0		*AA*	*A*−	−*A*	−*A*	—	*AA*	*AA*	*AA*
90th percentile hourly wages	18–20	0.14*	—	—	—	—	0.07	0.47**	−0.32**
	21–25	0.04	0.20*	−0.01	0.12	—	0.06	0.13	−0.10
	26–30	0.03	0.26*	0.01	0.12	—	0.07	−0.05	0.01

Test:[b] 18–30 = 0, 21–30 = 0		*AA*	−*R*	−*A*	−*A*	—	*AA*	*RA*	*AA*
Median annual earnings	18–20	0.13**	0.06	−0.15	—	—	0.12	0.03	0.12
	21–25	−0.02	−0.12	−0.05	0.00	0.06	−0.02	−0.03	−0.06
	26–30	−0.05	0.07	−0.02	−0.02	0.15	−0.03	−0.01	−0.00
Test:[b] 18–30 = 0, 21–30 = 0		*RA*	*AA*	*AA*	−*A*	−*A*	*AA*	*AA*	*AA*
10th percentile annual earnings	18–20	0.33*	—	—	—	—	0.47*	−0.83*	0.38
	21–25	0.03	−0.18	−0.30	0.44*	—	0.08	−0.00	−0.36
	26–30	0.01	0.16	0.13	0.30	—	0.11	0.39	−0.43
Test:[b] 18–30 = 0, 21–30 = 0		*AA*	−*A*	−*A*	−*A*	—	*AA*	*AA*	*AA*
90th percentile annual earnings	18–20	−0.00	—	—	—	—	−0.05	0.36*	−0.00
	21–25	0.03	0.06	−0.05	0.19	—	0.03	0.10	−0.03
	26–30	0.02	0.09	−0.10*	0.07	—	0.07	−0.07	−0.06
Test:[b] 18–30 = 0, 21–30 = 0		*AA*	−*A*	−*A*	−*A*	—	*AA*	*AA*	*AA*

a. A "*" superscript on an estimate indicates that the asymptotic t-statistic associated with it exceeds 1.96 in value (namely, P value for this estimate > .05), and a "**" superscript indicates that the value of the t-statistic is greater than 2.57 (namely, P value > .01).

b. These rows present findings for tests of the null hypotheses that all age-group coefficients included in the ranges 18–30 and 21–30 respectively are jointly equal to zero. An A indicates acceptance of the null at 98 percent confidence, whereas an R indicates rejection of the null; a dash indicates that the hypothesis test involves too few observations to be informative.

Table 7b
Differences in Annual and Hourly Earnings of Women Who Attrit in the Future for Various Education, Race-Ethnicity, and Age Groups[a]

Earnings Measure	Age Group	All	Education 11−	Education 12	Education 13–15	Education 16+	Whites	Blacks	Hispanic
Median Hourly wages	18–20	0.13*	−0.07	0.06	—	—	0.12	−0.10	−0.14
	21–25	0.02	0.23*	0.06	0.00	−0.05	−0.00	0.08	0.02
	26–30	−0.01	0.16	−0.07	0.03	0.02	−0.01	0.15	−0.12
Test:[b] 18–30 = 0, 21–30 = 0		*AA*	*AA*	*AA*	*AA*	*AA*	*AA*	*AA*	*AA*
10th percentile hourly wages	18–20	0.17	—	—	—	—	0.13	—	—
	21–25	−0.07	—	0.06	−0.14	—	−0.12	0.30	0.11
	26–30	0.20	—	−0.05	0.31	—	0.25	0.16	0.11
Test:[b] 18–30 = 0, 21–30 = 0		*AA*	—	−*A*	−*A*	—	*AA*	−*A*	−*A*
90th percentile hourly wages	18–20	0.12	—	—	—	—	0.11	—	—
	21–25	−0.02	—	−0.01	0.09	—	−0.08	0.17	0.30**
	26–30	−0.03	—	−0.09	0.00	—	−0.05	0.11	−0.05

Test:[b] 18–30 = 0, 21–30 = 0		*AA*	—	−*A*	−*A*	—	*AA*	−*A*	−*R*
Median annual earnings	18–20	0.10**	0.31	0.03	—	—	0.22*	0.25	−0.01
	21–25	0.08**	0.28	0.12*	0.03	−0.02	0.08	0.07	0.18*
	26–30	0.00	0.45	0.02	0.18**	0.13	−0.02	0.19	−0.01
Test:[b] 18–30 = 0, 21–30 = 0		*RR*	*AA*	*AA*	−*R*	−*A*	*AA*	*AA*	*AA*
10th percentile annual earnings	18–20	0.04	—	—	—	—	0.07	—	—
	21–25	−0.15	—	0.18	−0.52	—	−0.09	0.12	0.06
	26–30	0.19	—	0.18	−0.15	—	−0.10	0.30	0.43
Test:[b] 18–30 = 0, 21–30 = 0		*AA*	—	−*A*	−*A*	—	*AA*	−*A*	−*A*
90th percentile annual earnings	18–20	0.29**	—	—	—	—	0.31**	—	—
	21–25	0.11**	—	0.09*	0.16**	—	0.10*	0.23**	0.21**
	26–30	0.01	—	0.02	−0.01	—	0.01	0.13	0.04
Test:[b] 18–30 = 0, 21–30 = 0		*RR*	—	−*A*	−*A*	—	*RA*	−*R*	−*R*

a. A "*" superscript on an estimate indicates that the asymptotic t-statistic associated with it exceeds 1.96 in value (namely, P value for this estimate > .05), and a "**" superscript indicates that the value of the t-statistic is greater than 2.57 (namely, P value > .01).

b. These rows present findings for tests of the null hypotheses that all age-group coefficients included in the ranges 18–30 and 21–30 respectively are jointly equal to zero. An *A* indicates acceptance of the null at 98 percent confidence, whereas an *R* indicates rejection of the null; a dash indicates that the hypothesis test involves too few observations to be informative.

tions in the oldest group (31–34) are well below these thresholds, so no estimates are given for this group.

3. *Summary of Findings for Prospective Attrition*

Collectively, Tables 6a, 6b, 7a, and 7b contain only a few danger signals to the possible influence of attrition on measures of young men's and women's labor force participation, wages, and earnings in the NLSY. To draw the most extensive inferences possible from these tables, we vigorously extract any pattern of coefficients that can be given an interpretation. In doing so, we typically act as if statistically insignificant coefficients are zero. (If a coefficient estimate does not have a * superscript, then we read it as zero.) We are, of course, aware of the shortcomings of such a reading of estimates, but we want to avoid drawing inferences that do not have a sound statistical basis. We leave it to the reader to decide which particular patterns are convincing. We summarize what we have learned from these points at the end on this section.

a. Attrition Patterns for Men

The most striking feature of Table 6a comes in the first set of rows which strongly support the conclusion:

- Prospective attriters among men are more likely to come from the ranks of the nonemployed.

For the population at large listed in the first column (labeled ''All''), individuals aged 18–20 who are in school are 8 percentage points less likely to attrit in the subsequent year than their nonemployed counterparts, and employed persons are 3 percentage points less likely to attrit; the in-school difference is significant at the 5 percent level. Employed individuals aged 21–30 are about 10–12 percentage points less likely to attrit compared to the nonemployed; this difference is statistically significant.

Further, the results for ages below 18 indicate that:

- School attendance is unrelated to attrition during the period when young men are likely to be enrolled in high school.

All the in-school coefficients for the age bracket 14–17 are statistically insignificant at conventional levels of confidence. This implies that persons dropping out of high school are no more likely to attrit than those remaining in high school. After high school, we can also accept the hypothesis that the attrition rate from the in-school status is not significantly different than the rate from employment.

This pattern generally holds for breakdowns by education and race-ethnic groupings. For the college grads, the differential between the employed and the nonemployed after leaving school rises to 25 percentage points. Differentials tend to be smaller and not significant when considering Blacks and Hispanics.

Inspecting the results reported in the ''All'' column of Table 7a, which describes where prospective attriters come from in the wage and earnings distribution for the overall working population of young men, we see that:

- Among the working population of young men in their 20s, attrition occurs randomly across the wage and earnings distributions.

None of the coefficients associated with attrition in age groups 21–25 and 26–30 are either individually or jointly significant in the wages or earnings results. However, examination of the coefficients for the age group 18–20 indicates that:

- Among working teenage men (aged 18–20), attriters tend to come from the upper portions of the wage and earnings distributions.

The estimates for wages show that adding prospective attriters to the nonattrition population systematically fills in that part of the distribution above the median, leaving the 10th and the 50th percentiles unaffected and raising the 90th percentile by 14 percentage points. The estimates for earnings show that attriters tend to come from that part of the distribution concentrated just above the median—not from either tail of the distribution—leaving the 90th percentile unaffected and raising the 10th and the 50th percentiles by 33 and 13 percentage points respectively.

Considering findings in the second through the fifth columns describing sources of attrition in the various education groups, we see few differences across educational attainments. One distinctive pattern is the following:

- Among high-school dropout men, a disproportionate number of attriters have wages in the upper tail, even though attrition is random with respect to earnings.

For high-school dropouts, the estimates for attrition effects in the wage distribution for age brackets 21–25 and 26–30 show that prospective attriters raise the 90th percentiles by 20–26 percentage points, leaving the 10th and 50th percentiles unaltered to any statistically significant extent. In the 18–20 age group attriters appear to come from a wage distribution with a higher median, but none of the differences are significant at conventional levels. Similarly, none of the attrition-effect estimates for earnings are significant.

For the other education groups, we see little evidence of systematic attrition effects for either wages or earnings. Practically none of the estimates are statistically significant, but some of the point estimates are substantial. The number of prospective attriters among college graduates is so small that one cannot reliably draw inferences about the 10th or 90th percentiles.

The findings corresponding to different race-ethnic groups, presented in the last three columns, broadly support the conclusions for the nationally representative population. That is, prospective attriters above age 20 are drawn randomly from wage and earnings distributions. At the same time, these findings suggest differences across race-ethnic classifications for the youngest group where attrition effects are present.

Referring to the results for Whites, we see that:

- Attriters among teenage White men have wages similar to their nonattriting counterparts, but their earnings are less likely to be at the lowest values.

The only significant estimate for Whites corresponds to the coefficient linking the 10th percentile of the earnings distribution to the attrition effects; this coefficient indicates that prospective attriters have a 10th percentile that is 47 percentage points above the value associated with the nonattriting population.

Considering the results for Blacks reveals:

- Attriters among teenage Black men typically have higher wages than their nonattriting counterparts and have earnings placing them in either the lower or upper tails of the distribution.

Estimates of wage percentiles indicate uniformly higher wages for the attriters among Black men aged 18–20, with attriters' percentiles lying 23–47 percentage points above those corresponding to nonattriters. In the earnings distribution, the 10th percentile for attriters is 83 percentage points below the nonattriting population value, the median is virtually identical, and the 90th percentile is 36 percentage points above the value for the nonattriters. Thus, attriters tend to come from both tails of the earnings distribution, which implies greater earnings disparity among attriters than nonattriters.

Finally, inspection of the findings for Hispanics indicates:

- Attriters among teenage Hispanic men commonly have lower wages than their nonattriting contemporaries but show no difference in earnings.

The estimates of wage percentiles for Hispanic men indicate uniformly lower wages for the attriters, with attriters' percentiles lying 10–32 percentage points below those for nonattriting Hispanic teenage population. In contrast, the removal of attriters does not alter the earnings distribution in any statistically significant way.

b. *Attrition Patterns for Women*

As reported in Tables 6b and 7b, the influence of attrition on the wage and earnings distributions of young women mirrors patterns found for men in part, but there are some significant differences.

As is the case for men, the top rows of Table 6b indicate:

- Prospective attriters among women are more likely to come from the ranks of the nonemployed.

This finding generally holds for all ages, educational attainments, and race-ethnic classifications. The effects are somewhat less pronounced than for men after age 20. All estimates that are statistically significant imply that not working and not going to school at all in a calendar year is positively associated with the attrition rate. Employed women in the age range 18–30 have a likelihood of attriting in the next year that is 3–12 percentage points lower than their nonemployed counterparts. This range for the estimated employment effects on attrition roughly applies to the education groups with college experience, for both Whites and Blacks. For women who are high-school dropouts or graduates, employment appears to have no significant influence on attrition based on conventional hypothesis tests. This holds for Hispanic women as well.

The findings on the role of school attendance on attrition rates implies:

- School attendance is not linked to attrition for women in high school and college, with the exception of Black women participating in college (ages 18–20) who are less likely to attrit than their nonemployed contemporaries.

The basis for this finding comes from two sets of hypotheses tests: first, for all women's groups other than Black college participants, one can accept the hypotheses that each of the in-school coefficients equals zero; and, second, one can accept the hypotheses that each of the in-school coefficients equals the employment coefficients in their respective age brackets.[15] Combined, these test results imply that the attrition rates for those in-school are not significantly different from those of the nonemployed, and employed populations. Looking at the point estimates for the entire sample, the estimated in-school attrition rates lie between the estimated employed and the nonemployed rates. In the case of Black women attending college, the estimates and test statistics indicate that their attrition rate is similar to the rate for employed Black women in the same age range and is substantially below the rate for their nonemployed counterparts.

Table 7b shows the placement of prospective attriters in the wage and earnings distributions of the working population of young women. From this table, we see that:

- Attriters among working women have higher wages in their late teens, but their wages are not different after age 20. Further, these attriters have higher earnings in their late teens and early twenties, but their earnings are the same after age 25.

The wage estimates for ages 18–20 show that prospective attriters have percentiles uniformly above those of the nonattriting population, with 10th, 50th, and 90th deciles exceeding those of the nonattriters' by 12–17 percentage points. The wage estimates for the other age groups are not significant either individually or jointly. Estimates of earnings attrition effects for ages 18–25 reveal significantly higher values for the 50th and the 90th percentiles for attriters, with increases in the range of 8–10 percentage points for the medians, and as high as 29 points for the 90th percentiles. Estimates of the 10th percentile differentials are not statistically significantly different from zero for any age bracket, nor are any of the coefficients associated with attriters' differentials for ages 26–30.

Due to small sample sizes, the results describing sources of attrition in the various education groups (second through fifth columns) contain too sparse information to infer particular patterns across educational attainments. The results which are available support two conclusions, the first being:

- Attriters among high-school dropout women in their twenties typically have higher wages; attrition is otherwise random across wages for higher educational attainments.

Estimates of attrition effects in the wage distribution for high school dropouts aged 21–30 show that prospective attriters have medians exceeding those of nonattriters

15. These tests only apply to the age brackets 18–20 since this is the only bracket distinguishing schooling and employment as distinct statuses. Formal tests of the null hypothesis that the in-school coefficient equals the employment readily pass for each of the 18–20 brackets at conventional levels of significance.

by as much as 23 percentage points, which is a statistically significant differential. The wage estimate for age group 18–20 is not significant, nor are any of the wage estimates either large in magnitude or significant for the other education groups.

The second conclusion is:

- Among women high-school graduates and those with some college education, attriters tend to come from the upper half of the earnings distribution for some time after leaving school; otherwise, attrition is random for other ages and educational attainments.

The earnings estimates exhibiting significance are those associated with the 50th and 90th percentiles of the earnings distribution in the 21–25 age bracket for high-school graduates and the 21–30 bracket for those with less than four years college education. The gain in these percentiles for attriters falls in the range of 9–18 percentage points, indicating that attriters are drawn from the upper tail of the earnings distribution. None of the earnings attrition effects are statistically significant for high-school dropouts or college graduates, even though some of the estimated impacts on the median earnings for high school dropouts are large in magnitude.

Finally, the last three columns of Table 7b reveal:

- Attrition patterns for women within race-ethnic groups broadly mirrors those found in the overall women's population.

There are subtle differences, however. There is some evidence of a wage selection effect arising from Hispanic attriters aged 21–25 being drawn from the upper tail of their distribution, but only a single estimate is significant. The other differences are not noteworthy.

B. Who Returns to the NLSY?

Since many of the individuals who leave the NLSY sample eventually return, we can also explore how the labor-market experiences of these former attriters compare with the sample members who were not absent from the sample sometime in the past. As before, two questions must be answered to characterize these different experiences: (1) Do returnees have lower or higher employment rates than the nonattriters? (2) How do the wage and earnings distributions of the working returnees compare to those of the nonattriters? Straightforward modifications of the probit and the quantile specifications in the previous sections offer a framework for addressing these questions.

1. Empirical Model Describing the Employment Status of Returnees

To capture differences in the employment rates of nonattriters and returnees, the following probit model relates past attrition and the likelihood of employment in year t:

$$(3.4)\quad P[E_{it}|X_{it}] = \Phi(X_{it}\theta) \equiv \Phi(\delta_{it}^{*} D_{Ait}\theta_{A\delta} + K_{it}\theta_K),$$

where θ is a parameter vector, and K_{it} is once again a vector of controls incorporating a full complement of age and year dummies. (As the dummy variable δ_{it}^{*} signals whether individual i has ever been an attriter in a year prior to the current year t; the inclusion of this variable in the specification identifies the returnees in the sample interviewed in year t. The presence of the age bracket dummies D_{Ait} as interaction terms in (3.4) allows differences in the employment rates for returnees and nonattriters to vary with age.

Tables 8a and 8b summarize the results of estimating this probit model separately for men and women, respectively, using the same structure as Tables 6a and 6b. The model is estimated for the same eight sample compositions using 1979 weights, with the exception, of course, that nonrespondents are not included during years without interviews. In contrast to Tables 6a and 6b. Tables 8a and 8b include estimates for the oldest age bracket (31–34) in addition to the three brackets (18–20, 21–25, and 26–30) because the small sample size in the prospective attrition sample for this age bracket is not an issue in the analysis of the retrospective attrition sample considered here.

Tables 8a and 8b summarize the differences in the probability of employment between returnees and nonattriters. To translate the findings of the probit model into estimates of these differences, we adopt an approach similar to that above. In particular, for each age group we take the never-attrited population as the reference case by evaluating the distinct age effects, included among the covariates K, at the average for each bracket and the included time effects at the average over the period 1979–91. With $\theta^{\dagger} = K_{it}\hat{\theta}_K$ representing the fitted value evaluated in this manner, the differential in returnees' employment rates is

$$(3.6) \quad \Phi[\theta^{\dagger} + D_{A\cdot\cdot}\hat{\theta}_{A\delta}] - \Phi[\theta^{\dagger}].$$

These average differences in employment probabilities are the numbers reported in Tables 8a and 8b.

As a measure of the statistical significance of these differences, we report the significance of the coefficient $\hat{\theta}_{A\delta}$, assigning the "*" and "**" superscripts. The bottom row of each group reports the findings obtained for tests of two hypotheses relating to the joint significance of the attrition-effect coefficients: the first null restricts the coefficients $\theta_{A\delta}$ to be jointly zero for age brackets 18–20, 21–25, 26–30, and 31–34; and the second null hypothesis constrains these coefficients to be jointly zero for the three oldest brackets 21–25, 26–30, and 31–34. An A (R) in the first position indicates acceptance (rejection) of the first null at 2 percent level of significance, whereas an A (R) in the second position signifies acceptance (rejection) of the second null.

Analogous to the discussion in Section A, we wait to summarize the findings in Tables 8a and 8b concerning how returnees differ in their work status until after we present an empirical framework characterizing how returnees also differ in other determinants of earnings. We then present an overall review of how the labor-market experiences of returnees compare to other sample members.

2. *Empirical Model Describing Wage and Earnings of Returnees*

As before, we estimate quantile regressions for wages and earnings of returnees. The basic specification to summarize differences in wage and earnings distributions

Table 8a
Differences in Employment Rates for Men Who Return to Sample for Various Education, Race-Ethnicity, and Age Groups[a]

		Education						
Age Group	All	11−	12	13–15	16+	Whites	Blacks	Hispanic
18–20	—	−0.05	−0.22**	—	—	−0.18**	−0.04*	−0.20**
21–25	—	−0.08**	−0.12**	−0.02	−0.11*	−0.09**	−0.11**	−0.12**
26–30	—	−0.04*	−0.05**	−0.07**	−0.11**	−0.06**	−0.08**	−0.05**
31–34	—	−0.02	0.09**	0.01	−0.08	−0.05*	−0.09**	−0.03
Tests for no differences across ages:[b]								
18–34 = 0								
21–34 = 0	*RR*	*RR*	*RR*	−*R*	−*R*	*RR*	*RR*	*RR*

a. A "*" superscript on an estimate indicates that the asymptotic *t*-statistic associated with it exceeds 1.96 in value (namely, *P* value for this estimate > .05), and a "**" superscript indicates that the value of the *t*-statistic is greater than 2.57 (namely, *P* value > .01).

b. These rows present findings for tests of the null hypotheses that all age-group coefficients included in the ranges 18–30 and 21–30 respectively are jointly equal to zero. An "*A*" indicates acceptance of the null at 98 percent confidence, whereas an "*R*" indicates rejection of the null. Thus, *RR* in this row signifies rejection of both the coefficient restrictions 18–34 = 0 and 21–34 = 0. A dash means that the hypothesis test for 18–30 involves too few observations to be informative beyond the 21–30 test.

Table 8b

Differences in Employment Rates for Women Who Return to Sample for Various Education, Race-Ethnicity, and Age Groups[a]

Age Group	All	Education				Whites	Blacks	Hispanic
		11−	12	13–15	16+			
18–20	—	0.05	0.03	—	—	−0.08	0.01	−0.06
21–25	—	−0.08**	−0.10**	0.01	−0.13**	−0.12**	−0.11**	−0.15**
26–30	—	−0.11**	−0.10**	−0.06**	−0.12**	−0.13**	−0.13**	−0.11*
31–34	—	−0.11**	−0.12**	−0.09*	−0.01	−0.12**	−0.08*	−0.15**
Tests for no differences across ages:[b]								
18–34 = 0								
21–34 = 0	*RR*	*RR*	*RR*	*−R*	*−R*	*RR*	*RR*	*RR*

a. A "*" superscript on an estimate indicates that the asymptotic *t*-statistic associated with it exceeds 1.96 in value (namely, *P* value for this estimate > .05), and a "**" superscript indicates that the value of the *t*-statistic is greater than 2.57 (namely, *P* value > .01).

b. These rows present findings for tests of the null hypotheses that all age-group coefficients included in the ranges 18–30 and 21–30 respectively are jointly equal to zero. An "*A*" indicates acceptance of the null at 98 percent confidence, whereas an "*R*" indicates rejection of the null. Thus, *RR* in this row signifies rejection of both the coefficient restrictions 18–34 = 0 and 21–34 = 0. A dash means that the hypothesis test for 18–30 involves too few observations to be informative beyond the 21–30 test.

of returnees compared to their nonattriter counterparts relies on the following equation for wages:

$$(3.5)\quad \ln \omega_{it} = \delta^*_{it} D_{Ait} \pi_{A\delta} + K_{it}\pi_K + \varepsilon_{it},$$

where the π's are coefficients, and ε_{it} is an error term. To analyze earnings, we substitute $\ln y_{it}$ for $\ln \omega_{it}$ in (3.5). Again, we estimate formulations corresponding to the 10th, 50th, and the 90th percentiles to infer how distributions adjust to describe the experiences of returnees.

Tables 9a and 9b summarize the results of estimating these quantile regressions separately for men and women, respectively, with the structure of these tables matching Tables 7a and 7b. The first three sets of rows of Tables 9a and 9b present estimates of the $\pi_{A\delta}$, coefficients representing the percentage differences between the median, 10th and 90th percentiles of the hourly wage distributions faced by returnees verses their nonattriter contemporaries. The last three sets of rows characterize the analogous differentials for the distribution of annual earnings.

Tables 9a and 9b do not report medians when the number of retrospective attriters fails to exceed 20; they do not present estimated differentials for the 10th and 90th percentiles when the number falls below 50.[16] Like Tables 8a and 8b Tables 9a and 9b also present estimates for the oldest age bracket (31–34) in addition to the three age brackets 18–20, 21–25, and 26–30.

3. *Summary of Findings for the Returnee Population*

Tables 8a, 8b, 9a, and 9b present a relatively simple description of how returnees in the NLSY differ from nonattriter sample members. The tables reveal patterns that apply broadly for all education, race-ethnicity, and gender classifications. In interpreting findings, we once again follow the approach as applied in our discussion of Tables 6a, 6b, 7a, and 7b; we attempt to identify all patterns, often treating statistically insignificant coefficients as if they were zero.

a. *Labor-Market Experiences of Male Returnees*

The top rows of Table 8a strongly support the conclusion:

- Men reentering the NLSY—of all ages, educational attainments, and race-ethnicity—are less likely to work upon their return than are sample members who have never attrited.

The annual employment rate of returnees falls short of their respective population rates by as much as 20 percentage points, with many of these estimates statistically significant. The largest differentials occur for the youngest ages. After age 20, the estimated differences are of comparable size across educational attainments.

Examining results in the first column of Table 9a shows where retrospective attriters end up in the wage and earnings distributions of the overall population, we see:

16. In addition to the fact that the returnee samples for the age bracket 14–17 are typically small, the analysis reports no estimates for this youngest bracket due to the limited information conveyed by labor market outcomes for individuals in this age range.

- Among the working population of men above age 20, returnees receive wages and earnings locating them at the lowest values of the distributions.

The most substantial differentials for attriters occur at the 10th percentiles of wages and earnings, where attriters typically fall below the never-attriter population by 20–30 percentage points. Differentials also exist for attriters at the medians, but the shortfalls are typically below 10 percentage points in magnitude and are sometimes insignificant. Such findings support the conclusion that former attriters receive hourly wages and annual earnings placing them below the medians of their nonattriting counterparts and they are concentrated in the lower tail of the distributions.

The results for the various education groups in Table 9a offer one qualification to this conclusion:

- In contrast to the overall population of young men, returnees among high-school graduates and those with some college receive wages randomly dispersed in their respective distributions.

Although returnees are concentrated at the lowest wages for the high-school-dropouts and college-graduate groups—estimated differentials for the 10th percentiles reach as much as 40 percentage points for both groups—we see little evidence that the wages of returnees are different than their counterparts in the 12 and the 13–15 education groups. Despite these results for wages, it is still the case that the earnings of the returnees in these groups occupy the lower strata of their respective distributions.

The more noteworthy exceptions to the findings for the overall population show up in the results for the race-ethnic groups. For Blacks we see:

- Among the working population of Black men, former attriters receive the same wage and earnings as their nonattriter contemporaries.

Practically none of the estimates associated with returnees wages or earnings are either individually or jointly significant; the few significant results imply that returnees earn more (instead of less) than the never-attrit Black population.

In the case of Hispanics, we see:

- Among the working population of Hispanic men, former attriters receive uniformly lower wages and earnings than their nonattriter counterparts.

Not only are the 10th and the 50th percentiles of Hispanic attriters significantly below those of their population, so are their 90th percentiles. Thus, incorporating former attriters in the calculation of distributions does not merely shift mass from just above the median to the lower tails, it also shifts mass from the upper tail as well. This is consistent with former attriters having uniformly lower wages and earnings, meaning that even the highest earning returnees (namely, the highest 10 percent) are likely to be worse off than the highest-earning nonattriter.

b. Labor-Market Experiences of Female Returnees

The role of retrospective attrition on the wage and earnings distributions of young women shares the main features identified above for men. However,

Table 9a
Differences in Annual and Hourly Earnings of Men Who Return to NLSY for Various Education, Race-Ethnicity, and Age Groups[a]

Earnings Measure	Age Group	All	Education: 11−	Education: 12	Education: 13–15	Education: 16+	Whites	Blacks	Hispanic
Median Hourly wages	18–20	−0.05	−0.15	−0.12	—	—	−0.04	0.09	−0.23**
	21–25	−0.00	−0.13**	0.06*	0.06	−0.06	−0.01	0.01	−0.03
	26–30	−0.05**	−0.06	0.02	0.06	−0.07	−0.04*	0.03	−0.17**
	31–34	−0.13**	−0.16	0.10	−0.06	−0.08	−0.07	−0.13	−0.29**
Test:[b] 18–34 = 0, 21–34 = 0		*RR*	*RR*	*RR*	*−A*	*−A*	*AA*	*AA*	*RR*
10th percentile hourly wages	18–20	−0.08	−0.29	—	—	—	−0.15	—	−0.77**
	21–25	−0.21**	−0.31**	−0.08	0.17	−0.01	−0.18*	−0.20	−0.41**
	26–30	−0.16**	−0.46**	0.01	0.04	−0.40**	−0.16*	0.05	−0.56**
	31–34	−0.28**	−0.14	−0.29	−0.11	−0.34*	−0.33**	−0.02	−0.41
Test:[b] 18–34 = 0, 21–34 = 0		*RR*	*RR*	*−A*	*−A*	*−R*	*RR*	*−A*	*RR*
90th percentile hourly wages	18–20	0.10	0.21	—	—	—	0.15	—	0.08
	21–25	−0.01	0.08	−0.00	0.10	−0.17	−0.02	−0.00	−0.01
	26–30	0.00	0.05	0.13**	−0.02	−0.05	0.00	0.00	−0.12*
	31–34	0.08	−0.08	0.13	0.12	0.14	0.12	−0.00	0.08

Test:[b] 18–34 = 0, 21–34 = 0		*AA*	*AA*	−*R*	−*A*	−*A*	*AA*	−*A*	*AA*
Median annual earnings	18–20	0.20**	−0.02	−0.14	—	—	0.16	0.28*	−0.29**
	21–25	−0.02	−0.09*	0.02	−0.01	−0.14	−0.00	0.01	−0.07**
	26–30	−0.10**	−0.13**	0.02	0.03	−0.13*	−0.07*	−0.02	−0.22**
	31–34	−0.15**	−0.23**	0.02	−0.12	−0.19	−0.10	−0.10	−0.29**
Test:[b] 18–34 = 0, 21–34 = 0		*RR*	*RR*	*AA*	−*A*	−*A*	*RA*	*AA*	*RR*
10th percentile annual earnings	18–20	0.06	0.00	—	—	—	0.30	—	−0.37
	21–25	−0.30**	0.11	−0.47**	−0.08	−0.16	−0.18	−0.11	−0.42**
	26–30	−0.29**	−0.22	0.01	−0.06	−0.69**	−0.22**	−0.03	−0.87**
	31–34	−0.40**	−0.39	−0.68**	−0.20	−0.36	−0.41*	−0.18	−0.93**
Test:[b] 18–34 = 0, 21–34 = 0		*RR*	*AA*	−*R*	−*A*	−*R*	*RR*	−*A*	*RR*
90th percentile annual earnings	18–20	0.05	0.09	—	—	—	0.03	—	−0.16
	21–25	0.00	0.07*	0.01	−0.04	−0.04	0.01	0.12	−0.04
	26–30	−0.03	0.03	0.03	−0.03	−0.00	−0.03	0.12*	−0.16**
	31–34	−0.02	0.05	0.10	−0.09	−0.02	0.00	−0.00	−0.27*
Test:[b] 18–34 = 0, 21–34 = 0		*AA*	*AA*	−*A*	−*A*	−*A*	*AA*	−*A*	*RR*

a. A "*" superscript on an estimate indicates that the asymptotic t-statistic associated with it exceeds 1.96 in value (namely, P value for this estimate > .05), and a "**" superscript indicates that the value of the t-statistic is greater than 2.57 (namely, P value > .01).

b. These rows present findings for tests of the null hypotheses that all age-group coefficients included in the ranges 18–34 and 21–34 respectively are jointly equal to zero. An *A* indicates acceptance of the null at 98 percent confidence, whereas an *R* indicates rejection of the null; a dash indicates that the hypothesis test involves too few observations to be informative.

Table 9b
Differences in Annual and Hourly Earnings of Women Who Return to NLSY for Various Education, Race-Ethnicity, and Age Groups[a]

Earnings Measure	Age Group	All	Education 11−	12	13–15	16+	Whites	Blacks	Hispanic
Median Hourly wages	18–20	−0.06	−0.02	−0.10	—	—	−0.06	−0.13	−0.06
	21–25	−0.09**	−0.14*	−0.02	−0.04	−0.10	−0.10**	−0.03	−0.12*
	26–30	−0.05*	0.00	0.03	−0.04	−0.09*	−0.04	−0.08	−0.09*
	31–34	−0.01	0.11	−0.01	0.12	−0.08	−0.03	0.12	−0.06
Test:[b] 18–34 = 0, 21–34 = 0		*RR*	*AA*	*AA*	*−A*	*−R*	*RR*	*AA*	*RR*
10th percentile hourly wages	18–20	−0.02	—	—	—	—	−0.11	—	—
	21–25	−0.13	−0.03	0.05	−0.18	—	−0.17	0.04	−0.13
	26–30	−0.07	−0.13	0.08	−0.11	−0.08	−0.13	−0.08	0.08
	31–34	−0.29*	—	−0.78**	0.00	—	−0.39*	0.10	0.13
Test:[b] 18–34 = 0, 21–34 = 0		*AA*	*−A*	*−R*	*−A*	—	*AR*	*−A*	*−A*
90th percentile hourly wages	18–20	0.03	—	—	—	—	0.02	—	—
	21–25	−0.07*	0.10	−0.02	−0.01	—	−0.09*	0.17*	−0.04
	26–30	−0.04	0.15	0.08	−0.01	−0.10	−0.05	0.00	−0.10
	31–34	0.08	—	−0.06	0.16	—	0.13	0.11	−0.20*

Test:[b] 18–34 = 0, 21–34 = 0		*RR*	−*A*	−*A*	*AA*	—	*AA*	−*A*	−*A*
Median annual earnings	18–20	−0.07	0.05	−0.47**	—	—	−0.12	0.04	0.25
	21–25	−0.16**	0.10	−0.02	−0.14**	−0.12	−0.14**	−0.11	−0.14
	26–30	−0.08**	−0.17	0.06	−0.07*	−0.11	−0.07*	−0.04	−0.07
	31–34	−0.11**	−0.55*	−0.11	0.15*	−0.26**	−0.18*	0.04	−0.05
Test:[b] 18–33 = 0, 21–33 = 0		*RR*	*AA*	*RA*	−*R*	−*R*	*RR*	*AA*	*AA*
10th percentile annual earnings	18–20	−0.32	—	—	—	—	−0.09	—	—
	21–25	−0.33**	0.22	−0.28	−0.59	—	−0.46*	0.31	−0.03
	26–30	−0.02	−0.38	0.29	−0.34	−0.25	−0.10	0.00	0.15
	31–34	−0.99**	—	−0.91*	−0.30	—	−1.02**	−0.50	−0.29
Test:[b] 18–34 = 0, 21–34 = 0		*RR*	−*A*	−*R*	−*A*	—	*RR*	−*A*	−*A*
90th percentile annual earnings	18–20	−0.06	—	—	—	—	−0.06	—	—
	21–25	−0.07*	−0.17*	0.00	−0.15**	—	−0.08*	0.07	−0.16**
	26–30	−0.08**	−0.06	0.07*	−0.06	−0.08	−0.09**	−0.00	−0.17**
	31–34	−0.03	—	−0.13*	0.08	—	0.01	0.04	−0.09**
Test:[b] 18–34 = 0, 21–34 = 0		*RR*	−*A*	−*R*	−*R*	—	*AR*	−*A*	−*R*

a. A "*" superscript on an estimate indicates that the asymptotic t-statistic associated with it exceeds 1.96 in value (namely, P value for this estimate > .05), and a "**" superscript indicates that the value of the t-statistic is greater than 2.57 (namely, P value > .01).

b. These rows present findings for tests of the null hypotheses that all age-group coefficients included in the ranges 18–34 and 21–34 respectively are jointly equal to zero. An A indicates acceptance of the null at 98 percent confidence, whereas an R indicates rejection of the null; a dash indicates that the hypothesis test involves too few observations to be informative.

the results for women in Tables 8b and 9b show several points of departure. Three key conclusions are:

- Women returnees typically work less and earn less than women who have never attrited;
- Employment differentials for returnees are larger for women than for men, except in the teenage years when the reverse is true; and
- Returnee wage differentials are typically smaller for women than for men.

We consider the evidence supporting the last proposition first.

The estimates describing returnee-wage differentials in the ''All'' column of Table 9b for the nationally representative sample of women show that former attriters typically earn lower wages than nonattriter sample members. At least one of the differentials in either the 10th, 50th, or the 90th percentiles is statistically significant for all ages above 20. Comparison of these estimates with those in Table 9a indicates that larger losses occur in the lower tail of the wage distribution for male returnees than transpire for women. The reduction in the lowest decile for male returnees is about 10 percentage points greater than is the reduction faced by female returnees.

Comparing employment differentials for female and male attriters, the estimates in Tables 8a and 8b show that women returnees experience greater reductions in their annual employment rates than do men beyond age 20. However, results presented in the last three rows of Tables 9a and 9b indicate that female and male returnees experience similar earnings reductions. This finding for earnings, combined with the observation that women's wage losses are smaller than men's, implies that female returnees experience larger employment reductions than their male counterparts even in the working population. Thus, compared to men, former female attriters are less likely to work at all during the year than nonattriters, and when they do work, they work fewer hours at sightly lower wages.

The findings for various education and race-ethnic groups support some additional conclusions:

- The bulk of the wage losses for female returnees occurs for the lowest and highest levels of education, just as in the case of men.[17]

The median earnings estimates seem to infer that high-school-dropout returnees experience more substantial earnings losses than do former attriters in other education groups, but the lowest deciles show that returnees in these other groups face a substantially lower tail in their earnings which more than makes up for the rosier picture based on the median estimates.

When considering women, the special returnee patterns based on race-ethnic groups discovered above for men apply for Blacks but not for Hispanics. In particular, we see in Table 9b that:

- Among working Black women, former attriters receive wages and earnings similar to their nonattriter contemporaries.

17. The only exception is the large negative estimate for the 10th percentile of the high-school graduates in the oldest age group.

Only one coefficients associated with returnees' wages or earnings is statistically significant, and it has the opposite sign of the effects relevant for the population at large (namely, it suggests that 21–25 year old Black female returnees face a wage distribution with a higher upper tail).

On the other hand, in contrast to the experiences of Hispanic men,

- The wages of returnees among working Hispanic women are lower only in the middle of the distribution, and their earnings are lower only in the upper tail.

The estimates of the Hispanic returnee differentials for the 10th percentiles are never significant for either earnings or wages. The medians for returnees' wages are about 10 percentage points below those of all Hispanic women, and the 90th percentiles are rarely significant. The estimated differentials for the medians of earnings are never significant, but the 90th percentiles are lower by 9–17 percentage points. Thus, the shortfalls experienced by returnees in the female Hispanic population primarily occur near the medians of the distributions.

C. Further Empirical Analyses

In addition to the empirical work described above, we undertook a variety of other exercises designed to improve our understanding of the attrition process in the NLSY. Here we briefly summarize our findings for four such exercises.

We first reran most of the prospective attrition specifications, this time distinguishing who is a returnee and who is not. Before they leave the sample, attriters tend to have higher or no difference in earnings and wages, and upon their return they have distinctly lower values of compensation. These results appear conflicting. If returnees are different than those who leave the sample for good, however, then this could explain why prospective attriters as a group appear to come from the high earnings segments of the population and why those who reenter the sample return as low earners. Therefore, we included dummy variables in specifications (3.1) and (3.3) indicating who returns to the sample in some future year and who does not, in addition to incorporating the attrition-age interactions already in these specifications. We found little evidence to support a conjecture that returnees differ from permanent attriters in this manner. Although we occasionally found differences with the returnee dummy variables included in specifications (3.1) and (3.3), these dummy variables were generally insignificant determinants of labor market outcomes prior to attrition. When they were significant, the inclusion of these variables typically did not alter the conclusions listed above. Our overall impression from this analysis is that future returnees look much like their nonreturnee counterparts prior to attrition.

To explore prospective attrition further, we broadened the definition of who is classified as a prospective attriter by including anyone who leaves the sample within the subsequent two years as opposed to only the first year. This modification generally weakened all estimated effects, but it did not alter our previous conclusions. The evidence supports the idea that prospective attriters are most different just prior to leaving the sample, rather than several years before leaving.

We also considered another measure of earnings in carrying out both the prospective and retrospective analyses. Instead of reported income used above, we considered the imputed earnings variable described in Section II as the measure of earnings for assessing the consequences of attrition. In large part, this modification in the dependent variable did not alter our main conclusions concerning the make up of either the prospective-attriter or the returnee populations.

Finally, given the availability of the imputed income variable, we could determine what circumstances the members of the returnee population faced while they were absent from the sample, compared to those who remained in the sample. For those who attrited and returned, the NLSY permits construction of an imputed earnings variable during the periods of absence. In earnings specification (3.3) we used the imputed earnings variable as the dependent variable and replaced the returnee-dummy-variable δ with a dummy variable signaling when a person is absent in the observation year (namely, when this person was not interviewed in the year). The results of this analysis generally indicate that the earnings of returnees were higher or equal during the absent years. This finding is difficult to reconcile with the result that they have lower earnings after their return. A more likely explanation for this finding is that returned attritters tend to answer questions in the NLSY about earnings from employers in earlier years by reporting what their current earnings are from these employers or what these earnings were when they last worked for these employers.[18]

D. Summary: The Attrition Process in the NLSY

In developing a general characterization of attriters in the NLSY, we find substantial similarity in patterns between men and women, but important differences in the characteristics of prospective versus retrospective attriters.

Prospective attriters are more likely to be nonemployed—as opposed to being in-school or employed—prior to leaving the sample. Among working teenagers, these attriters typically have higher wages and earnings than their contemporaries, but after age 20 attrition occurs randomly across wage and earnings distributions. Among working high-school dropouts, prospective attriters tend to have higher wages than nonattriters throughout their twenties. Attrition is mostly random for the other educational attainments, with the possible exception being high-school-graduates and some-college-educated women in their twenties who are systematically selected from higher earnings levels. Prospective attrition patterns within race-ethnic groups broadly mirrors those found in the overall population, if we ignore the experiences of Black and Hispanic men in their teenage years. Attriters among teenage Black men typically have higher wages than their nonattriter counterparts and have earnings placing them in either the lower or upper tails of the distribution. Attriters among teenage Hispanic men commonly have lower wages but no difference in earnings.

Returnees or retrospective attriters to the NLSY—of all ages, educational attainments, and race-ethnicity—typically work less and earn less than sample members who have never attrited. Employment differentials for returnees are larger for women

18. In a comprehensive comparison of the various earnings measures, Cameron, Gritz, and MaCurdy (1989) find that the imputed earnings measure is typically higher and more disperse than the reported income variable in the NLSY. These authors speculate that respondents tend to answer questions about earnings averaged during the tenure of a job by citing what their current earnings are on the job at the time of the interview. Because wages grow rapidly for youth, this method of answering results in exaggerating the earnings actually received in the previous year.

than for men, except in the teenage years when the opposite is true. However, returnee wage differentials are typically smaller for women than for men. Wage losses for returnees occur for high-school dropouts and college grads, but not for high-school graduates or those with some college. Former attriters among Blacks have the same wages and earnings as their nonattriter contemporaries; whereas former attriters among Hispanics receive lower wages and earnings than their nonattriter counterparts.

IV. Comparisons of the NLSY and CPS

The analysis above suggests that attrition may have a small effect on the representativeness of the NLSY because attriters and returnees have somewhat different employment and earnings experiences compared to nonattriters. This section examines the representativeness of the NLSY for each year of the survey. A popular procedure for evaluating the point-in-time representativeness of a data set such as the NLSY is to compare its contents to another source deemed to provide a reliable picture of labor market experiences. No one knows whether such a data source exists for young adults, but the CPS regularly serves as a reference case.

The subsequent analysis compares an assortment of labor-market measures in the NLSY and CPS under a number of sample definitions. The variables of primary interest are those measuring the labor supply and earnings of individuals during a year, which include annual earnings, wages (average hourly earnings), annual labor force participation rates, and annual hours of work. We consider a wide variety of measures designed to assess the extent to which labor-market variables differ in the NLSY and the CPS. These measures include indications of both the location and dispersion of wage and earnings distributions, as well as how these distributions evolve over time. We carry out this empirical analysis controlling fully for age and gender, and we examine the degree to which variables match across data sets when selecting on educational attainment as well. Appendix A presents details of the data constructions and the estimation procedures applied in this section.

A. Differences Between the NLSY and CPS Across Cohorts, Years, and Ages

To begin our evaluation, we create graphs comparing wage, earnings and employment distributions of the cohorts common in the NLSY and CPS. After summarizing these findings, we formulate an empirical model that identifies how the NLSY and the CPS distributions compare across years and ages in an attempt to detect patterns in differences. The next major subsection (Section B) addresses how the NLSY and CPS compare when viewing variables organized by educational attainment.

1. Figures Describing Differences Across Cohorts

Figures 1a, 1b, 2a, 2b, 3a, 3b, 4a, and 4b summarize how a number of wage and earnings percentiles evolve over time according to the NLSY and CPS. Each figure plots the evolution of five percentiles: 10th, 25th, 50th, 75th, and 90th. (These are the statistics $P10$, $P25$, $P50$, $P75$, and $P90$ in the notation of Section II.) In both the

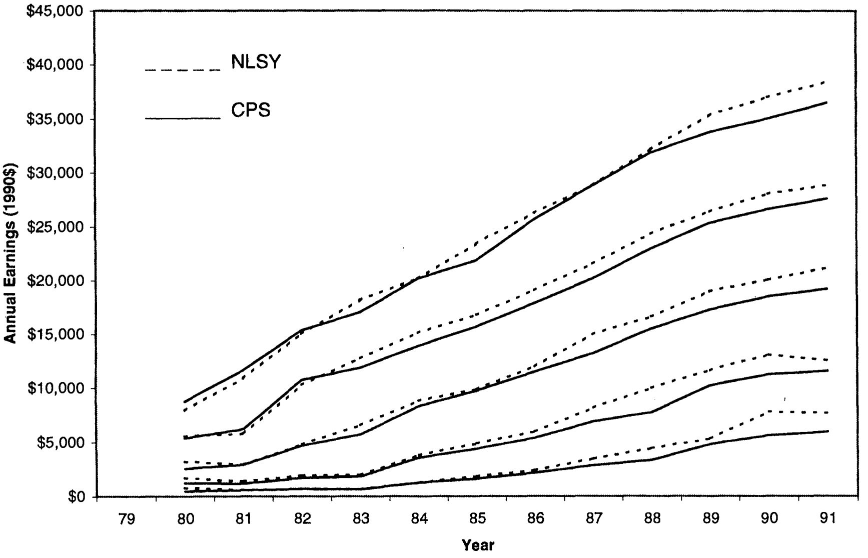

Figure 1a
Total Earnings—Cohorts 14 to 17, Men

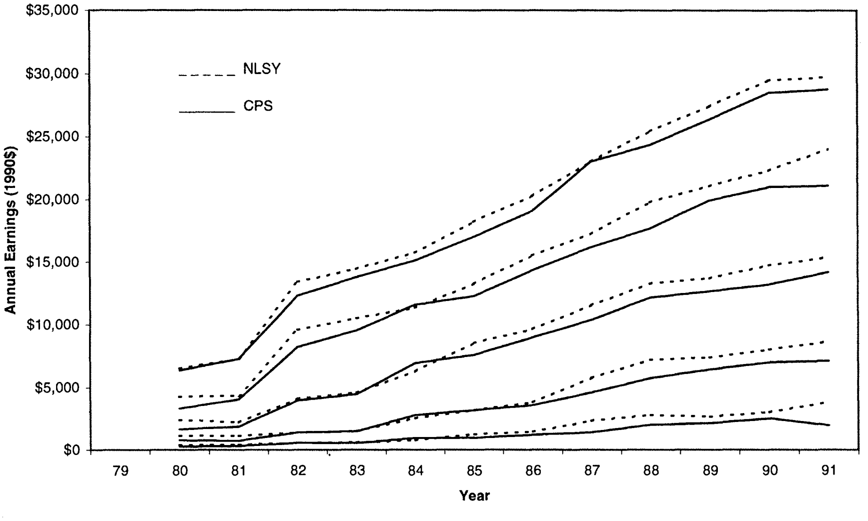

Figure 1b
Total Earnings—Cohorts 14 to 17, Women

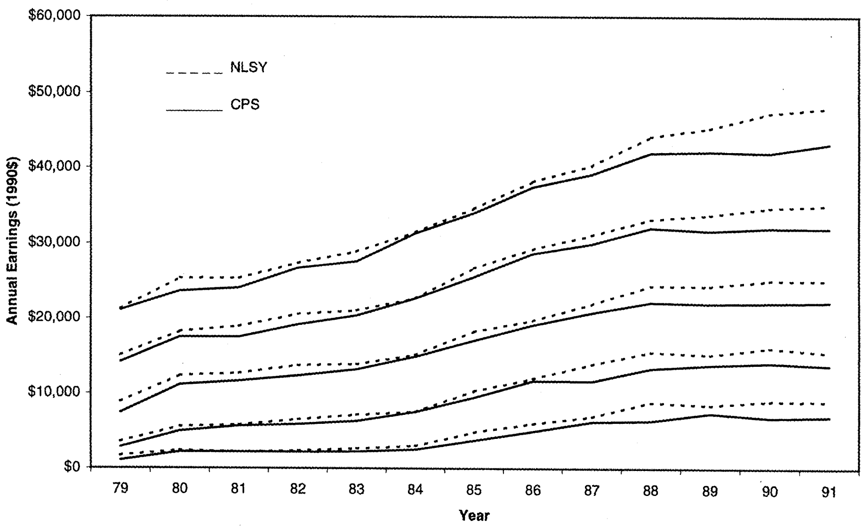

Figure 2a
Total Earnings—Cohorts 18 to 21, Men

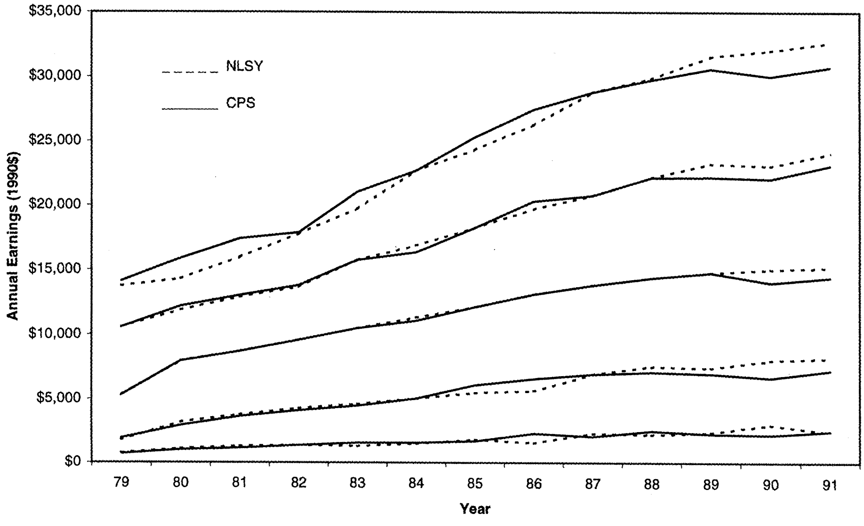

Figure 2b
Total Earnings—Cohorts 18 to 21, Women

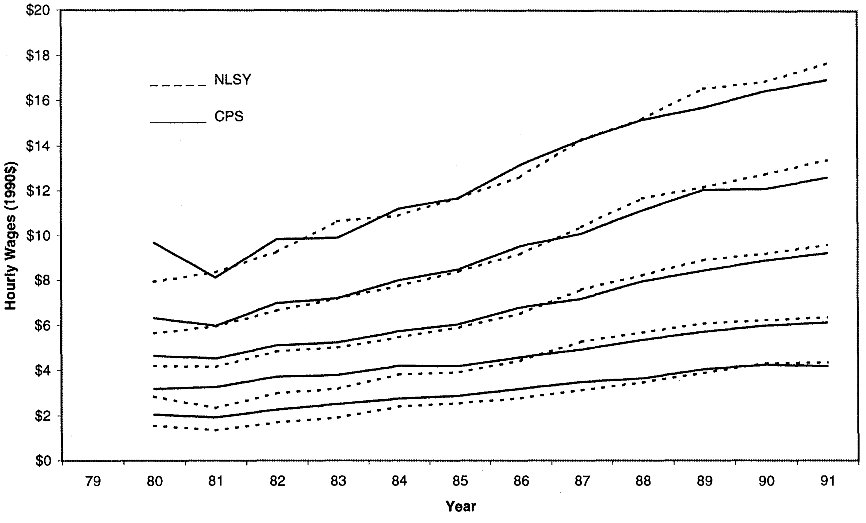

Figure 3a
Average Hourly Earnings—Cohorts 14 to 17, Men

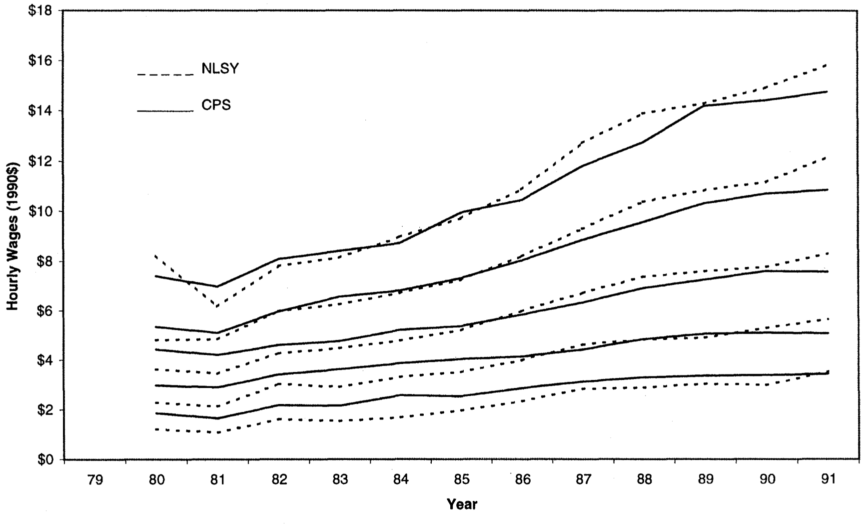

Figure 3b
Average Hourly Earnings—Cohorts 14 to 17, Women

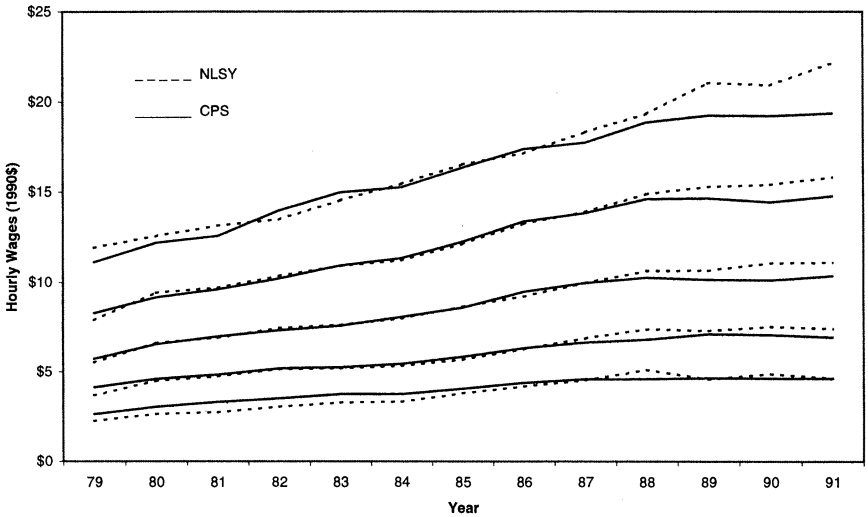

Figure 4a
Average Hourly Earnings—Cohort 18 to 21, Men

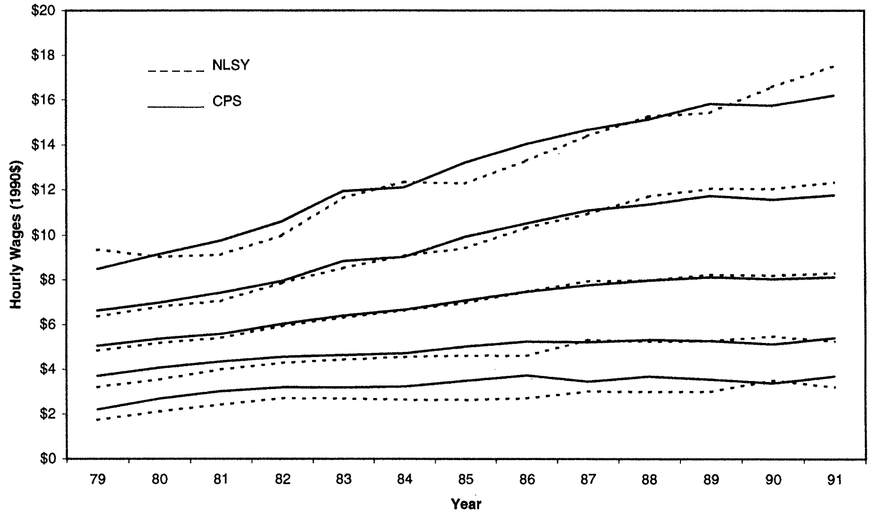

Figure 4b
Average Hourly Earnings—Cohorts 18 to 21, Women

NLSY and the CPS, the calculation of percentiles uses weighting procedures with samples restricted to only those individuals who are interviewed and who report positive hours and earnings during the relevant year. The dashed lines in the figures represent the profiles for the NLSY, and the solid lines designate profiles for the CPS. One should be careful to note that the *y*-axis scales are different across figures, so vertical distances represent different magnitudes that are not readily comparable across figures.

The "a" figures (for example, 1a) refer to men, and the "b" figures (for example 1b) refer to results for women. Figures 1 and 2 plot results for real annual earnings over the years covered by the NLSY, and Figures 3 and 4 graph quantities describing real wages (namely, average hourly earnings), both measured in 1990 dollars. Figures 1 and 3 plot percentiles for the cohorts aged 14–17 in 1979; and, Figures 2 and 4 graph percentiles for cohorts aged 18–21 in 1979. Both sets of figures include only those individuals who had positive hours and earnings during the year.

To compare labor force participation rates and hours of work implied by the NLSY and the CPS, Figures 5 and 6 graph the two measures of employment rates for cohort groupings 14–17 and 18–21. Let $ER_{At}(h > H)$ be the weighted proportion of persons in cohort A in year t who work more than H hours. The first employment measure is ER_{At} $(h > 0)$ (namely, the proportion of individuals working more than zero hours during the year); and the second measure is ER_{At} $(h > 1{,}000)$ (namely, the fraction of persons working more than 1,000 hours during the year). We refer to the condition $h > 1{,}000$ as "beyond-part-time employment." For us to consider an individual to have positive hours of work, we require him or her to have positive earnings as well. Figures 5 and 6 plot the weighted proportion of persons in a cohort group in year t who meet the relevant work conditions with separate results for men and women. The higher curves in the graphs represent the less stringent and more familiar employment rate ER $(h > 0)$, and the lower curves plot the values of the beyond-part-time employment rate ER $(h > 1{,}000)$.

2. *Summary of Differences Across Cohorts*

The graphs of the percentiles of annual earnings (Figures 1 and 2) show that:

- The earnings distributions of the NLSY and CPS are close to one another until about 1985 for all cohorts for both men and women. After 1985, the NLSY generally predicts higher earnings for all groups, with differentials widening after 1987.

All percentiles tell the same story. The NLSY percentiles typically lie above those of the CPS. The differentials are small before 1985–87. All differentials grow steadily after this period. implying that the entire distribution of the NLSY shifts upwards relative to that of the CPS. This domination appears smaller for women and for the younger cohorts. These graphs suggest that there are few differences in the trends in earnings across the two samples.

The wage-percentile graphs (Figures 3 and 4) reveal that:

- There is a crossover in the NLSY and CPS wage distributions around 1985 for both men and women. Prior to 1985, the CPS distributions predict higher

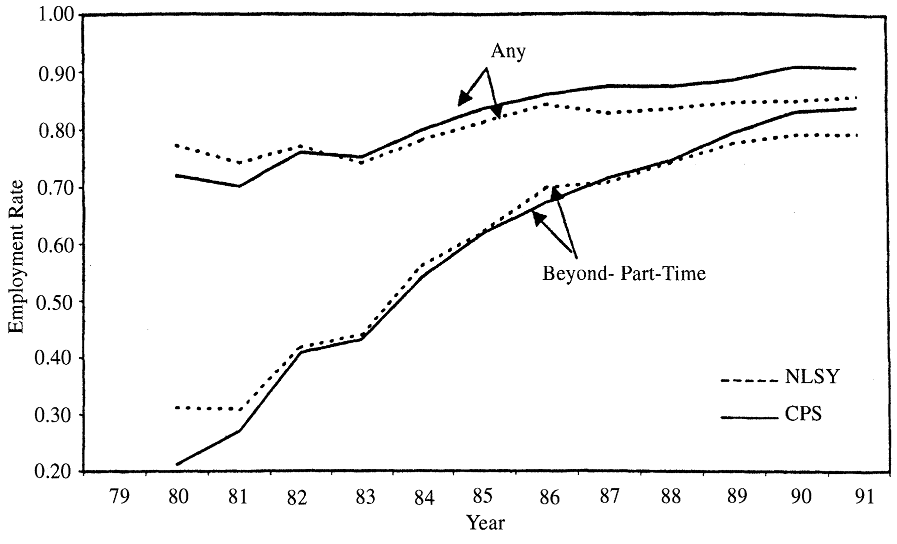

Figure 5a
Employment Rate—Cohorts 14 to 17, Men

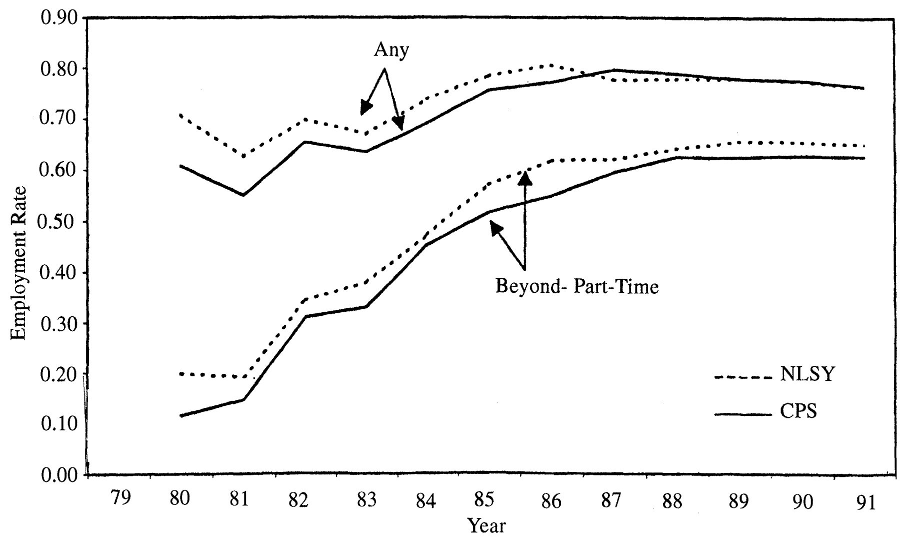

Figure 5b
Employment Rate—Cohorts 14 to 18, Women

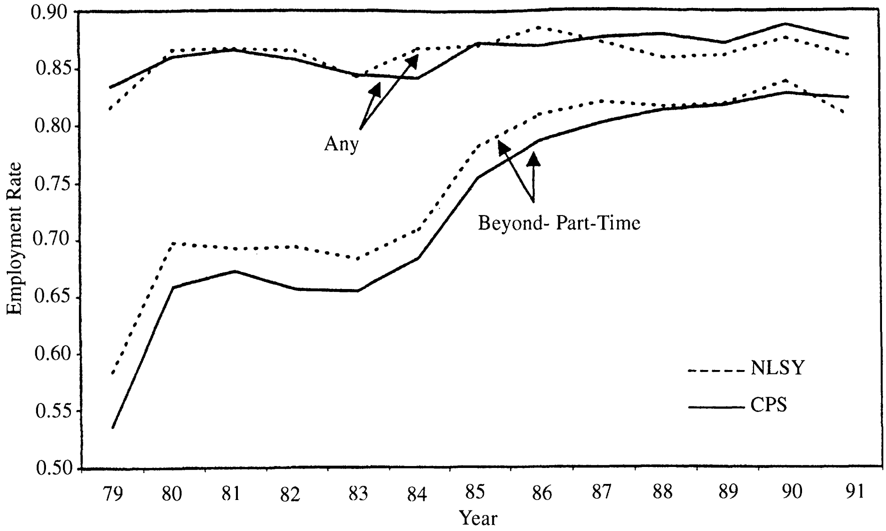

Figure 6a
Employment Rate—Cohorts 18 to 21, Men

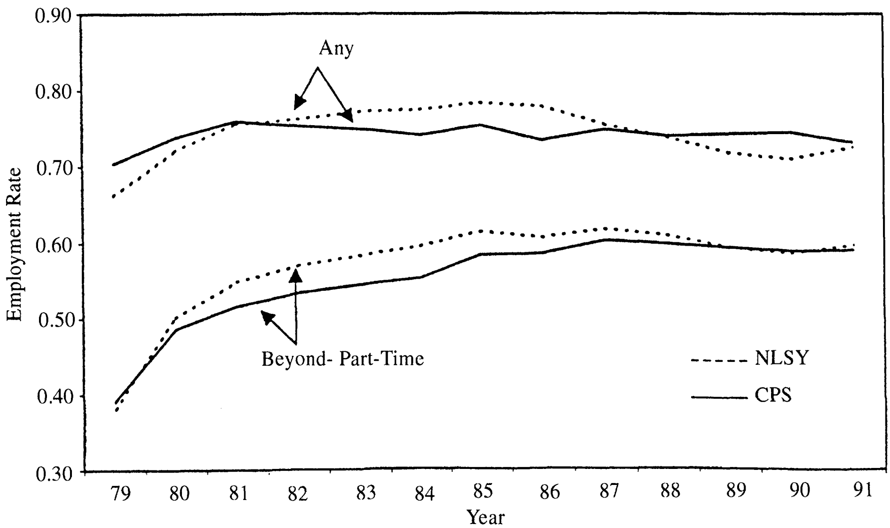

Figure 6b
Employment Rate—Cohorts 18 to 21, Women

wages, especially at the lower percentiles for the younger cohorts. After 1986, the NLSY predicts higher wages, especially at the upper percentiles.

For almost all of the earliest years, the NLSY percentile points lie below those of the CPS at all but the highest decile. Around 1985, the two distributions appear to coincide. Following 1986, however, the NLSY wage distributions dominate those of the CPS except at the percentiles below the median which are below or match the CPS. The largest difference between the two samples in the later years is for the oldest male cohort. Although the bulk of the trends appear fairly similar across the two samples, the NLSY implies slightly greater wage growth for all groups.

Considering the implications of Figures 1–4 together, the findings suggest that at young ages (or during the early years) the NLSY measures more hours of work than the CPS for those workers not in the higher tail of the earnings distribution. During the later years and at older ages, the NLSY appears to be measuring more earnings per worker, which is reflected in the higher wage and earnings distributions.

Figures 5 and 6 give a clear picture of the relative rankings of the NLSY versus the CPS in measuring the beyond-part-time rate (working more than 1,000 hours per year), but the story is somewhat mixed for the overall employment rate (indicating any work during the year):

- The NLSY signals higher beyond part-time rates than the CPS, with discrepancies tending to narrow in the latest years of the sample.

This finding is most pronounced for women, at least until the latest years of the sample. On the other hand, this result does not hold for the youngest cohorts of men where rates implied by the NLSY and CPS track one another closely throughout the period. No clear patterns exist for the overall employment rates for either women or men. Overall, the trends in the various employment rates conveyed by the two data sources match closely for all cohorts and genders. What differences arise in employment rates come in the form of slight shifts in the profiles.

3. Estimation of Distributional Differences Across Years and Ages

Although these figures provide an enlightening picture of the differences between the distributions of labor-market variables implied by the NLSY and CPS, they offer no guidance concerning which differences are statistically significant. Nor do these figures permit an accurate assessment of whether differences exist in the dispersions of distributions. Exploring these issues requires a statistical framework.

We use a simple regression analysis to summarize the extent that earnings, wage, and hours data from the NLSY deviate from those in the CPS. Define μ^n_{at} as the statistic μ_{at} computed for age cell α in year t using data from the NLSY, and μ^c_{at} as the corresponding statistic μ_{at} computed using data from the CPS. Our calculations of μ^n_{at} and μ^c_{at} use the NLSY and CPS weights to account for stratified sampling. Denote $Z_{at} = \mu^n_{at} - \mu^c_{at}$ as the difference between these statistics. Consider the regression model:

(4.1) $Z_{at} = d_t + D_{a\cdot\cdot} + v_{at},$

where d_t and $D_{\alpha..}$ are year and age dummies defined in the previous section, and v_{at} is an error term. The ranges of values for the years t and the ages α used as observations depend upon the sample composition under investigation. The number of observations for each regression consists of 88 cell differences (for each gender group) with estimated year effects for 1979 to 1991 and age effects from age 18 to age 30. As described in Appendix A, we apply weighted least squares adjusting for the sample sizes of each cell to limit the influence of cells with a small number of observations in either the NLSY or the CPS. In estimation, we always restrict the sum of the age effects over the relevant age range for each regression to equal zero. The estimated age effects therefore convey information about the life-cycle profile differences in the two data sets after removal of an overall average effect. In all instances, we estimate separate year and age effects, and all of the regressions are estimated separately for men and women.

In our analysis of the cohort differences across cells in the NLSY and the CPS, we examine six separate statistics μ_{at} for each gender group. First, we examine the difference between logarithms of the medians of annual earnings (namely, $\mu_{at} = \ln P50_{at}(y_{it})$ using the notation of Section II). Second, we examine the difference between the logarithms of the cell medians of wages or average hourly earnings (namely, $\mu_{at} = \ln P50_{at}(\omega_{it})$).

The third and fourth comparisons examine differences in earnings and wage inequality across the two data sets by forming measures of the logarithms of the ratio of the 90th percentile to the 10th percentile of the earnings and wage distributions for each cell. In the case of earnings, we calculate $\mu_{at} = \ln P90_{at}(y_{it}) - \ln P10_{at}(y_{it})$; in the case of wages, $\mu_{at} = \ln P90_{at}(\omega_{it}) - \ln P10_{at}(\omega_{it})$.

Finally, the fifth and sixth comparisons look at the proportion of individuals working more than zero hours and more than 1,000 hours during the previous year. We consider the statistics $\mu_{at} = ER_{at}\ (h > H)$, for $H = 0$ and $H = 1{,}000$, where $ER_{at}(h > H)$ designates the weighted proportion of persons of age α in year t who work more than H hours.

4. *Summary of Distributional Differences Across Years and Ages*

Tables 10a and 10b present estimates of the coefficients of Equation 4.1 for the six statistics μ_{at} defined in Section 3 for men and women. Listed in order by columns from left to right, the six statistics include: the differences between logarithms of the medians of annual earnings and wages; the differences in earnings and wage inequality (namely, differences in the logarithms of the ratio of the 90th percentile to the 10th percentiles); and the differences in the employment rates ($ER(h > 0)$) and the beyond-part-time employment rates ($ER(h > 1{,}000)$). For each labor-market measure, we report estimates for two specifications of the regression model (4.1): the left portion of each column shows the coefficient estimates obtained with only year effects incorporated in (4.1); and the right portion presents estimates with the inclusion of both year and age effects in the model. The "*" and the "**" superscripts attached to estimates indicate the ranges associated with their asymptotic t-statistics. The last row of each table reports whether one can reject the hypothesis that the year effects are zero at a 95 percent level of significance.

The first columns in Table 10 report findings for median annual earnings. The results indicate:

- The NLSY predicts higher median earnings for both men and women, with deviations of ten percentage points over the CPS not uncommon; the differentials tend to grow over time.

All of the year effects for men are positive and statistically significant after 1984. This situation does not change after inclusion of the controls for age. These controls in the women's specification render all the year effects significant and positive after 1986. Generally, controlling for age does little to diminish the year trends for differences in earnings for men, and it enhances the effects for women. No age pattern is apparent for men. For women, there is some evidence that the NLSY predicts higher earnings for ages below 25 and lower after 25. However, this finding may reflect a statistical artifact arising from an underlying interaction between cohorts and the time trend that is present in expanding the NLSY differential.

For average hourly earnings, the second columns of Tables 10a and 10b generally indicate:

- The NLSY predicts lower wages before 1986 with larger discrepancies for women, and it matches the CPS thereafter; differentials rarely exceed five percentage points.

Inspection of the estimates for men reveals that all of the year coefficients are positive and statistically significant after 1985, but this statistical significance vanishes after the inclusion of age controls. We are left with a few negative and significant coefficients in the mid-1980s, suggesting no substantive trend in the differential between the NLSY and the CPS wages for men. In the case of women, the inclusion of age controls produces negative and significant coefficients before 1986; the NLSY predicts lower wages, and the difference diminishes from 1979 until it vanishes around the mid-1980s. Examining the age effects in the lower half of the table reveals underestimation by the NLSY at the youngest age, but otherwise small age effects.

The third columns of Tables 10a and 10b present estimates obtained when the differences in the 90–10 dispersion measure of earnings serves as the dependent variable, and the fourth columns list estimates for the differences in the 90–10 dispersion measure of wages.

Turning to the differences in earnings dispersion, one striking feature emerges:

- The NLSY predicts dramatically less dispersion in earnings relative to the CPS in the last half of the survey years, with some differentials exceeding 25 percentage points for both men and women.

With the inclusion of age controls, almost all of the year effects are negative and most are significant after 1984 for men and after 1986 for women. The differentials are exceptionally large at the end of the sample period. Turning to the age effects in the bottom panels, younger ages appear to have relatively less dispersion in the NLSY and the older ages appear to have more dispersion. This pattern may indicate a complicated underlying process at work for women, given the size and significance of the estimates. However, it is difficult to infer much from this pattern since the

Table 10a
Differences in Earnings and Employment Statistics Between the NLSY and CPS for Men[a]

Coefficients	Median Earnings		Median Wages		90–10 Dispersion of Earnings		90–10 Dispersion of Wages		Employment Rates		Beyond Part-Time Work Rates	
Year effects												
79	.12	.11*	−.06	−.02	−.34**	−.27**	.06	−.06	−.02	.00	.06	.06**
80	.08	.06	−.02	.01	−.15	−.10	.12**	.03	.02	.04**	.06**	.05**
81	.05	.04	−.02	−.01	.01	.04	.18**	.11*	.01	.03*	.02	.02
82	.12**	.10*	−.01	.00	−.07	−.04	.10	.05	.01	.03**	.03**	.03**
83	.05**	.05	−.03	−.04*	−.08*	−.06	.10**	.07**	−.01	.01	.03**	.04**
84	.03	.04	−.02	−.04*	−.09	−.08	.06	.05	.01	.02**	.02	.03**
85	.06*	.06**	−.01	−.03**	−.12	−.12*	.04	.06	−.01	.00	.02	.03
86	.05**	.06**	−.03*	−.05**	−.12**	−.12**	.05	.08**	.00	.00	.02**	.03**
87	.11**	.12**	.05**	.03*	−.12	−.13*	.02	.07*	−.03*	−.03**	.00	.01
88	.11**	.12**	.03*	.01	−.27**	−.30**	−.04	.03	−.03**	−.04**	.00	.00
89	.09**	.10**	.04**	.03	−.10	−.14	.02	.09**	−.03*	−.04**	−.01	−.01
90	.09**	.11**	.04*	.03	−.17**	−.23**	−.03	.06	−.05**	−.07**	−.03**	−.04**
91	.12**	.13**	.06*	.04	−.07	−.15**	.05	.14**	−.04**	−.07**	−.04**	−.04**

Age effects												
18	—	.11*	—	−.14**	—	−.20**	—	.12*	—	.03**	—	.04*
19	—	−.03	—	−.04	—	−.15*	—	.20**	—	−.02	—	−.01
20	—	−.04	—	.02	—	.05	—	.14**	—	−.05**	—	−.02
21	—	−.02	—	.01	—	.00	—	.05	—	−.04**	—	−.02
22	—	.04	—	.01	—	−.01	—	.01	—	−.02*	—	.01
23	—	.01	—	.03**	—	.01	—	.03	—	−.02**	—	−.02
24	—	.06*	—	.04**	—	−.05	—	−.05**	—	−.01	—	.00
25	—	−.05**	—	−.00	—	−.01	—	−.06*	—	.00	—	.00
26	—	−.04	—	.00	—	.00	—	−.09**	—	.01	—	−.01
27	—	−.03	—	.02	—	.04	—	−.09**	—	.02**	—	.00
28	—	−.01	—	.03	—	.04	—	−.06	—	.03**	—	.00
29	—	−.01	—	.02	—	.11	—	−.12**	—	.03**	—	.01
30	—	.02	—	−.00	—	.16**	—	−.07**	—	.04**	—	.02
Trend?[b]		Yes		Yes		Yes		No		Yes		Yes

a. A ''*'' superscript on an estimate indicates that the asymptotic t-statistic associated with it exceeds 1.96 in value (namely, P value for this estimate > .05), and ''**'' superscript indicates that the value of the t-statistic is greater than 2.57 (namely, P value > .01).

b. Row indicates whether one accepts the hypothesis of equal year effects based on a Wald statistic assuming 95 percent level of confidence. A ''yes'' signifies rejection of the null, which indicates the existence of a trend. A ''no'' signifies the opposite.

Table 10b
Differences in Earnings and Employment Statistics Between the NLSY and CPS for Women[a]

Coefficients	Median Earnings		Median Wages		90–10 Dispersion of Earnings		90–10 Dispersion of Wages		Employment Rates		Beyond Part-Time Work Rates	
Year effects												
79	.03	−.03	−.06	−.04	−.02	.06	.16**	.07	−.04**	−.05**	−.01	−.02*
80	.04	−.03	−.08**	−.08**	−.05	.05	.18**	.12**	.01	.01	.03	.02
81	.01	−.03	−.07**	−.07**	−.23**	−.10	.10**	.07*	.02	.02**	.04**	.03**
82	.03	−.01	−.03	−.04*	−.05	.07	.12**	.10*	.02	.03*	.04**	.03**
83	.05**	.02	−.04	−.06**	.06	.16**	.12**	.13**	.03**	.03**	.04**	.03**
84	−.01	−.02	−.02	−.05**	.05	.14	.20**	.23**	.04**	.05**	.04**	.03**
85	.03	.03	−.03**	−.05**	−.07	−.01	.14**	.18**	.03**	.04**	.04**	.04**
86	.02	.03	−.01	−.03	.06	.07	.18**	.22**	.04**	.05**	.04**	.05**
87	.08*	.11**	.04*	.02	−.23**	−.25**	.07*	.13**	−.01	.00	.02**	.02**
88	.02	.06*	.02	.02	−.10	−.16*	.16**	.22**	.00	.00	.01**	.02**
89	.03	.08**	.02	.02	−.11	−.21*	.06	.12**	−.01	−.01	.01	.02**
90	.08**	.15**	.01	.01	−.18*	−.32**	.05	.12**	−.01	−.01	.02*	.03**
91	.09**	.16**	.08**	.09**	−.34**	−.50**	.00	.07*	.00	.00	.02**	.04**

Age effects												
18	—	.11	—	−.13**	—	−.13	—	.11**	—	.04	—	.00
19	—	.10**	—	.01	—	.01	—	.29**	—	.04	—	.03
20	—	.06	—	.02	—	−.15*	—	.03	—	.00	—	.02*
21	—	.00	—	.02	—	−.02	—	.06	—	−.02*	—	−.01
22	—	.08*	—	.04**	—	−.13	—	.00	—	−.02**	—	.01*
23	—	.00	—	.03	—	−.19**	—	−.03	—	−.02*	—	.00
24	—	.04	—	.03*	—	−.14	—	−.07*	—	.00	—	.02*
25	—	−.03	—	.01	—	−.02	—	−.08*	—	.00	—	.01
26	—	−.04	—	.02	—	.06	—	−.07**	—	−.01	—	−.02
27	—	−.11**	—	−.01	—	.15**	—	−.06	—	.00	—	−.02
28	—	−.05*	—	.01	—	.27**	—	−.03	—	.01	—	−.02*
29	—	−.09**	—	−.03	—	.16*	—	−.08*	—	.00	—	−.01
30	—	−.06	—	−.03	—	.13	—	−.08	—	−.01	—	−.02*
Trend?[b]		Yes		Yes		Yes		Yes		Yes		Yes

a. A "*" superscript on an estimate indicates that the asymptotic t-statistic associated with it exceeds 1.96 in value (namely, P value for this estimate $>$ .05), and "**" superscript indicates that the value of the t-statistic is greater than 2.57 (namely, P value $>$ .01).

b. Row indicates whether one accepts the hypothesis of equal year effects based on a Wald statistic assuming 95 percent level of confidence. A "yes" signifies rejection of the null, which indicates the existence of a trend. A "no" signifies the opposite.

sample is aging with time and these trends in age effects will tend to offset the year trends in dispersion.

Turning now to the differences in wage dispersion, we see:

- The NLSY tends to predict more dispersion in wages throughout the sample period, especially in the case of women where differentials typically exceed ten percentage points.

There is no instance where an estimated year effect indicates less wage dispersion in the NLSY than in the CPS for women, and the majority of these estimated effects are statistically significant. Only two of the 13 year effects for men are negative, and these two are small and insignificant. The differences in wage dispersion are far less severe for men than for women. No general trend appears in these estimates of the differences in wage dispersion across time.

Controlling for age effects in wage dispersion enlarges most of the estimated values of year differentials for both men and women. Examination of the estimated age effects suggests:

- There is a tendency for the NLSY to exhibit more wage dispersion at younger ages and less at older ages than appears in the CPS, for both men and women.

The age coefficients are always positive for men and women at the younger ages and are always negative in the mid-twenties and beyond. Most are statistically significant, especially in the case of men. In the late teens the NLSY shows more dispersion by 11 to 29 percentage points, while by the late twenties the NLSY shows somewhat less dispersion than the CPS, with the differences typically under 10 percentage points. This relatively greater dispersion in the NLSY at younger ages suggests that the NLSY might be better able to measure extreme values of incomes and hours of work than the CPS during teenage years.

The fifth pair of columns of Table 10a and 10b summarize the differences in the employment rates between the NLSY and CPS for the overall measure $ER(h > 0)$. Summarizing the findings:

- For the overall employment rates of both men and women, there is a distinct drop in the NLSY rate relative to the CPS rate after 1986;
- For men, this drop shows up with the NLSY predicting slightly higher employment rates in the early years of the sample and marginally lower rates in the later years;
- For women, this drop takes the form of the NLSY forecasting somewhat higher employment during the early 1980's and matching the CPS after 1986.

Year and age trends for differences in employment appear relatively small. For men, all of the year coefficients are positive and many are significant before 1985, and these coefficients are all negative and significant after 1986. Thus, the NLSY appears to predict higher male employment during the early years of the sample, and lower rates during the later years, although these differentials are typically small and only reach seven percentage points at their maximum. There are no obvious trends in the differences in employment for women in the NLSY and the CPS. The year effects do indicate a slight relative increase in employment rates for women from 1981 to

1986, with most of the significant and positive year effects occurring during these years. This middle year "bulge" may be nothing more than a statistical artifact, but it is curiously persistent.

Corresponding to the year trends influencing men's overall employment rate are the offsetting age trends. Controlling for age effects we see:

- Men at younger ages appear to have relatively lower overall employment rates according to the NLSY, and higher rates at older ages.

Most estimated age coefficients are negative and many are significant before age 25, and these coefficients are all positive and mostly significant after this age. In the case of women, age effects are quite small and mostly insignificant.

The last pair of columns in Tables 10a and 10b report differentials in the beyond-part-time employment rates captured by the measure $ER(h > 1{,}000)$. These findings indicate:

- For men, the overall and beyond-part-time rates behave similarly in that the NLSY exhibits slightly higher rates in the early years and marginally lower rates in later years
- For women, the NLSY forecasts slightly higher incidence of beyond-part-time employment throughout the sample period.

All of the estimated year effects for men are positive and mostly significant before 1987 and are negative after this year. The relative overprediction by the NLSY declines and reverses sign by the later years of the sample. For women, the NLSY shows higher beyond-part-time employment rates in all years; all coefficients are positive, and all but one are statistically significant after 1980. These differences in beyond-part-time work rates for women, as in the comparisons of employment rates, have no significant trends in calendar time or age space. The differentials in the beyond-part-time rates are generally small for both men and women, typically hovering between two to five percentage points. Controlling for age effects produces nothing of interest in the NLSY-CPS differentials for the beyond-part-time employment rate; there exists no pronounced age trend for either men or women.

B. Differences Between the NLSY and CPS Across Educational Attainments

How do the distributions of labor-market variables compare in the NLSY and CPS for various levels of educational attainments? Do the wages of low-educated workers in the NLSY look like those in the CPS? What about differences in the returns to receiving a college education?

1. Figures Describing Differences Across Education Groups

To answer these questions, the subsequent analysis carries out this comparison for men and women classified into four levels of educational attainments: high-school dropouts (11−); high-school graduates (12); individuals with some college (13–15); and college graduates (16+).

To study how returns to education differ between the NLSY and the CPS, Figures

7–14 summarize how various wage and earnings percentiles evolve over time for the four education groups. The structure of these figures is identical to Figures 1–4. The percentiles are calculated with the weighting procedures described in Appendix A. Once again, when comparing findings across figures, one should be careful to note that vertical scales differ, so a larger vertical distance in one graph does not necessarily translate into a bigger divergence for one education group versus another.

Figures 7–9 plot results for annual earnings over the years covered by the NLSY, and Figures 12–14 graph quantities describing wages. Figures 7 and 11 plot percentiles for high-school dropouts; Figures 8 and 12 graph percentiles for high-school graduates; Figures 9 and 13 present profiles for those with some-college education; and Figures 10 and 14 chart percentiles for college graduates.

To examine differences in the overall trends in educational returns, Figures 15a and 15b compare returns to education in terms of average hourly earnings across education groups, and Figures 16a and 16b compare returns from measures of total earnings. The high-school-completion return is defined as the difference in ln(median earnings) for those with exactly 12 years of school and those with fewer than 12 years of schooling during each year. Only individuals older than 19 are used when defining median earnings for constructing the high-school return. The college return measures the difference in ln(median earnings) for those with 16 or more years of schooling completed and those with exactly 12 years of schooling completed. For the college return, we only use individuals older than age 22. Both of these numbers are weighted averages of age groups in each year where the weights, as discussed above, partially adjust for the sampling variability of the estimated medians. These measures do not adjust for the changing age composition across years.

2. *Summary of Differences Across Education Groups*

Examining the graphs of annual earnings percentiles (Figures 7–10) for men discloses a striking relationship:

- For men, the NLSY predicts uniformly higher earnings for high-school dropouts and graduates than does the CPS; for the college educated, it closely tracks the CPS earnings until the later sample years when it grows more rapidly.

The patterns for women are quite different from those of men. In particular, the uniform divergence in the earnings distributions of the two lower education groups is not present. However, the comparisons for women do share some similarities with those of men in that they show a divergence that begins in the mid-1980s. The earnings percentiles for women indicate:

- NLSY earnings profiles for women with different educations closely track those in the CPS until the later sample years when NLSY earnings grow more rapidly; the size of the differentials in these later years is inversely related to education.

In contrast to the situation faced by men, the greater growth in earnings for the less-educated women primarily shows up in faster growing upper percentiles for the

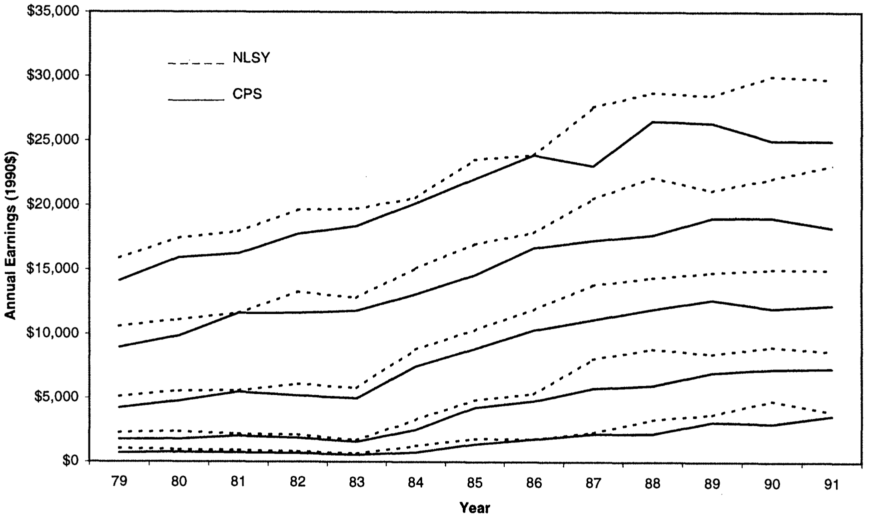

Figure 7a
Total Earnings—Grades 11−, Men

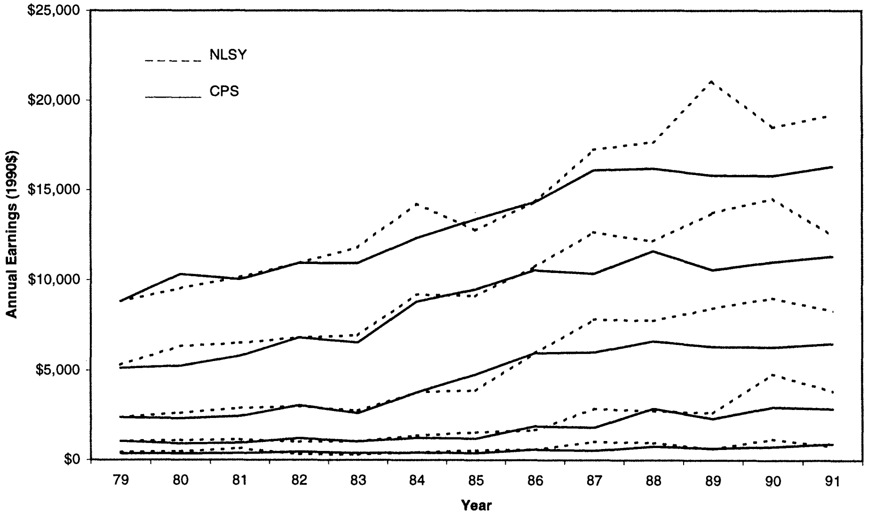

Figure 7b
Total Earnings—Grades 11−, Women

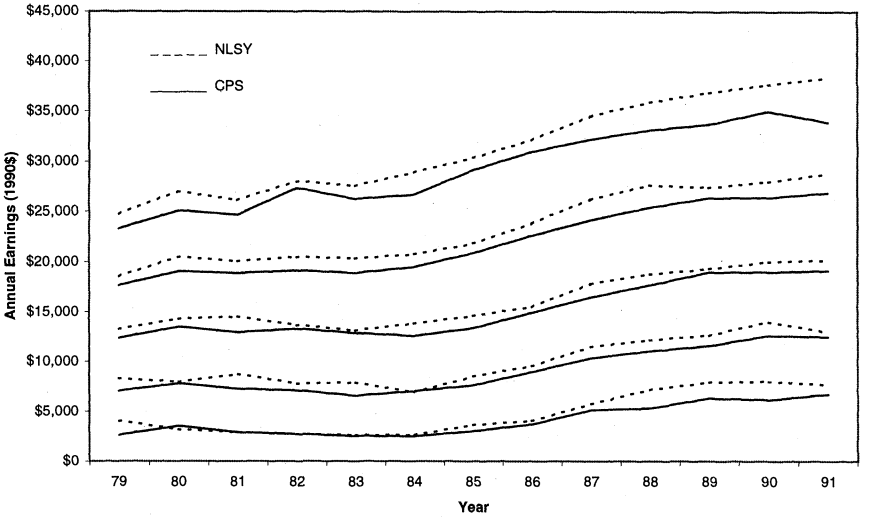

Figure 8a
Total Earnings—Grade 12, Men

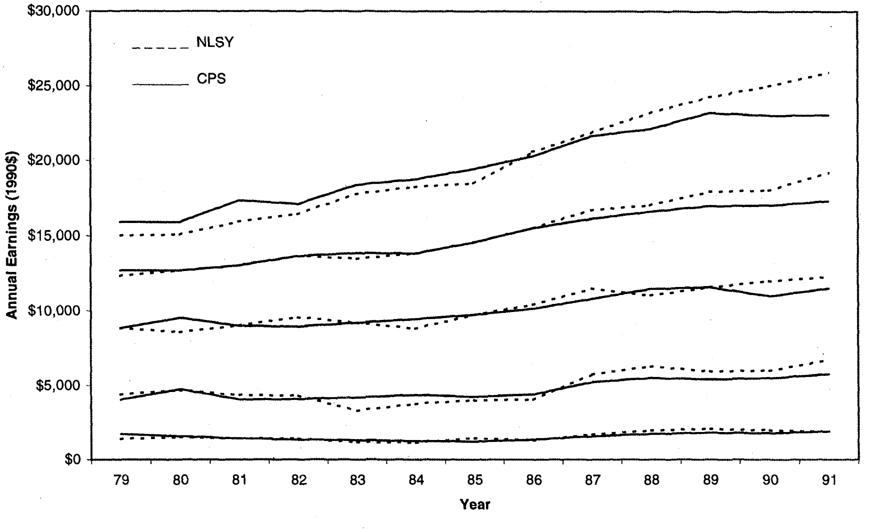

Figure 8b
Total Earnings—Grade 12, Women

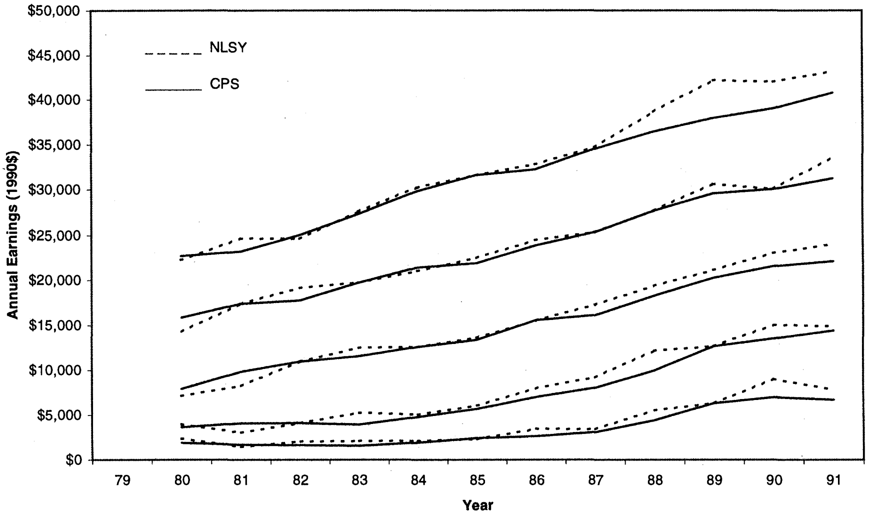

Figure 9a
Total Earnings—Grades 13 to 15, Men

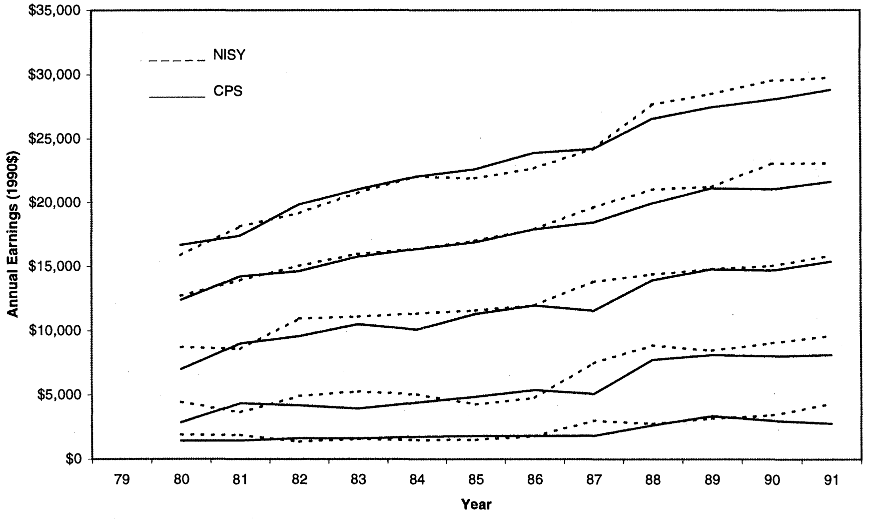

Figure 9b
Total Earnings—Grades 13 to 15, Women

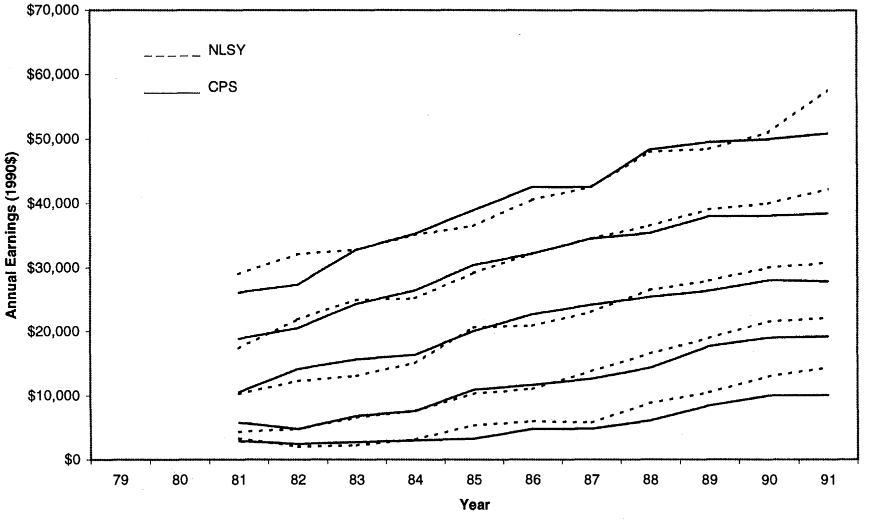

Figure 10a
Total Earnings—Grades 16+, Men

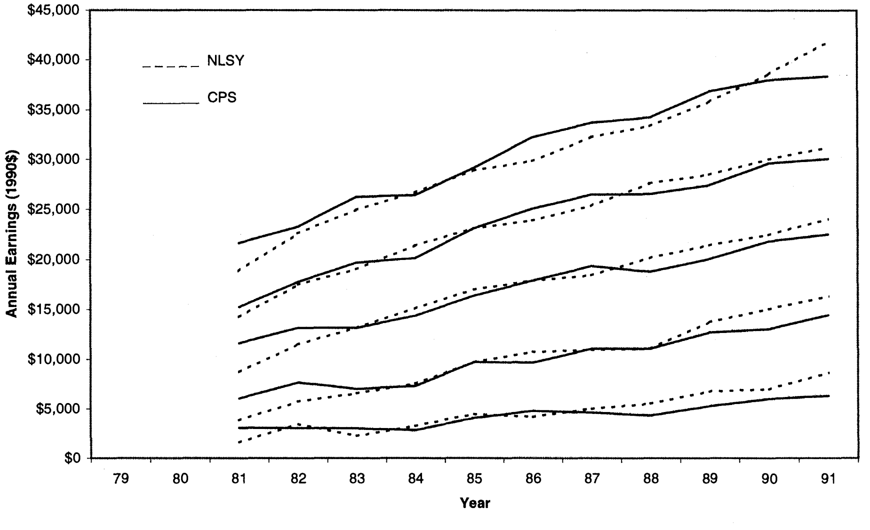

Figure 10b
Total Earnings—Grades 16+, Women

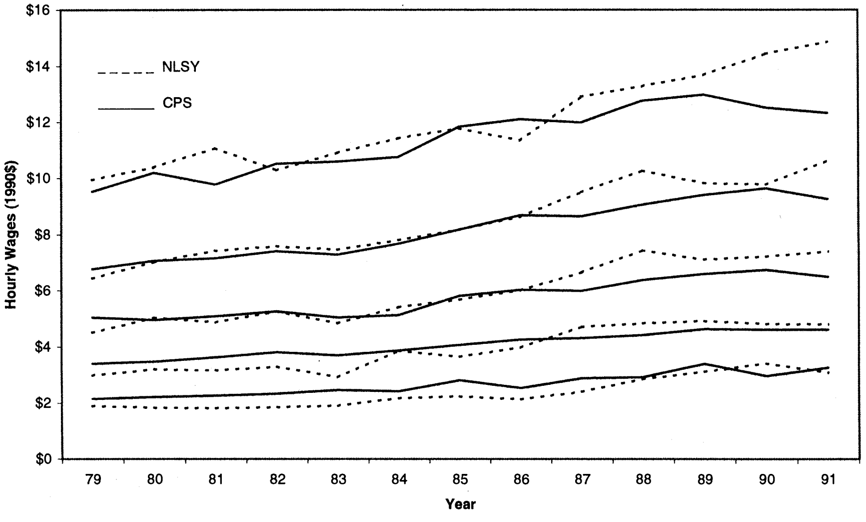

Figure 11a
Average Hourly Earnings—Grades 11−, Men

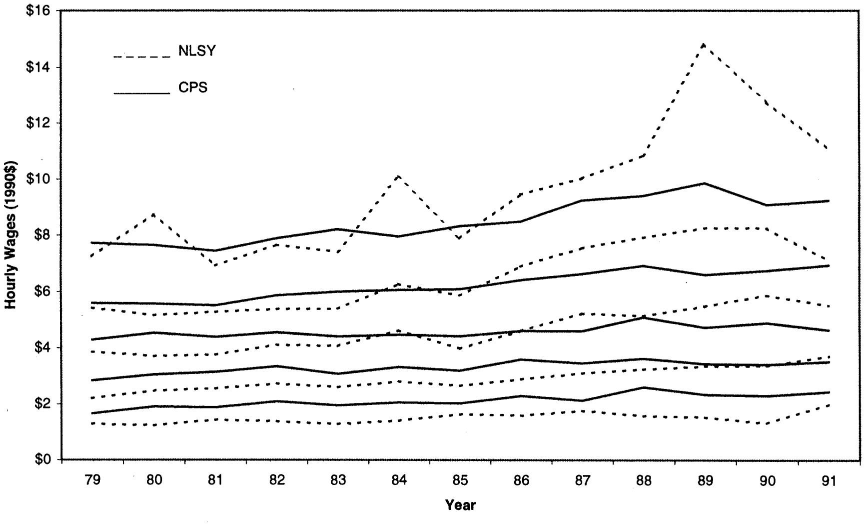

Figure 11b
Average Hourly Earnings—Grades 11−, Women

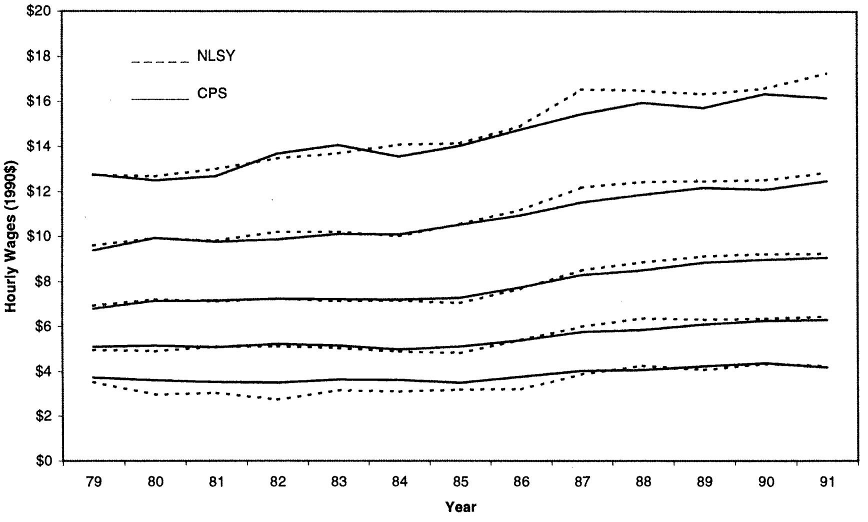

Figure 12a
Average Hourly Earnings—Grade 12, Men

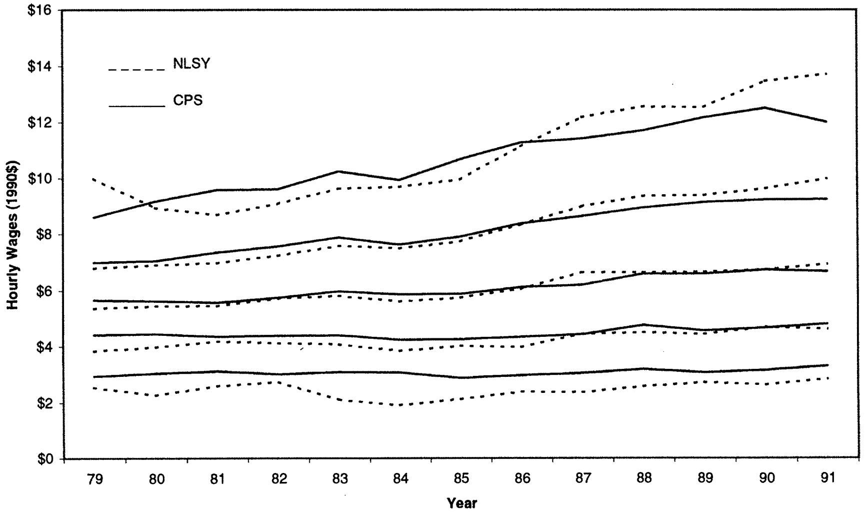

Figure 12b
Average Hourly Earnings—Grade 12, Women

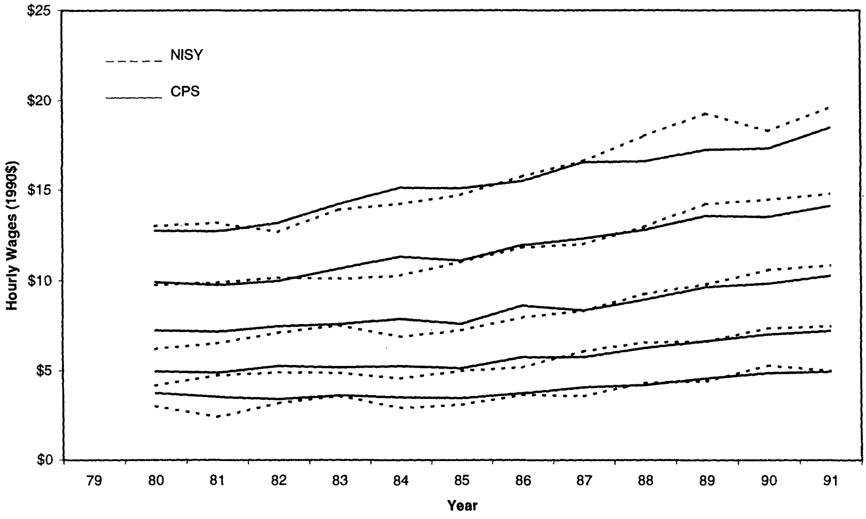

Figure 13a
Average Hourly Earnings—Grades 13 to 15, Men

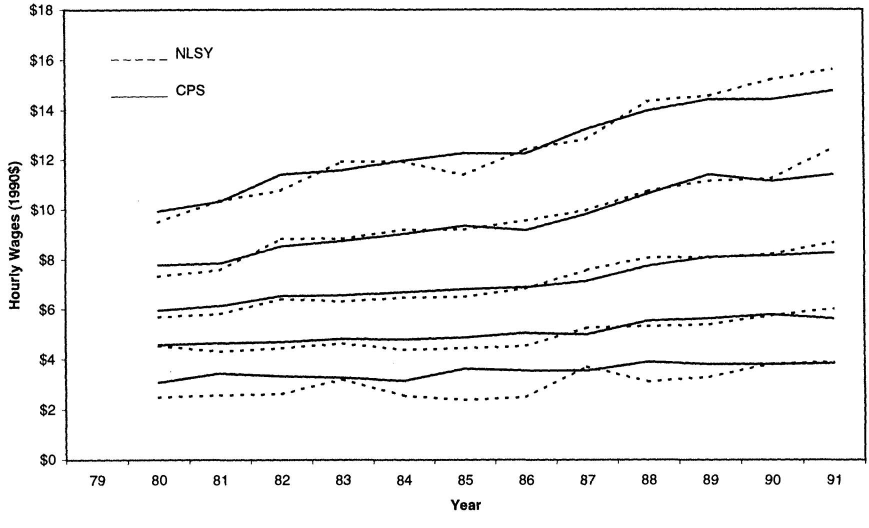

Figure 13b
Average Hourly Earnings—Grades 13 to 15, Women

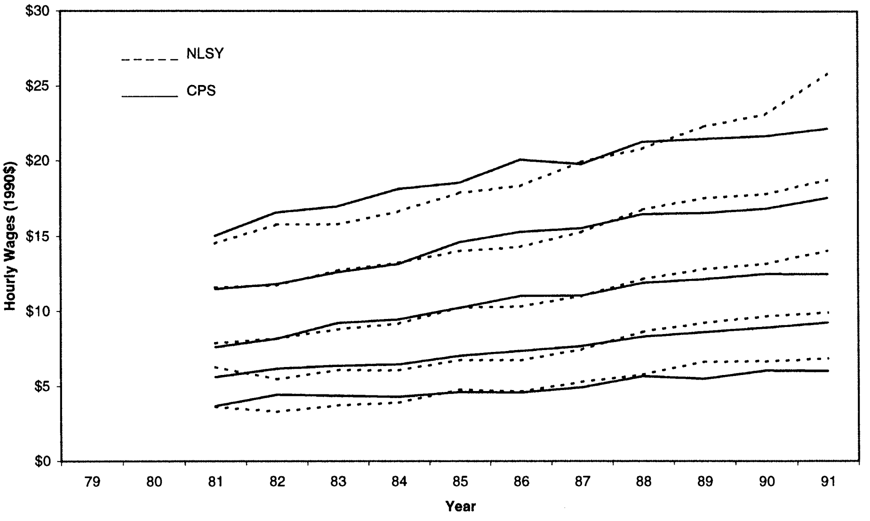

Figure 14a
Average Hourly Earnings—Grades 16+, Men

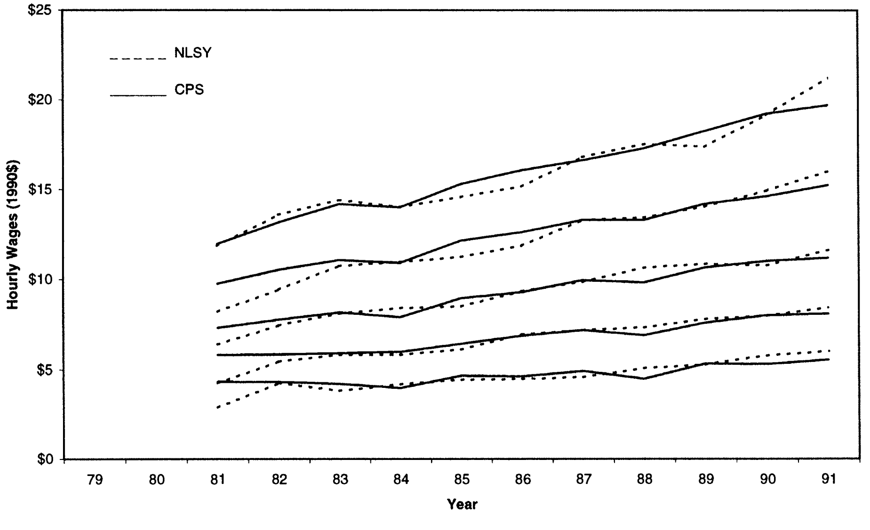

Figure 14b
Average Hourly Earnings—Grades 16+, Women

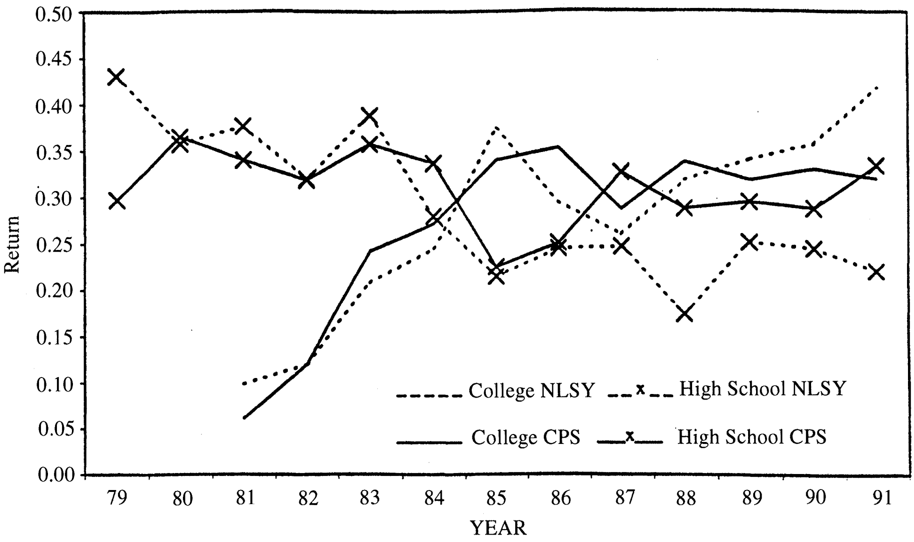

Figure 15a
Returns to Education—Wage, Men

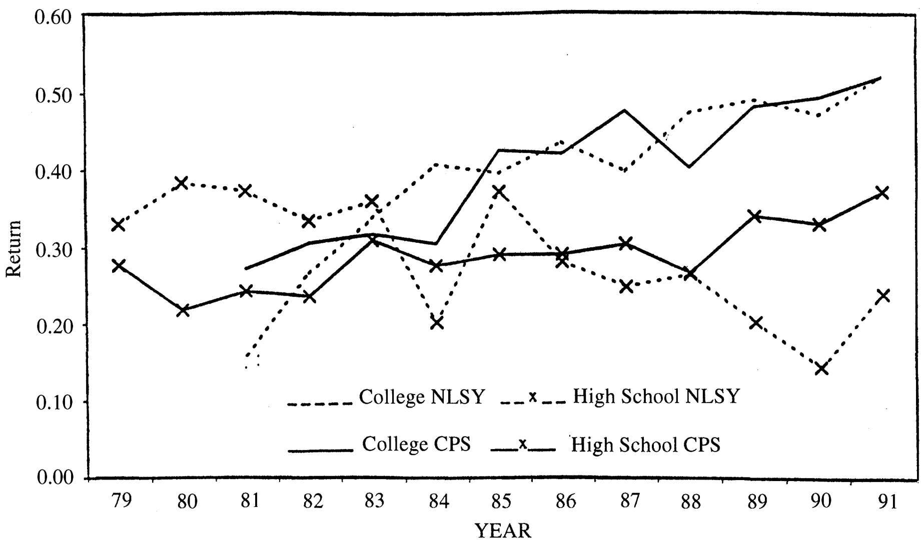

Figure 15b
Returns to Education—Wages, Women

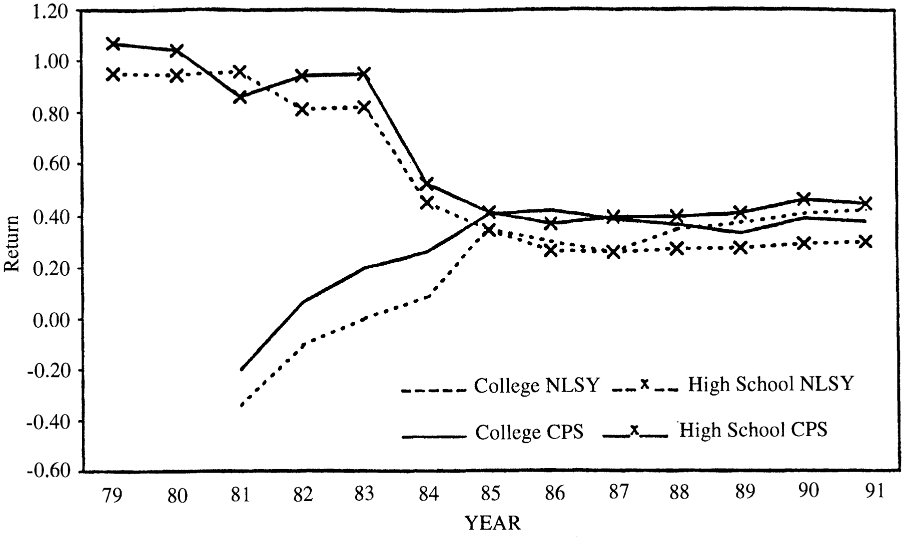

Figure 16a
Returns to Education—Earnings, Men

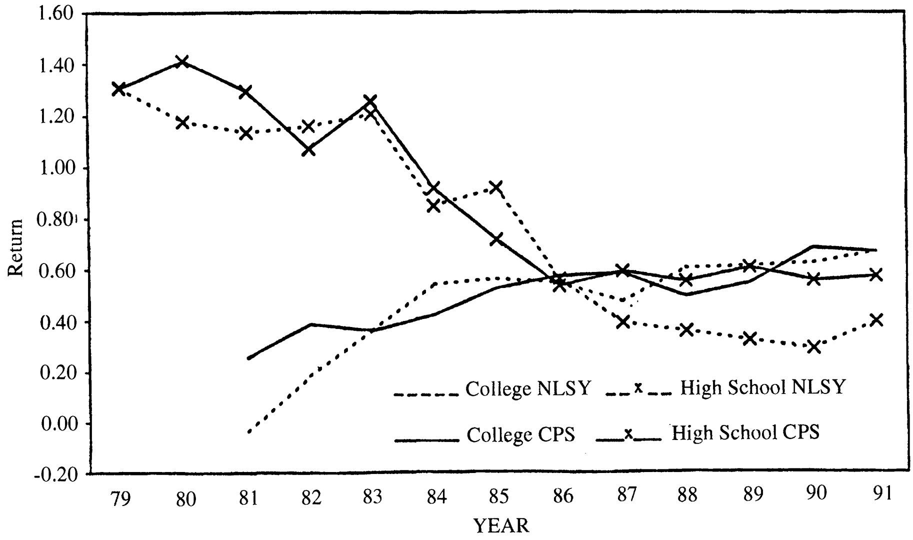

Figure 16b
Returns to Education—Earnings, Women

NLSY, whereas the larger growth for the more educated groups typically occurs in the lower percentiles.

Inspecting the wage-percentile graphs (Figures 11–14) reveals:

- The wages for all education groups for both men and women are more dispersed in the NLSY than in the CPS.

Furthermore, comparing wage levels we see different patterns for men and women:

- For men, NLSY wages for all education levels tend to track those in the CPS until the later years when NLSY wages grow more rapidly.
- For women, wages for all education groups start below those in the CPS, and they either match or exceed the CPS values by the end of the sample period; the largest relative gains accrue to the lowest education levels.

The lower percentiles for women's wages start out lower in the NLSY than in the CPS for all education groups with the largest discrepancy for high-school dropouts. Over the course of the sample period, the NLSY profiles gain ground to either match or surpass the corresponding CPS percentiles by the end of the period. This upsurge in the lower percentiles occurring in the mid to late 1980s is also reflected in the other wage percentiles for all levels of educational attainment.

Figures 15a and 15b indicate that the rates of return to a college degree are similar across the two samples.[19] There is, however, a slight tendency for the NLSY to predict somewhat lower returns to college for men, but the overall trends in the college return are similar across the samples. The returns to high school in the two samples, however, do appear to be different. They match up well during the early years of the sample, but the NLSY predicts lower returns during the last few years. Except for the divergence of the high school returns in the late 1980's, overall, the trends in the returns in the two samples are quite similar.

3. Estimation of Differences in Educational Returns

To assess the extent to which discrepancies in educational returns are significantly different from one another in the NLSY and CPS, we apply regression framework (4.1) to estimate these differences. The dependent variable $Z_{\alpha t} = \mu^{n}_{\alpha t} - \mu^{c}_{\alpha t}$ is used to measure differences in returns to high school and college, considering both wages and annual earnings as the relevant forms of compensation. The ranges of values for the years t and the ages α used as observations in constructing the time and age dummies d_t and $D_{\alpha..}$ in (4.1) depend on the particular educational composition under investigation.

We consider two measures of returns to high-school: one uses wages to calculate returns, and the other uses earnings. Using wages as a basis for computing these returns, we calculate the weighted medians $P50_{\alpha t}(\omega_{it}|edu = 11-)$ and $P50_{\alpha t}$ $(\omega_{it}|edu = 12)$, where the notation $\omega_{it}|edu = j$ designates that computation occurs

19. Care must be used in interpreting the low rates of return to college in the early years of the samples because these results do not control for the age composition of the sample. Further, the high returns to college for women measured by total earnings most likely reflect the tendency for college educated women to be less likely to work part-time than high-school educated women.

only over observations in the age-year cell (α, t) who have completed j years of schooling. The group of high-school dropouts (j = 11−) includes individuals who are old enough to have graduated from high school and who still only have 11 or fewer years of schooling. High-school graduates (j = 12) are the group who are at least two years beyond the age of graduating from high-school who report 12 years of education. Our measure of the return to a high-school education is given by

$$\mu_{\alpha t} = \ln P50_{\alpha t}(\omega_{it} | edu = 12) - \ln P50_{\alpha t}(\omega_{it} | edu = 11-).$$

This quantity can be interpreted as the percent difference in the return to a high school degree (relative to being a high school dropout) in the two samples. As an alternative to using wages as the basis for calculating educational returns, we consider the statistic

$$\mu_{\alpha t} = \ln P50_{\alpha t}(y_{it} | edu = 12) - \ln P50_{\alpha t}(y_{it} | edu = 11-)$$

which assumes that annual earnings is the relevant measure of compensation. This quantity can be interpreted as the percent difference in the return to a high school degree (relative to being a high school dropout) in the two samples.

Similarly, we consider two measures of returns associated with the completion of college over a high school degree. Analogous to the above high school returns, we construct the statistics

$$\mu_{\alpha t} = \ln P50_{\alpha t}(\omega_{it} | edu = 16+) - \text{In } P50_{\alpha t}(\omega_{it} | edu = 12)$$

$$\mu_{\alpha t} = \ln P50_{\alpha t}(y_{it} | edu = 16+) - \text{In } P50_{\alpha t}(y_{it} | edu = 12).$$

College graduates (j = 16+) are the group beyond age 23 and who report 16 or more years of education.

We estimate these equations using the procedures and sample compositions described in Section A3 and Appendix A. Imposing the age selection implied by the educational groupings, for the high school returns there are 77 cell differences, a full set of year effects, and age effects from age 20 to age 30. For the college returns, there are 54 cell differences, year effects from 1981 to 1991, and age effects from age 23 to age 30.

4. Summary of Differences in Educational Returns

Tables 11a and 11b report estimates of the differences in the returns to education between the NLSY and CPS. The layout of these tables mirrors Tables 10a and 10b. The left group of columns presents the results based on wages, and the right group lists estimates using earnings to compute the return measures.

Inspection of the findings in Table 11a reveals:

- The wage returns to high-school completion for young men are lower in the NLSY at young ages, higher at older ages, and lower in the later sample years; the earnings returns are lower in the NLSY than in the CPS in nearly every year.

With the inclusion of age controls in men's wage specifications, the estimated age coefficients for high-school returns are all negative and mostly significant before age 25 (typically ranging between five to ten percentage points); the age effects are positive and usually significant thereafter (reaching above ten percentage points at the oldest ages). The estimated year effects for wages are all negative after 1986 (exceeding ten percentage points after 1987) and most are significant. These age and year patterns combine in the wage equation to produce negative estimates for the year effects without age controls for all but two years. The inclusion of age controls does little to alter the time pattern of the estimated high-school earnings returns for men; the year effects are consistently negative and often significant throughout the sample period. The extent of underestimation by the NLSY does not seem to follow much of a trend, varying between zero and 21 percentage points.

For both hourly wages and annual earnings, the estimated year effects for men's college returns suggest:

- The NLSY tends to predict lower wage and earnings returns to college for men in the early years of the sample, higher returns at young ages, and lower returns at older ages than the CPS.

The evidence for this conclusion is somewhat more mixed than is the support for the above inferences about the high-school returns. Although the year effects for earnings are all negative and mostly significant before 1989 (with about half exceeding 15 percentage points), the year effects for wages are negative only in the mid-1980s and are much smaller. The age effects do tend to be positive at young ages and negative at old ages, but few are significant. The NLSY differentials in men's returns to college as measured by wages move over time and across ages in the opposite direction of the men's high-school returns.

The findings for young women presented in Table 11b suggest:

- The NLSY predicts lower wage and earnings returns for high-school completion for women in the later years of the sample; otherwise, the NLSY and the CPS agree in their predictions for high-school and college returns.

The only consistent pattern found among the coefficients in Table 11b are the estimated year effects in the later years of the NLSY sample in both the wage and earnings equations for high-school completion. The returns to a high school degree follow roughly the same time patterns as the returns for men, but the trends are less striking and are not as consistent. The age effects for the differences across the two data sets in the returns to high-school for young women are also less evident. If anything, the results suggest that the differential is decreasing in age, which is the opposite effect from that found for young men.

The differentials in returns to college degrees for women are different from those found for men. There is no compelling evidence for either an increasing or decreasing difference in the measured returns with time, and the age effects are typically small and insignificant. Overall the returns to college for women appear to agree in the NLSY and the CPS.

Table 11a

Difference in Returns to Education Between the NLSY and CPS for Men[a]

Coefficients	Hourly Wages				Annual Earnings			
	High School Returns		College Returns		High School Returns		College Returns	
Year Effects								
79	−.02	.09	—	—	−.06	.00	—	—
80	−.15**	−.05	—	—	−.25**	−.21**	—	—
81	.03	.13	.03	.06	.01	.05	−.08	−.09
82	−.05**	.04	.05**	.01	−.15**	−.10*	−.22*	−.31*
83	−.01	.07	−.04	−.09	−.04	−.01	−.11	−.19*
84	−.02	.04	−.06**	−.09*	−.11*	−.08	−.17**	−.23**
85	−.01	.04	−.01	−.04	−.04	−.01	−.04	−.09*
86	.03	.06	−.08*	−.10**	−.13**	−.11**	−.03	−.06
87	−.03	−.04	−.07**	−.08**	−.08	−.09	−.15*	−.17**
88	−.13**	−.17**	−.04	−.04	−.11	−.12*	−.02	−.03
89	−.08	−.14**	.03	.03	−.05	−.07	−.01	−.00
90	−.04	−.12*	−.01	.00	−.13*	−.16*	.00	.03
91	−.16**	−.24**	.09**	.11**	−.17**	−.21**	.07**	.12**

Age Effects								
18	—	—	—	—	—	—	—	—
19	—	—	—	—	—	—	—	—
20	—	−.07	—	—	—	.00	—	—
21	—	−.16**	—	—	—	−.12	—	—
22	—	−.08*	—	—	—	−.01	—	—
23	—	−.07*	—	−.03	—	−.05	—	.02
24	—	−.05	—	.09**	—	−.04	—	.16**
25	—	.02	—	.06*	—	.01	—	.06
26	—	.05	—	.01	—	.02	—	.03
27	—	.05	—	−.01	—	−.04	—	−.02
28	—	.03	—	−.04	—	.04	—	−.05
29	—	.17**	—	−.04	—	.13	—	−.09**
30	—	.11**	—	−.06	—	.06	—	−.10*
Trend?[b]		Yes		Yes		No		Yes

a. A "*" superscript on an estimate indicates that the asymptotic t-statistic associated with it exceeds 1.96 in value (namely, P value for this estimate > .05), and "**" superscript indicates that the value of the t-statistic is greater than 2.57 (namely, P value > .01).

b. Row indicates whether one accepts the hypothesis of equal year effects based on a Wald statistic assuming 95 percent level of confidence. A "yes" signifies rejection of the null, which indicates the existence of a trend. A "no" signifies the opposite.

Table 11b
Difference in Returns to Education Between the NLSY and CPS for Women[a]

Coefficients	Hourly Wages				Annual Earnings			
	High School Returns		College Returns		High School Returns		College Returns	
Year Effects								
79	−.09	−.17*	—	—	−.30**	−.21	—	—
80	.15**	.08*	—	—	−.07	.01	—	—
81	.08	.01	−.06	−.03	−.27*	−.19	−.26	−.19**
82	.05	−.02	−.04	−.01	.12	.17	−.12	−.05
83	.02	−.03	.01	.04	.00	.02	.06	.13*
84	−.01	−.06	.06	.10	.08	.07	.08	.12**
85	.10*	.07	−.01	.02	.15	.15	.03	.06
86	.06	.06	.01	.02	.18	.16	.02	.04
87	.00	.00	−.09*	−.08**	−.27**	−.33**	−.06	−.06
88	.03	.06	.08*	.08**	−.17**	−.20*	.05	.05
89	−.18**	−.13**	.04	.03	−.23	−.27*	.08*	.07
90	−.17**	−.11*	.00	−.01	−.24**	−.26**	−.02	−.05
91	−.17*	−.11	−.05	−.07	−.28	−.28	.02	−.02

Age Effects								
18	—	—	—	—	—	—	—	—
19	—	—	—	—	—	—	—	—
20	—	.09*	—	—	—	−.22*	—	—
21	—	.06	—	—	—	.04	—	—
22	—	.07	—	—	—	−.04	—	—
23	—	.07	—	−.03	—	−.01	—	−.07
24	—	.03	—	−.04	—	.03	—	−.08*
25	—	−.04	—	−.03	—	.12	—	−.04
26	—	.05	—	−.02	—	.24*	—	.00
27	—	−.09*	—	.00	—	−.15*	—	.06
28	—	−.09	—	.08*	—	−.10	—	.07
29	—	−.09	—	.01	—	.17	—	.07
30	—	.05	—	.04	—	−.09	—	−.00
Trend?[b]		Yes		Yes		Yes		Yes

a. A ''*'' superscript on an estimate indicates that the asymptotic t-statistic associated with it exceeds 1.96 in value (namely, P value for this estimate > .05), and ''**'' superscript indicates that the value of the t-statistic is greater than 2.57 (namely, P value > .01).

b. Row indicates whether one accepts the hypothesis of equal year effects based on a Wald statistic assuming 95 percent level of confidence. A ''yes'' signifies rejection of the null, which indicates the existence of a trend. A ''no'' signifies the opposite.

C. Overall Summary of Differences Between the NLSY and CPS

What do the various findings detailed above tell us about how the NLSY and the CPS differ in their characterizations of the labor-market experiences of men and women? The following discussion integrates the information in the various bullets to summarize differences in four measures of labor-market outcomes: (i) earnings, (ii) wages, (iii) returns to high-school and college educations, and (iv) employment rates.

1. Differences for Men

Refering to the bullets concerning men's earnings, we see that the NLSY predicts uniformly higher earnings for men for high-school dropouts and graduates than does the CPS. For the college educated, NLSY earnings closely track CPS earnings until the later sample years when those from the NLSY grow more rapidly. These educational differences readily account for our finding in the cohort analyses that the NLSY systematically forecasts higher earnings throughout the sample period, with differentials that tend to grow over time. Furthermore, the cohort analysis determines that the NLSY predicts dramatically lower dispersion in earnings relative to the CPS in the last half of the survey years.

Examining men's wages, the educational breakdowns suggest that NLSY wages for all educations roughly track those in the CPS until the later years when NLSY wages grow more rapidly. In a cohort analysis, the evidence supports the conclusion that the NLSY forecasts slightly lower wages before 1986 and closely follows the CPS thereafter. These apparently conflicting findings can be reconciled by recognizing the influences of differing educational compositions of the NLSY and the CPS at young ages. Both the educational and cohort analyses clearly show that the NLSY produces a higher dispersion in wages throughout the sample period.

For returns to a high-school education, the NLSY overwhelmingly portrays a much lower advantage for completing high school. The earnings returns to completing high-school over dropping-out for young men are lower in the NLSY than in the CPS in all years. The wage returns to high-school completion over dropping-out for young men start out lower in the NLSY than in the CPS at young ages and become higher at older ages, a result consistent with the view that life-cycle growth in the returns to high-school completion is higher in the NLSY.

Turning to returns to college education, the NLSY once again depicts lower returns to education over much of the sample period, although the discrepancies with the CPS are smaller than for high school and disappear by the end of the period. In particular, the NLSY tends to predict lower wage and earnings returns to college for men than the CPS until the latest years of the sample. The age effects reveal relatively less life-cycle growth in the NLSY than the CPS for both the wage and earnings returns to college over high school completion.

The points comparing the NLSY and the CPS for young men's employment indicate fairly close matches throughout the sample period. Looking at the discrepancies, the NLSY predicts slightly higher rates of employment for both the overall and the beyond-part-time rates in the early years of the sample and marginally lower rates in the later years. Considering ages instead of years, the results indicate that men at younger ages appear to have relatively lower overall employment rates according to the NLSY, and relatively higher rates at older ages.

2. *Differences for Women*

From the summary points for women's earnings, we see that the NLSY earnings profiles for women with different education levels closely track those in the CPS until the later sample years when NLSY earnings grow more rapidly. The size of the differentials in these later years is inversely related to education. These educational differences are fully consistent with our findings in the cohort analyses indicating that the NLSY forecasts higher earnings for women in the later years of the sample. As in the case of men, the cohort analysis also shows that the NLSY predicts sharply lower earnings dispersion relative to the CPS in the last half of the survey years.

For women's wages, the results distinguishing educational attainment indicate that the NLSY wages for all groups start below those in the CPS, and that they either match or exceed the CPS values by the sample period's end. The largest gains accrue to the lowest education levels. Again this result is entirely consistent with the cohort findings which reveal that the NLSY predicts lower wages for women before 1986 and matches the CPS thereafter. Both the educational and cohort analyses support the proposition that the NLSY tends to predict higher dispersion in women's wages throughout the sample period.

Compared to men's educational returns, the results for women show a similar association between the NLSY and the CPS for the completion of high-school but greater conformability for college returns. As in the case of men, the NLSY predicts lower wage and earnings returns for high-school completion over dropping-out for women in the later years of the sample. The NLSY and the CPS broadly agree in their predictions for women's high-school returns over the life cycle. Further, they agree on how the returns to college evolve over time and across ages for young women.

Finally, the summary points related to the employment rates of women indicate that the NLSY generally predicts higher rates. The NLSY forecasts somewhat higher overall employment for women during the early 1980s. It projects slightly higher incidence of full-time employment for women compared to the CPS throughout the sample period.

3. *Common Differences*

Taking the entire set of graphs together, the NLSY and CPS convey similar pictures of the overall trends in earnings and wages for the various education groups. The earnings and wage profiles show a considerable amount of variation across ages and over time, and both data sets broadly trace out the basic features of these trends.

The majority of differences between the NLSY and CPS show up primarily in three areas: measuring returns to education, earnings inequality, and differentials in labor-market outcomes that tend to begin or end sometime in the mid-1980s (usually in the 1986–87 period). For educational returns, the NLSY suggests that high-school dropouts fare relatively better over time than indicated in the CPS. This difference arises from the fact that earnings and wage measures for the lowest education group in the NLSY tend to grow relative to those in the CPS. The returns to college education measures, on the other hand, appear roughly consistent in the two data sets. By our measures of inequality, earnings inequality tends to be smaller in the NLSY than in the CPS, while wage dispersion is definitely larger in the NLSY. Finally, the relative differentials in wage, employment, and dispersion variables often seem to change in character just after 1985.

These findings largely support those of Gottschalk and Moffitt (1992) in their comparisons of the NLSY and March CPS data sets. Their analysis is limited because they rely on only three years of data. Nonetheless, they reach four principle conclusions: (i) mean earnings in age-education cells differed substantially, with differences appearing to narrow with age; (ii) the annual labor force participation rates (LFPR) are higher in the NLSY than the CPS, especially at younger ages; (iii) the NLSY exhibits smaller variations in earnings within years; and (iv) both the LFPRs and earnings dispersions grow faster in the CPS than the NLSY. Our findings confirm the first part of (i), but do not lend support for the second part declaring a narrowing of the differences; our examination of more years and the removal of time trends does not produce much of a pattern in age effects. Our results definitely affirm (ii) and (iii). Although (iv) is consistent with our findings, its meaning is somewhat debatable. The greater growth in employment rates and earnings dispersion in the NLSY mostly occurs as a consequence of sharp breaks in the NLSY and CPS series after the 1985–86. period.

4. Possible Reasons for the Differences

The comparisons of the labor market outcomes between the NLSY and the CPS are difficult to interpret. Ideally, one would like to be able to treat the CPS as a reliable sample of individuals from the same population as the NLSY with accurately measured earnings and work variables. If this were the case, then the CPS would be an appropriate cross-sectional baseline that could be used to judge the cross-sectional validity of the NLSY over time; statistically significant deviations of the NLSY from the CPS would then indicate the evolution of potential problems with the NLSY sample.

There are, however, major drawbacks to using the CPS as a baseline. One obvious drawback is that the relevant population for the CPS is the population in the United States at each interview date, while the relevant population for the NLSY is the population as of 1979. Demographic changes due to such factors as immigration during the 1980's make the relevant populations for the CPS and the NLSY quite different, and it is not possible to control for this difference in the baseline populations. A more important drawback concerns the ability of the CPS to accurately measure labor market experiences for youth. Many youth are not actually interviewed in the CPS, and the proxy responses in the CPS may not accurately measure the youths' labor market experiences. Also, nonresponse in the CPS are typically replaced by hot deck procedures. The little evidence available on nonresponse in the CPS suggests that nonrespondents are not a random subset of the population. Despite these drawbacks, we are forced to use the CPS as a baseline for comparison, since there is no other standard available for judging how well the NLSY matches the United States population at each point in time.

Even with these considerations in mind, significant differences exist between the NLSY and the CPS data sets in measuring educational returns and earnings inequality. We must express apprehension about the ability of the two data sets to characterize the rates of changes in some labor market trends. On the basis of these findings we cannot yet provide a definitive answer to the question of whether the NLSY measures the labor market experiences of youth accurately. In some instances, we can think of good reasons why the two data sets might be different with the NLSY providing the more accurate information. In other instances, we cannot offer an adequate explanation.

One promising source for explaining discrepancies between the NLSY and the CPS involves sample designs and data collection procedures. The greater dispersion in average hourly earnings found in the NLSY no doubt reflects a larger variability in hours of work in the NLSY versus the CPS. Such a discrepancy is not surprising given the fact that the NLSY asks questions about hours-of-work separately for each employer and the CPS asks about hours worked in a typical week. Furthermore, the CPS underwent changes in its methods of data collection and design shortly after 1985. Moreover, in 1989 three major modifications were introduced by the Census Bureau in the income imputation procedures used to process the public-use sample. The modifications were: (i) an extension of the hot-deck procedure to reflect the full detail of the March income supplement rather than a more limited set used in the old procedure; (ii) the simultaneous imputation of earnings, unearned income and noncash benefits for an individual compared to separate imputations for each component in the old system; and, (iii) the old system occasionally replaced reported information with imputed values and the new system never replaces reported values.[20] While these new procedures were applied retrospectively to the March 1988 survey and all subsequent surveys, an analysis conducted by the Census Bureau suggests that this modification resulted in only minor changes in the distribution of earnings. Hence, these changes to the CPS only explain a small part, if any, of the divergence between the NLSY and the CPS observed after the 1985–1986 period.

There were also changes over time in the collection of data in the NLSY. In 1987 only, the survey was conducted using a telephone interview instead of an in-person interview. More importantly, the NLSY underwent a change in its fielding period (namely, the horizen during which interviews are conducted). Prior to 1988, this period was February-May in each calendar year; whereas in 1988 and thereafter, the fielding period shifted to August-September. This delay in the fielding period may introduce more recall error in respondents' reporting of their annual earnings in the previous calendar year because of the additional six months (on average) of recall required to report annual income from the previous calendar year. For example, if there is growth in respondents' earnings over these additional six months, respondents are likely to overstate their earnings in the previous calendar year when asked in August and September compared to February and March. Recall errors

20. See *Current Population Survey: March 1989 Technical Documentation CPS-89-3* for a complete discussion of these modifications.

of this type provide one possible explanation for the observed patterns described above.

V. Conclusion

How representative is the NLSY in its depiction of youths' labor market experiences? How does attrition affect this representativeness? Our findings show that about 10 percent of the original respondents are regularly missing in the later years of the NLSY, and as much as 20 percent of the sample members in these late years are returnees (i.e. respondents missing at least one interview in previous years). This study conducts two types of analysis to evaluate the potential consequences of these features. The first examines how the earnings and work activities of attriters and returnees differ from NLSY respondents who remain as sample members. The second analysis compares the NLSY and the CPS in their depictions of the labor market experiences for youths in the NLSY cohorts. Together these analyses provide substantial information on the characteristics of the NLSY. However, answering the questions posed above still requires a considerable amount of speculation.

In our first analysis, we identify the extent to which nonrespondents come disproportionately from particular segments of the labor market. We examine attrition in two ways: "prospective attrition," where we evaluate the experiences of attriters prior to their exit from the sample; and "retrospective attrition," where we evaluate the experiences of returnees after they reenter the sample. For both groups, the attrition patterns are quite similar for men and women.

Prospective attriters are more likely to be nonemployed, as opposed to being in-school or employed, prior to leaving the sample. Among working teenagers, these attriters typically have higher wages and earnings than their nonattriting contemporaries, but after age 20 attrition occurs randomly across wage and earnings distributions. Among working high-school dropouts, prospective attriters tend to have higher wages than sample stayers throughout their twenties. Attrition is mostly random for the other educational attainment groups.

Returnees typically work less and earn less than sample members who have never attrited. Returnee wages are lower for returnees who are high-school dropouts and college grads, but not for high-school graduates or those with some college. The employment differentials are larger for women than for men, except in the teenage years when the opposite is true, but wage differentials are typically smaller for women than for men.

Our second analysis compares the NLSY and the CPS. We find that these data sets convey similar pictures regarding the overall trends of young men's and women's earnings and wages during the 1980s. In three important respects, however, the NLSY and CPS do not match well.

First, significant differences exist in measuring returns to education. The NLSY suggests that high school dropouts fare relatively better over time than does the CPS. The NLSY predicts uniformly higher earnings for male high-school dropouts and graduates than the CPS. For both men and women, earnings and wage measures for the lowest education group in the NLSY tend to grow relative to the CPS. Comparing wage returns to college education, the NLSY tends to predict lower wage and earnings returns to college for men until the latest years of the sample than does the CPS. For women, the NLSY and the CPS generally agree in their predictions of college returns.

Second, earnings and wage inequality differ in the NLSY and the CPS. For both men and women, the NLSY shows a much smaller dispersion in earnings during the last half of the survey years. On the other hand, the wages for both men and women for all education groups are *more* dispersed in the NLSY than in the CPS over the entire sample period.

Third, and finally, relative changes between the NLSY and the CPS generally occur in the mid-1980s (usually in the 1986–87 period), with differentials either widening or narrowing at this crossover point. There are several examples of this phenomenon. The shrinkage in earnings dispersions occurs after 1987. NLSY wages for men of all educations track those in the CPS until the mid-1980s when NLSY wages grow more rapidly; NLSY earnings of college-educated men and women of all educational attainments also follow this pattern. The NLSY predicts higher employment rates during the early years of the sample for both men and women.

There are technical differences in the sampling and compositions of the two surveys which may explain some of these disparities. The NLSY asks questions about hours of work by employer, while the CPS asks about hours worked on average. This factor readily justifies a larger variability in hours of work in the NLSY and, thus, a greater dispersion in average hourly earnings found in the NLSY. Furthermore, the population sampled for the CPS is youth in the United States at each interview date, while the relevant population for the NLSY is the population as of 1979. Demographic changes arising from immigration during the 1980s makes these populations different, even with weighting. Also, many youth are not actually interviewed in the CPS, and the proxy responses in the CPS generated by ''hot deck'' procedures may not accurately measure the youths' labor market experiences. Finally, both the NLSY and the CPS underwent changes in their methods of data collection and design shortly after 1985—the NLSY shifted its fielding periods, and the CPS made several major modifications in the income imputation procedure used to report its data. These changes might assist in explaining the shifts in the NLSY and the CPS trends in the late 1980's.

Even with these considerations, some of the differences in the NLSY and the CPS are difficult to reconcile. The relatively high wages and earnings for the lowest education group is especially puzzling, and we must express some apprehension in the ability of the NLSY to characterize some labor-market trends if the CPS is accepted as truth. Of course, many believe it is the NLSY and not the CPS that presents the more accurate picture of youths' earnings and work experiences for it is the NLSY that surveys youth directly.

Do the attrition findings obtained in our first analysis assist in explaining the discrepancies between the NLSY and the CPS? In our judgment, these findings cannot explain the NLSY-CPS discrepancies associated with educational returns, or with the enhanced growth in wages and earnings observed in the NLSY. Our attrition results tell us that attriters tend to have higher wages and earnings before they leave the sample, and they have lower wages and earnings upon their return. Losing the higher compensated workers from the NLSY suggests that its compensation profiles should understate the actual profiles, assuming these workers tend to hold their positions in the distributions in future years. The fact the attriters who return to the NLSY are at the bottom of the wage and earnings distributions further contributes to a downward influence on NLSY profiles, for growth would be *higher* without the inclusion of returnees. Thus, the disparities between the CPS and NLSY would be greater, not smaller, in the absence of attrition.

Nor do the consequences of attrition on employment seem to explain the NLSY-CPS discrepancies. Our attrition findings tell us that attriters have lower employment before they leave the sample and lower employment upon their return. Assuming that persons with lower employment rates come disproportionately from the segments of the population with the lower wages and earnings, the loss of prospective attriters tends to heighten wage and earnings growth in the NLSY, while the inclusion of the returnees tends to do just the opposite. Since our task is to explain higher growth rates in the NLSY, we can ignore the influence of returnees on the NLSY as a candidate for reconciling NLSY-CPS differentials. This leaves the loss of low-employment prospective attriters as a potential explanation for the differences in wages and earnings. However, this attrition effect has no chance of reconciling the differentials in employment rates observed in the NLSY and the CPS. The NLSY predicts higher employment rates in the early years of the sample and matches the CPS thereafter. With the loss of low-employment prospective attriters, however, the NLSY employment rate should grow relative to the CPS, not fall.

Despite the differences in the two surveys, our analysis offers little basis for suspecting that the NLSY presents an inaccurate picture of youths' labor market experiences. In many instances we can think of good reasons why the NLSY provides more accurate information than the CPS. We present evidence suggesting that attrition from the NLSY is not a random process, but we have little support for contending that this process induces biases of any consequence. Indeed, our analysis shows that the NLSY tracks the major features of the CPS trends for youth, a monumental feat given the vigorous evolution of these trends both over time and across ages.

What are our recommendations for using the NLSY and for interpreting its findings? The NLSY–CPS differences and the nature of attrition uncovered in this work should undoubtedly be sources of concern for analyses using the NLSY. However, we are not convinced that one should exercise more caution with the NLSY than with most other multi-purpose data sets. Our analysis in Section II shows minimal repercussions associated with using alternative sample compositions as long as one applies weights in estimation. The distributions of earnings, wages, and hours look essentially the same regardless of whether one computes them using the entire sample of the NLSY, just the nationally representative (namely, random) component of the sample, or subsamples constructed including only persons who never attrit from the NLSY. It matters whether one weights the data in estimation, but it appears not to matter whether one uses the first-year weights calculated to account for the original stratified sampling character of the NLSY or the current year weights that adjust these original weights for shifts in the demographic compositions attributable to attrition.

In reaching these conclusions, we must emphasize that the weights currently supplied with the NLSY are inapplicable for use with most sample compositions analyzed in the literature. We construct new weights for use with the cross-sectional sample and for the combination of this sample with the Black and Hispanic supplemental components of the NLSY. We hope to see both of these weights added to future releases of the NLSY data, as well as weights designed for use with samples based on continuously-interviewed individuals. One may be able to devise sophisticated weights that adjust for the sorts of attrition patterns discovered in this study, but much conceptual work needs to be done to ascertain on how to construct such weights.

Appendix A

Sample Compositions and Estimation Procedures Used in Comparisons of the NLSY and CPS

In the calculations of both our figures and estimation results in Section IV, we restrict the sample under investigation to persons not likely to have been in school as of March 12 of each year. We impose this restriction by limiting the ages under investigation; these age restrictions depend upon the number of years of schooling completed and age as of March 12 in the NLSY, and as of the interview date in the March CPS. In particular, for persons reporting 11 or fewer completed years of schooling, we include only observations aged 18 or older as of March 12. For those with 12 years of education we drop observations younger than age 20. For those with 13–15 years of schooling, we restrict the sample to those age 22 and older, and, for those reporting 16 or more years of schooling, we restrict the analysis to persons at least 23 years old as of March 12. We use these education dependent age restrictions even when the analysis does not specify an educational breakdown. Due to the small number of NLSY cohorts which are older than age 30 in 1991, we also restrict the analysis to those persons aged 30 or younger. In addition, we restrict the sample to those aged 14–21 as of March 12, 1979 in the NLSY. For the CPS, we impose the restriction that the person's age in 1979, based upon the age as of the March interview date in each year, would be in the range 14–21 in 1979.

The hours of work measure in the NLSY is the measure constructed by the Center for Human Resource Research as the sum of hours of work from the work histories. In the CPS, annual hours of work is the product of weeks worked last year and usual hours of work per week last year. Our analysis is not restricted to those without self-employment and farm incomes. Since the hours of work measures can include self-employment or farm hours, this average hourly earnings definition is not an hourly wage measure, but it is consistently defined across the two data sets. We define wages as wage and salary income in the prior year divided by annual hours of work, and we only include observations with positive earnings and positive annual hours.

The variables $\mu^n_{\alpha t}$ and $\mu^c_{\alpha t}$ are summary statistics for particular gender, age, and year cells from the two data sources. To reduce the impact of imprecisely measured cell summary statistics, we apply weighted regressions to limit the influence of cells based upon a small number of observations in either the NLSY or the CPS. The justification for these weights comes from an analogy to a simple examination of differences in means from two independent random samples from the same population. Let $n_{n\alpha t}$ be the number of observations in the age-year cell from the NLSY and

$n_{c\alpha t}$ be the number of observations in the age-year cell from the CPS. The variance of the difference in the estimated cell means is $\sigma^2(1/n_{n\alpha t} + 1/n_{c\alpha t})$, and the implied regression weight would be $w_c = 1/(1/n_{n\alpha t} + 1/n_{c\alpha t})$ if all cells had the same population variance.

We use these cell weights to control for the differing sample sizes in the two samples across cells and construct heteroscedasticity-consistent standard errors to control for the across-cell variations in the population variance. Note that this approach does not control for the non-ransom nature of the NLSY and CPS samples. A similar correction for sample sizes would be appropriate for differences in estimated percentiles from the two samples. When examining rates of return to education, the dependent variable is measured as the difference between the difference in two summary statistics from the NLSY and the difference in two summary statistics from the CPS. In an analogous fashion, we define the regression weights as $1/(1/n_{n(1)\alpha t} + 1/n_{n(2)\alpha t} + 1/n_{c(1)\alpha t} + 1/n_{c(2)\alpha t})$, where the terms in parentheses indicate the particular subsample from the NLSY and the CPS.

Most of our comparisons between the NLSY and the CPS examine differences between cell percentiles from the two data sets. We construct these statistics from frequency adjusted percentile estimators. In the NLSY we use the year specific weights to define the frequencies for the adjusted percentiles, and in the CPS we use the March supplemental weights. In those instances when we examine mean differences by cell for the two data sets, we construct weighted means using the same weights as above.