

## THE NLSY79

The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14 to 22 years of age when first surveyed in 1979. During the years since that first interview, these young people typically have finished their schooling, moved out of their parents' homes, made decisions on continuing education and training, entered the labor market, served in the military, married, and started families of their own. Data collected from the NLSY79 respondents chronicle these changes, providing researchers with a unique opportunity to study in detail the life-course experiences of a large group of adults representative of all men and women born in the late 1950s and early 1960s and living in the United States when the survey began.

The cohort was interviewed annually through 1994. Since 1994, the survey has been administered biennially. Interviewing for round 21 of the NLSY79 began in January 2004 and ended in the winter of 2005. This chapter reflects the contents of the round 21 interview, although data for that survey will not be available to the public until early 2006.

Since 1986, detailed information on the development of children born to women in the NLSY79 cohort has supplemented the data on mothers and children collected during the main NLSY79. Chapter 4 describes the biennial surveys of this group, referred to as the "Children of the NLSY79."

### Chapter organization

This chapter supplies an overview of the main NLSY79 cohort and the types of data present within this data set. It provides information about the main and supplemental NLSY79 samples and includes a discussion of multiple-respondent households in the NLSY79, an important feature of the original sampling design. This chapter also gives information on interviews with the NLSY79 men and women, as well as information relevant to assessing how attrition has affected the sample sizes over time. Major data elements for the NLSY79 are then highlighted. The chapter concludes with a brief summary of the data files and documentation available for the NLSY79.

### The NLSY79 samples

The NLSY79 sample design enables researchers to analyze the disparate life course experiences of groups such as women, Hispanics, blacks, and the economically disadvan-

taged. Three subsamples make up the NLSY79:

- A cross-sectional sample of 6,111 youths designed to be representative of noninstitutionalized civilian youths living in the United States in 1979 and born between January 1, 1957, and December 31, 1964 (ages 14 to 21 as of December 31, 1978)
- A supplemental sample of 5,295 youths designed to oversample civilian Hispanic, black, and economically disadvantaged nonblack/non-Hispanic youths living in the United States during 1979 and born between January 1, 1957, and December 31, 1964
- A military sample of 1,280 youths designed to represent the population born between January 1, 1957, and December 31, 1961 (ages 17 to 21 as of December 31, 1978), and enlisted in one of the four branches of the active military forces as of September 30, 1978

All members of the cross-sectional sample and the Hispanic and black supplemental samples have been eligible for interviewing during each round of the NLSY79. Funding constraints limited the numbers of military sample members interviewed after 1984 and of economically disadvantaged nonblack/non-Hispanic supplemental sample members interviewed after 1990. (See table 3.1.)

### Multiple-respondent households

In accordance with the NLSY79 sample design, all eligible individuals ages 14 to 21 who resided in a surveyed household at the end of 1978 were selected as respondents. As a result, the 11,406 civilian respondents interviewed in 1979 originated from 7,490 unique households; 2,862 households included more than one NLSY79 respondent.

At the time the surveys began, the primary types of relationships between respondents living in multiple-respondent households were those of spouse and biological, step-, or adopted sibling. Table 3.2 lists the number of cohort members living in multiple-respondent households during the initial survey by type of relationship.

### The NLSY79 interviews

Interviews with NLSY79 respondents were conducted on an annual basis from 1979 through 1994. After 1994, the survey

was placed on a biennial schedule. The original interview schedule, calling for primarily in-person interviews, was maintained from 1979 to 1986. In 1987, budget constraints dictated that almost 86 percent of respondents participated in a more limited telephone interview, while others were administered the shortened survey in person. Personal interviews were again the dominant type of interview for the 1988 round through the 2000 interview. Telephone efforts again became a major part of the field strategy beginning with the 2002 interview.

From 1979 to 1986, NLSY79 interviews generally were conducted in the first half of the year. Beginning in 1987, the fielding period was shifted to the summer and fall months. This continued until the 2002 survey when the fielding began in January with a telephone effort. Researchers should carefully note interview dates and fielding periods when using variables that may be affected by the time of year of the interview.

In 1993, computer-assisted personal interviewing (CAPI) replaced the traditional paper-and-pencil interview (PAPI)

**Table 3.1. Distribution of NLSY79 respondents by sample type, race, and sex**

Sample type	1979	1984	1990	1994	2000	2002
Total sample	12,686	12,069	10,436	8,891	8,033	7,724
Cross-sectional sample	6,111	5,814	5,498	5,457	4,949	4,775
Males	3,003	2,839	2,664	2,648	2,356	2,270
Nonblack/non-Hispanic	2,439	2,303	2,157	2,150	1,926	1,852
Black	346	329	318	310	272	268
Hispanic	218	207	189	188	158	150
Females	3,108	2,975	2,834	2,809	2,593	2,505
Nonblack/non-Hispanic	2,477	2,365	2,271	2,243	2,065	1,999
Black	405	393	365	363	343	328
Hispanic	226	217	198	203	185	178
Supplemental sample	5,295	5,040	4,755	3,256	2,921	2,792
Males	2,576	2,442	2,280	1,599	1,412	1,353
Economically disadvantaged nonblack/non-Hispanic	742	699	664	1—	1—	1—
Black	1,105	1,055	979	973	847	828
Hispanic	729	688	637	626	565	524
Females	2,719	2,598	2,475	1,657	1,509	1,440
Economically disadvantaged nonblack/non-Hispanic	901	851	819	1—	1—	1—
Black	1,067	1,034	984	987	912	879
Hispanic	751	713	672	670	597	561
Military sample	1,280	1,215	<sup>2</sup> 183	178	163	157
Males	824	774	168	164	152	147
Nonblack/non-Hispanic	609	575	70	68	64	61
Black	162	151	68	62	61	60
Hispanic	53	48	30	34	27	26
Females	456	441	15	14	11	10
Nonblack/non-Hispanic	342	331	7	7	6	6
Black	89	86	5	5	3	3
Hispanic	25	24	3	2	2	1

<sup>1</sup>The economically disadvantaged nonblack/non-Hispanic supplemental sample was dropped after the 1990 interview.

<sup>2</sup>After the military sample was dropped in 1985, 201 sample

members were retained for continued interviewing; 183 of these respondents were interviewed in 1990.

**Table 3.2. Number of NLSY79 civilian respondents by type of multiple-respondent household (1979)**

Type of household	Number of respondents
Sibling	5,914
Two siblings	3,398
Three siblings	1,743
Four siblings	624
Five siblings	125
Six siblings	24
Spouse	334

NOTE: Siblings may be biological, step, or adopted. Some households may include both siblings and spouses, as well as respondents with other relationships not presented in this table.

instruments used during the 1979–92 interviews. Experimental CAPI administrations had taken place during the 1989 and 1990 interviews, with documented improvements in data quality. (See Baker and Bradburn, 1992; Bradburn, Frankel et al., 1992; Olsen, 1991; Speizer and Dougherty, 1991.) For the 2004 interview, Web-based survey instruments were used for the first time in conducting telephone interviews.

Table 3.3 presents the type of survey (personal or telephone) and the completion rate for each year. Table 3.4 lists the number of interviewed respondents by sample type and the resulting retention rates for each NLSY79 round.

### Attrition and sample drops

As table 3.4 shows, retention rates for NLSY79 respondents considered eligible for interview remained close to 90 percent during the first 16 interview rounds and were approximately 85 percent for rounds 17 and 18. Retention rate, expressed as a percentage, is calculated by dividing the number of respondents interviewed by the number of respondents remaining eligible for interview. All base-year respondents, including those reported as deceased, are considered eligible for interview; those sample types that have been permanently dropped from the sample are not considered eligible for interview. In the round 20 (2002) survey, 7,724 civilian and military respondents out of the 9,964 eligible were interviewed, for an overall retention rate of 77.5 percent.

The number of respondents interviewed also can be expressed as a percentage of the number of base-year respondents not known to be deceased. This number, called response rate, is reported for each survey in table 3.5. As of 2002, 346 respondents (more than 3 percent of the respondents eligible for interview) had been reported as deceased. Thus, the response rate for those believed to be alive is 80.9 percent.

Funding constraints required that selected respondents in the military and supplemental subsamples be dropped from interviewing following the 1984 and 1990 surveys, respectively. After the 1984 survey, interviewing ceased for 1,079 members of the military subsample; 201 respondents were retained for continued interviewing and were randomly selected from the entire military sample. After the 1990 survey, all 1,643 economically disadvantaged nonblack/non-Hispanic respondents in the supplemental sample were dropped from the interview rolls.

### Sampling weights

The NLSY79 is a complex longitudinal survey composed of multiple nationally representative samples. Because of this complexity, the staff members working with NLSY79 data create a set of cross-sectional weights for each survey round. Using these weights provides a simple method that analysts may use to correct the raw data for the complex survey design in a particular year. If an analyst is using data from multiple rounds of the survey, the survey weight developed for any particular round will provide an accurate adjustment so long as the test for inclusion also is based data from the same round.

If a research project spans multiple survey rounds, one can create a custom set of survey weights that adjust both for the complexity of the survey design and for the use of data from multiple rounds. The custom weighting program can be found by going to the NLS Web site ([www.bls.gov/nls](http://www.bls.gov/nls)), clicking on the description of the NLSY79 cohort, and then clicking on the link to “create a set of custom weights.”

After the initial round, sampling weights were constructed to adjust for differential nonresponse and for the black, Hispanic, and economically disadvantaged oversamples. These weights provide the researcher with an estimate of the number of individuals in the United States who are represented by each respondent’s answers. The sampling weights are recalculated after each survey to account for noninterviews.

### Major data elements

Each round of the NLSY79 contains core sets of questions on the following topics: Labor market experiences, training investments, schooling, family income, health conditions, geographic residence and environmental characteristics, household composition, and marital and fertility histories. Although the surveys have collected information on these topics on a regular basis, users should note that the number of questions on a given topic, as well as the wording and universe(s) for each question, may differ from year to year.

In selected years, the surveys have included additional sets of questions on a variety of factors potentially affecting labor force attachment. The initial survey, for example, collected information on each respondent’s family background,

Table 3.3. NLSY79 interview schedule

Year	Personal		Telephone		Not available		Interviewed	Not interviewed	
1979	11,863	93.5%	548	4.3%	275	2.2%	12,686	—	—
1980	11,493	94.7	648	5.3	0	—	12,141	545	4.3%
1981	11,541	94.6	654	5.4	0	—	12,195	491	3.9
1982	11,066	91.3	1,054	8.7	3	<sup>1</sup>	12,123	563	4.4
1983	11,897	97.3	324	2.7	0	—	12,221	465	3.7
1984	11,422	94.6	646	5.4	1	<sup>1</sup>	12,069	617	4.9
1985	9,941	91.3	953	8.7	0	—	10,894	<sup>2</sup> 713	6.1
1986	9,726	91.3	929	8.7	0	—	10,655	952	8.2
1987	1,126	10.7	8,998	85.8	361	3.4	10,485	1,122	9.7
1988	9,494	90.7	920	8.8	51	0.5	10,465	1,142	9.8
1989	PAPI: 8,832 CAPI: 252	83.3 2.4	1,469 49	13.9 0.5	3	<sup>1</sup>	10,605	1,002	8.6
1990	PAPI: 6,972 CAPI: 2,145	66.8 20.6	1,032 285	9.9 2.7	2	<sup>1</sup>	10,436	1,171	10.1
1991	7,773	86.2	1,241	13.8	4	<sup>1</sup>	9,018	<sup>3</sup> 946	9.5
1992	7,848	87.0	1,164	12.9	4	<sup>1</sup>	9,016	948	9.5
1993	7,917	87.9	1,081	12.0	13	0.1	9,011	953	9.6
1994	7,948	89.4	933	10.5	10	0.1	8,891	1,073	10.8
1996	7,594	87.9	1,042	12.1	0	—	8,636	1,328	13.3
1998	6,330	75.4	2,069	24.6	0	—	8,399	1,565	15.7
2000	5,420	67.5	2,613	32.5	0	—	8,033	1,931	19.4
2002	2,317	30.0	5,407	70.0	0	—	7,724	2,240	22.5

<sup>1</sup>Less than 0.05%.

<sup>2</sup>The total number of NLSY79 civilian and military respondents eligible for interview beginning in 1985 was 11,607. A total of 201 military respondents were retained from the original sample of 1,280.

<sup>3</sup>The total number of NLSY79 civilian and military respondents eligible for interview beginning in 1991 was 9,964. The 1,643 economically disadvantaged nonblack/non-Hispanic male and fe-

male members of the supplemental subsample were not eligible for interview as of the 1991 survey year.

NOTE: PAPI interviews are those conducted with paper survey instruments and pencil-entered responses; CAPI interviews are administered using a laptop computer and an electronic questionnaire that captures respondent-, interviewer-, and machine-generated data.

his or her knowledge of the world of work, and a detailed retrospective evaluation of the respondent's labor market experience. Subsequent surveys have included questions on subjects such as job search methods, migration, school discipline, health, childcare, self-esteem, time use, delinquency, knowledge about AIDS, attitudes toward work, childhood residences, neighborhood problems, drug and alcohol use, educational/occupational aspirations, and prenatal and postnatal health behaviors.

In addition to these core and supplemental data collections, several special surveys have focused on NLSY79 respondents. These studies include the 1980 high school survey, the 1980–83 high school transcript data collections, and the 1980 administration of the *Armed Services Vocational Aptitude Battery (ASVAB)*.

This section briefly describes major data elements for the NLSY79. The categories encompass information available not only from the NLSY79 main files but also from the

Table 3.4. NLSY79 retention rates by sample types

Year	Cross-sectional sample		Supplemental sample		Military sample		Total sample	
	Total	Retention rate <sup>1</sup>	Total	Retention rate <sup>1</sup>	Total	Retention rate <sup>1</sup>	Total	Retention rate <sup>1</sup>
1979	6,111	—	5,295	—	1,280	—	12,686	—
1980	5,873	96.1	5,075	95.9	1,193	93.2	12,141	95.7
1981	5,892	96.4	5,108	96.5	1,195	93.4	12,195	96.1
1982	5,876	96.2	5,036	95.1	1,211	94.6	12,123	95.6
1983	5,902	96.6	5,093	96.2	1,226	95.8	12,221	96.3
1984	5,814	95.1	5,040	95.2	1,215	94.9	12,069	95.1
1985	5,751	94.1	4,957	93.6	<sup>2</sup> 186	92.5	<sup>2</sup> 10,894	93.9
1986	5,633	92.2	4,839	91.4	183	91.0	10,655	91.8
1987	5,538	90.6	4,768	90.1	179	89.1	10,485	90.3
1988	5,513	90.2	4,777	90.2	175	87.1	10,465	90.2
1989	5,571	91.2	4,853	91.7	181	90.0	10,605	91.4
1990	5,498	90.0	4,755	89.8	183	91.0	10,436	89.9
1991	5,556	90.9	<sup>3</sup> 3,281	89.9	181	90.0	<sup>3</sup> 9,018	90.5
1992	5,553	90.9	3,280	89.8	183	91.0	9,016	90.5
1993	5,537	90.6	3,293	90.2	181	90.0	9,011	90.4
1994	5,457	89.3	3,256	89.2	178	88.6	8,891	89.2
1996	5,290	86.6	3,171	86.8	175	87.1	8,636	86.7
1998	5,159	84.4	3,065	83.9	175	87.1	8,399	84.3
2000	4,949	81.0	2,921	80.0	163	81.1	8,033	80.6
2002	4,775	78.1	2,792	76.5	157	78.1	7,724	77.5

<sup>1</sup> Retention rate is defined as the percentage of base-year respondents remaining eligible who were interviewed in a given survey year; deceased respondents are included in the calculations.

<sup>2</sup> A total of 201 military respondents was retained from the original sample of 1,280; 186 of the 201 participated in the 1985 interview. The total number of NLSY79 civilian and military respondents

eligible for interview beginning in 1985 was 11,607.

<sup>3</sup> The 1,643 economically disadvantaged nonblack/non-Hispanic male and female members of the supplemental subsample were not eligible for interview as of the 1991 survey year. The total number of NLSY79 civilian and military respondents eligible for interview beginning in 1991 was 9,964.

NLSY79 work history and geocode constructed data files. This listing is by no means comprehensive, and not all data elements are necessarily present for all respondents in all survey years. To determine whether the specific variables of interest are available for the year(s) and respondent universe(s) of interest, prospective users should examine several sources. They first should look at the detailed variable table in this chapter, which presents a more comprehensive look at the types of variables available on the NLSY79 main data files. Additional information can be found in the NLSY79 questionnaires, the *NLSY79 User's Guide* (available at [www.bls.gov/nls](http://www.bls.gov/nls)), and the documentation files in the NLSY79 data sets.

Data elements for the NLSY79 are discussed in the order presented in figure 3.1.

### 1. Labor market experiences

For the years 1979 through 2002, information is available for all respondents on their current labor force status (that is, whether the respondent was employed, unemployed, or out of the labor force during most of the week preceding the survey). Those who report being employed supply job- and company-specific data, such as occupation, industry, firm size, availability of fringe benefits, job satisfaction, and number of hours worked each week. For all surveys, detailed job information is available on up to five employers for whom the respondent worked since the last interview. Respondents are asked about the dates on which the job started and stopped, the number of hours worked, the reason why the respondent left the job, details on up to four gaps in employment while associated with an employer, occupation, class

**Table 3.5. NLSY79 response rates (excluding deceased respondents) by sample types**

Year	Cross-sectional sample			Supplemental sample			Military sample			Total sample		
	Total	# deceased	Response rate <sup>1</sup>	Total	# deceased	Response rate <sup>1</sup>	Total	# deceased	Response rate <sup>1</sup>	Total	# deceased	Response rate <sup>1</sup>
1979	6,111	0	—	5,295	0	—	1,280	0	—	12,686	0	—
1980	5,873	4	96.2	5,075	5	95.9	1,193	0	93.2	12,141	9	95.8
1981	5,892	15	96.7	5,108	14	96.7	1,195	0	93.4	12,195	29	96.3
1982	5,876	24	96.5	5,036	19	95.5	1,211	1	94.7	12,123	44	95.9
1983	5,902	27	97.0	5,093	26	96.7	1,226	4	96.1	12,221	57	96.8
1984	5,814	30	95.6	5,040	33	95.8	1,215	4	95.2	12,069	67	95.6
1985	5,751	36	94.7	4,957	43	94.4	<sup>2</sup> 186	0	92.5	<sup>2</sup> 10,894	79	94.5
1986	5,633	43	92.8	4,839	51	92.3	183	1	91.5	10,655	95	92.6
1987	5,538	51	91.4	4,768	56	91.0	179	3	90.4	10,485	110	91.2
1988	5,513	56	91.0	4,777	68	91.4	175	3	88.4	10,465	127	91.2
1989	5,571	60	92.1	4,853	78	93.0	181	3	91.4	10,605	141	92.5
1990	5,498	67	91.0	4,755	82	91.2	183	3	92.4	10,436	152	91.1
1991	5,556	75	92.1	<sup>3</sup> 3,281	65	91.5	181	4	91.9	<sup>3</sup> 9,018	145	91.8
1992	5,553	81	92.1	3,280	71	91.6	182	4	92.4	9,015	156	91.9
1993	5,537	90	92.0	3,293	83	92.3	181	4	91.9	9,011	177	92.1
1994	5,457	104	90.8	3,256	96	91.6	178	4	90.4	8,891	204	91.1
1996	5,290	129	88.4	3,171	109	89.5	175	5	89.3	8,636	243	88.8
1998	5,159	152	86.6	3,065	118	86.7	175	5	89.3	8,399	275	86.7
2000	4,949	170	83.3	2,921	136	83.1	163	7	84.0	8,033	313	83.2
2002	4,775	188	81.2	2,792	151	80.6	157	7	81.6	7,724	346	80.9

<sup>1</sup>Response rate is defined as the percentage of base-year respondents remaining eligible and not known to be deceased who were interviewed in a given survey year.

<sup>2</sup>A total of 201 military respondents was retained from the original sample of 1,280; 186 of the 201 participated in the 1985 interview. The total number of NLSY79 civilian and military respondents eligible for interview (including deceased respondents)

beginning in 1985 was 11,607.

<sup>3</sup>The 1,643 economically disadvantaged nonblack/non-Hispanic male and female members of the supplemental subsample were not eligible for interview as of the 1991 survey year. The total number of NLSY79 civilian and military respondents eligible for interview (including deceased respondents) beginning in 1991 was 9,964.

**Figure 3.1. Data elements in the NLSY79**

1	Labor market experiences
2	Training investments
3	Schooling, school records, and aptitude information
4	Military experience
5	Income and assets
6	Health conditions, injuries, and insurance coverage
7	Alcohol and substance use, criminal behavior
8	Attitudes and aspirations
9	Geographic residence information
10	Family background and demographic characteristics
11	Household composition
12	Marital and fertility histories
13	Child care

of worker, rate of pay, the types of fringe benefits available, and collective bargaining activity in setting wages. In the 1994–2002 surveys, the information collected about each employer was expanded to include the job search method(s) that the respondent used to obtain the job. Also beginning with the 1994 survey, questions were included on whether the respondent participated in a pension plan available through that employer, and whether the respondent received severance pay when leaving the job. The 1994–98 surveys also asked whether the position held with the employer was permanent, temporary, or as an independent contractor.

Beginning in 2002, the employer supplement section of the questionnaire includes questions tailored to self-employed respondents and respondents with nontraditional

employment arrangements such as contract or temporary workers. Similar information is collected for these types of jobs and for regular employment, so that comparability is maintained, but the specialized questions more accurately reflect the job's characteristics.

Also available is information on the activity of the respondent during periods when he or she was not associated with an employer or in the military (that is, weeks not working, weeks spent looking for work, reasons not looking for work). Data are collected on job search behavior for unemployed respondents and on plans to seek employment for those out of the labor force. The surveys also have gathered information on the type and success of various job search methods used by unemployed respondents to find work. The 1979 interview included hypothetical job offer and reservation wage information.

The Center for Human Resource Research (CHRR) at The Ohio State University has constructed a week-by-week longitudinal work record for each respondent from January 1, 1978, through the most recent survey date; this record is present in the work history area of interest (found only on the main file with the work history data set). These weekly data are arranged in three primary arrays:

- A STATUS array of the respondent's labor force/military activity each week, beginning in January 1978
- An HOUR array of the usual hours worked per week at all jobs, beginning in January 1978
- ADUALJOB array, containing up to four additional job numbers for respondents who worked at more than one job simultaneously in any week, beginning in January 1978

Work history variables include information about the number of weeks of a respondent's active military service, the number of weeks that the respondent spent in various labor force states, and detailed information regarding up to five civilian jobs reported by the respondent during each survey year. The respondent identification number permits researchers to link these data with the separate NLSY79 child data.

## 2. Training investments

The surveys have regularly collected detailed information about the types of non-government-sponsored vocational or technical training programs in which a respondent had enrolled since the last interview. In general, the scope of information gathered includes the occupation in which a respondent received training, dates of enrollment and completion, the type of school providing the training, whether the respondent received any subsequent training, the types of certificates or licenses he or she earned, and how the respondent paid for the training.

The 1979–86 surveys asked an extensive series of ques-

tions about government-sponsored training programs in which the respondent had enrolled since the last interview. This series includes program-specific information regarding the respondent's current enrollment status, dates/hours of participation, periods of nonparticipation, and whether the program was part of a federally funded training program such as the Job Training Partnership Act (JTPA)/Comprehensive Employment and Training Act (CETA) or the Work Incentive Program (WIN). These surveys also collected information on the types of occupational or on-the-job training received, the types of classroom training and supportive services provided, and the respondent's rate of pay during training. For jobs identified as government-sponsored, information includes the job's occupational code, the types of classroom training received, and supportive services provided in conjunction with the job and job placement information. Surveys conducted after 1986 posed substantially fewer questions about respondents' participation in government training and employment programs.

In 1993 and 1994, the interviews included a series of questions on (1) the informal methods (for instance, classes/seminars, supervisors, coworkers, self-study) used to learn skills required on the respondent's current job and (2) the potential transferability of skills acquired during various on- and off-the-job training programs. The 1993 survey also fielded a set of high school course relevancy questions. This series included questions on whether the skills learned in a training program added to those the respondent had acquired during high school. For respondents who did not take such high school courses, questions attempted to determine how much of what was learned in the training program could have been learned in high school. Beginning in 1996, questions on informal training on current jobs in the last 4 weeks replaced the questions about informal training used to learn the job. In 2002 and 2004, this series was limited to questions on the use of self-training materials.

*3. Schooling, school records, and aptitude information*  
NLSY79 respondents regularly answer questions about their schooling investments since the date of last interview. Information available includes respondents' current school enrollment status, the highest grade they have attended or completed, whether they have earned a high school diploma or GED equivalency, the type of their high school curriculum, their college enrollment status, their major field of study at college, and the type(s) of college degrees they have earned. The 1994 survey collected retrospective information about respondents' participation in Head Start or another preschool program.

The 1980 high school survey gathered data from the records of the last secondary school attended by NLSY79 respondents; data are available for more than 8,500 respondents. This set of variables includes both respondent-spe-

cific and school-specific information about factors such as (1) the respondent's school enrollment status, highest grade attended, remedial classes taken, and scores for various intelligence and aptitude tests that had been administered during the youth's schooling and (2) each school's total enrollment, grading system, types of curricula offered, dropout rate, student body composition, and staffing characteristics.

Transcript data collected during the 1980–83 survey years are available for more than 9,000 NLSY79 respondents who had completed or were expected to complete high school. Information about as many as 64 high school courses includes course titles, final grades, and credit received. The transcript data collection also provided respondents' attendance records, class rank, and scores on various aptitude and achievement tests.

In addition to the aptitude and intelligence scores collected during the high school and transcript surveys described above, scores from the *ASVAB* are available for 11,914 NLSY79 respondents. These data include individual respondent raw scores, standard scores, scale scores, and standard errors for each of the 10 test sections (for example, general science, arithmetic reasoning, word knowledge, mechanical comprehension). Two Armed Forces Qualifications Test (AFQT) scores also are available in the data files.

#### 4. *Military experience*

The survey provides information regarding the respondent's enlistment intentions, attitudes toward the military, dates of military service/reserve duty, branch of service, military occupation, pay grade, income, education/training received, and reasons for leaving the military or reenlisting. Surveys after 1985 use an abbreviated set of military experience questions.

#### 5. *Income and assets*

The interviews record information about the sources and amounts of income received during the past calendar year by the respondent and his or her spouse or partner. Data on income sources of an opposite-sex partner are available beginning in 1990; if a respondent volunteers such information for a same-sex partner, that information also is recorded. Income sources identified include the respondents' and their spouses' or partners' wages and salaries, income from military service, profits from a farm or business, Social Security, pensions and annuities, and alimony/child support. Surveys between 1993 and 2000 asked child support questions that allow comparisons to be made between amounts of money to which a respondent or spouse was legally entitled and amounts of child support received. Finally, the surveys ask respondents to report the total income received by other family members living in the household. In some surveys, the sources of this income are recorded, but the specific

amounts received from each source are not.

Income received through various government programs also is recorded. The survey asks about monthly income received by the respondent and spouse or partner from unemployment compensation, Aid to Families with Dependent Children/Temporary Assistance to Needy Families (AFDC/TANF), Food Stamps, and other public assistance, including Supplemental Security Income (SSI). In response to the 1996 welfare reform legislation, the 1998 and 2000 surveys included questions about the effect of reform on respondents. In 1998, for example, respondents reported whether they had enrolled in training, looked for a job, or gotten a job to fulfill new welfare requirements, and reported in detail on receipt of targeted benefits. In 2000, a shortened targeted benefits series asked about receipt of job or skills training, clothing assistance, or food assistance other than Food Stamps, WIC, or school meals.

In addition to the information described above, the data include a set of created variables detailing the respondent's history of participation in unemployment compensation (respondent and spouse), AFDC/TANF, SSI, and Food Stamp programs. For each type of reciprocity program, variables include: (1) A monthly dollar value of the benefits, (2) a yearly indicator of receipt or nonreceipt, and (3) a yearly dollar value of the benefits. In addition, there are two yearly summary variables that indicate whether the respondent received any form of public assistance during the year and the total dollar value of all benefits received from AFDC/TANF, SSI, and Food Stamp programs.

Asset and liability information collected during the 1985–90, 1992–2000, and 2004 survey years includes: (1) The types and total market value of property owned by the respondent; (2) the value of other assets, including vehicles and savings accounts; and (3) the total amount of debts owed, including mortgages, back taxes, and other debts over \$500. Although not included in the 2002 survey, asset information was asked in 2004 and will continue to be collected on a regular basis in future rounds.

#### 6. *Health conditions, injuries, and insurance coverage*

The surveys provide data regarding the respondent's height and weight, as well as the presence and duration of health conditions that prevent or limit labor market activity. In some surveys, respondents answer questions about the type(s) of work-related injuries or illnesses they had experienced and their impact on the respondent's work activity. The *International Classification of Diseases* coding system identifies health conditions and work-related injuries reported by respondents during selected survey years. In some years, mothers provide data on prenatal healthcare, infant feeding practices, infant illnesses and treatment, and well-baby care, as well as accidents and injuries to their children. Recent surveys have included questions on the respondent's physi-



cal activity and health behaviors.

Most surveys since 1989 have collected information on whether NLSY79 respondents, their spouses, and their children are covered by a private or governmental healthcare/hospitalization plan and the source(s) of the coverage (current or previous employer, direct purchase, public assistance healthcare program, or other). The 1994 survey asked about specific periods without coverage for respondents, their spouses, and their children; surveys since 1994 have asked about specific periods without coverage only for respondents.

A set of health questions first asked in the 1998 survey established baseline health data for respondents age 40 and older. Each subsequent survey addresses the health module to respondents who have reached age 40 but have not yet answered these questions. This extensive module records general health status, the ways in which respondents' health influences daily activities, and their emotional well-being in the past 4 weeks. Respondents also state whether they have experienced any of a list of potential health problems, such as heart disease, cancer, diabetes, back pain, and so forth. A series about respondents' exercise habits was included in the 2000 survey; this was replaced in 2002 by a short series of questions on general health behaviors. Finally, limited data are collected on the health of the respondents' parents, including their life status, current state of health, major health problems, and cause of death, if deceased.

### 7. Alcohol and substance use, criminal behavior

Alcohol use questions, asked during selected survey years, cover the respondent's consumption of alcohol, his or her frequency of use, quantity usually consumed, whether such use has affected the respondent's schoolwork or job performance, and whether any relatives have alcohol-related problems. Respondents also have provided information on their use of cigarettes, including the age at which they began smoking, how frequently they smoke, and the age at which they stopped smoking, if applicable. Substance use data include the respondent's age at first use; extent of use of marijuana/hashish, amphetamines, barbiturates, cocaine, and heroin; and whether the respondent has used such substances on the job. Female NLSY79 respondents have regularly reported on their use of alcohol, cigarettes, marijuana, or cocaine during pregnancy.

A self-report supplement to the 1980 survey collected data on the respondent's participation in and income from various delinquent and criminal activities such as truancy, alcohol/marijuana use, vandalism, shoplifting, drug dealing, and robbery. A second set of questions in this supplement measured involvement with the criminal justice system, including reported arrest records and police contact.

### 8. Attitudes and aspirations

Attitude variables available for selected survey years include the *Internal-External Locus of Control Scale* (Rotter, 1966), the *Mastery Scale* (Pearlin et al., 1981), the *Self-Esteem Scale* (Rosenberg, 1965), and the *Center for Epidemiologic Studies Depression (CES-D) Scale* (Radloff, 1977). In some years, the survey has included questions on attitudes toward women and work; occupational aspirations; work commitment; knowledge of the world of work; perceived problems in getting a good job; and future expectations about marriage, education, and employment. In the initial survey, respondents reported on the attitude of the most influential person in their life toward certain key career, occupational, residence, and childbearing decisions.

### 9. Geographic residence information

The main NLSY79 data files contain general information on the larger environment in which the respondent lives and works. This information includes details on each respondent's region of residence at birth and at age 14, region of residence at each interview date, whether the current residence is urban or rural, and whether the residence is located in a metropolitan statistical area. The 1992-2000 surveys collected neighborhood characteristics information from NLSY79 women, asking them to rate their neighborhood as a place to raise children and to identify the extent of neighborhood problems such as crime and violence, lack of police protection, neighborhood apathy, and joblessness.

The restricted-release, detailed geocode data files provide additional geographic information about NLSY79 respondents. These files are distributed only to NLS researchers who complete the BLS geocode application process. The geocode file includes information on the respondent's State, county, and metropolitan statistical area of residence; the location of the most recent college that the respondent attended; and selected environmental variables from the U.S. Census Bureau's *County and City Data Books* for the respondent's county or metropolitan statistical area of current residence. More information about how to apply for access to the NLSY79 geocode file is available at [www.bls.gov/nls/geocodeapp.htm](http://www.bls.gov/nls/geocodeapp.htm).

A separate data set, the NLSY79 Women's Support Network data file, offers measures of the geographic proximity of the residences of relatives, friends, and acquaintances to female NLSY79 respondents interviewed during the 1983 through 1985 surveys.

### 10. Family background and demographic characteristics

Selected surveys have collected information on respondents' sex (interviewer-coded), racial/ethnic identification, date of birth, State or country of birth, number of siblings, religious affiliation, childhood residences from birth to age 18, and

immigration/visa status, as well as background information about their parents' birthplace, education, and work experiences. The 1993 survey asked respondents to report the age, education, and fertility of as many as 13 biological siblings. For those respondents identifying a sibling of the same age, follow-up questions in 1994 confirmed whether the paired relationship was that of an identical or fraternal twin, triplet, or quadruplet.

### 11. Household composition

Basic demographic information is available for each person living in the respondent's household at the time of the survey. These data include each resident's sex, relationship to the respondent, age, highest grade of schooling completed, and work experience during the past year.

### 12. Marital and fertility histories

Regularly collected information includes respondents' marital status at the survey date, changes in marital status since the last interview, and dates of each marital status change. Beginning in 2002, information about the respondent's cohabitations since the last interview has been recorded, replacing questions that asked only about cohabitations that were ongoing as of the interview date. Certain surveys have gathered information about the respondents' spouses (and, on a more limited basis, partners), including birth and death dates, occupation, educational attainment, labor force status, religious affiliation, previous marriages, and health limitations. Since 1994, "partner" is generally considered equal to "spouse" for the purposes of administering questions.

Beginning with the 2000 release, the data include a new series of constructed variables for each survey year that provide information about the respondent's relationship status. All previous surveys were examined to match names of spouses and partners, resulting in the creation of two variables for each year. The first variable provides information about the respondent's total number of spouses and partners since 1979 and can range from 0 to 9. The second variable is a unique code assigned to each new spouse/partner at the interview date. If the same spouse or partner resides with the respondent during the next survey round, his or her code remains the same. If the respondent has a new spouse or partner, the next available number is given to that person. If the respondent is reunited with a previous spouse or partner in a later survey round, the total number of spouses/partners is not increased, and the code for that year reflects the recurring spouse/partner's original number. For more details about these created variables, see the *NLSY79 User's Guide*.

Fertility data collected for NLSY79 women include information about all pregnancies resulting in live births, a cumulative roster of all children, residence status of all children, contraceptive methods used, birth expectations and

wantedness information, confidential abortion reports, and ages at menarche and first sexual intercourse. Fertility data collected for men in the NLSY79 cohort are more limited. Information includes birth data for the respondent's children, the residence status of his children, his age at first sexual intercourse, and contraceptive methods used by the respondent. Additionally, a set of questions asked in 1998 and 2000 recorded the frequency of the male respondent's various interactions with his oldest and youngest children.

A set of edited fertility variables added to the main data file (area of interest Fertility and Relationship History/Created) provides revisions to dates of birth, sex, and usual living arrangements for all respondents' children, as well as flag variables indicating the types of revisions made. Constructed variables include the beginning and ending dates of marriages, ages at first marriage and first birth, spacing between births, and spacing between marriage and first birth for the respondent's children. Because they have been corrected, variables in this area of interest are more accurate than the raw data. Users are strongly encouraged to use these edited and created variables. Finally, the database contains a variable evaluating the consistency of each female respondent's longitudinal fertility record between the 1979 and 1982 survey years. Male fertility data is available for download separately and will be included in the data for the next round.

### 13. Childcare

Select surveys provide information about the types of childcare used by female respondents (for instance, care by relatives, care by nonrelatives, daycare center, nursery or preschool, child cares for self), the types of childcare payments incurred, and the number of hours or days that the respondents' children spent in childcare. Retrospective data collected in recent surveys include childcare experiences and arrangements in the first 3 years of life for all children at least 1 year old.

### Data files and documentation

Data for the NLSY79 from 1979–2002 are distributed in two data sets:

- (1) NLSY79 Main/Work History File 1979–2002 data set, which excludes the restricted geocode data but includes the created work history data; and
- (2) NLSY79 Geocode File 1979–2002 CD with all main/work history file data and geocode data.

The combined main/work history data set is available for download from the [www.bls.gov/nls](http://www.bls.gov/nls) Web site.

Each data set includes documentation files and is accompanied by search and extraction software that enables users to easily peruse, select, and extract variables.

Chapter 8 provides more technical information on the NLS data format and extraction software, as well as descriptions of supplementary documentation available for the NLSY79.

### References

- Baker, Reginald P. and Bradburn, Norman M. "CAPI: Impacts on Data Quality and Survey Costs." In: *Proceedings of the 1991 Public Health Conference on Records and Statistics*. Washington, DC: U.S. National Center for Health Statistics (1992): 459–64.
- Bradburn, Norman M.; Frankel, Martin R.; Baker, Reginald P.; and Pergamit, Michael R. "A Comparison of Computer-Assisted Personal Interviews (CAPI) with Paper-and-Pencil (PAPI) Interviews in the National Longitudinal Study of Youth." *NLS Discussion Paper 92-2*. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics, 1992.
- Olsen, Randall J. "The Effects of Computer Assisted Interviewing on Data Quality." Columbus, Ohio: Center for Human Resource Research, The Ohio State University, 1991.
- Pearlin, Leonard I.; Lieberman, Morton A.; Menaghan, Elizabeth G.; and Mullan, Joseph T. "The Stress Process." *Journal of Health and Social Behavior* 22 (December 1981): 337–56.
- Radloff, Lenore Sawyer. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement* 1,3 (Summer 1977): 385–401.
- Rosenberg, Morris. *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press, 1965.
- Rotter, Julian B. "Generalized Expectancies for Internal vs. External Control of Reinforcements." *Psychological Monographs* 80,1 (1966): 1–28.
- Speizer, Howard and Dougherty, Doug. "Automating Data Transmission and Case Management Functions for a Nationwide CAPI Study." In: *Proceedings of the 1991 Annual Research Conference*. Washington, DC: U.S. Census Bureau (1991): 389–97.

### Selected variables by survey year: NLSY79

Table 3.6 depicts selected variables from the NLSY79 main data files. Additional raw and constructed variables on a number of topics (such as geographic-environmental data and detailed individual work histories) can be found on the separately released NLSY79 geocode CD and within the work history area of interest. Detailed information on the Children of the NLSY79 can be found in the various handbooks, user's guides, and evaluation reports that document the child data; more information is available in chapter 4.

The top row of the table lists the years in which the NLSY79 has been administered; the first column gives some of the variables for which information has been collected. An asterisk (\*) indicates that information for the designated variable was obtained for the given cohort in the year indicated. "R" used in variable descriptions stands for "respondent." Users should be aware that, in some instances, the variable descriptions in the table represent a group of questions and not single response items.

**Table 3.6. Selected NLSY79 variables by survey year: Respondents ages 14 to 22 in 1979**

Variable	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	96	98	00	02	04
<b>I. LABOR MARKET EXPERIENCE VARIABLES</b>																					
<b>A. Current labor force and employment status and characteristics of current/most recent job<sup>1</sup></b>																					
Survey week labor force and employment status	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Occupation (DOT code)	*																				
Hours worked in survey week	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Hours per week usually worked	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Shift worked (in 2002/2004 for CPS job only)	*	*	*	*	*	*	*			*	*	*	*	*	*	*	*	*	*	*	*
Promotion (varies with year)						*	*	*	*	*	*	*					*	*	*	*	*
Commuting time to current job	*	*	*							*						*	*				
Availability of fringe benefits (beginning in 1994 for multiple jobs)	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Global job satisfaction item	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Job satisfaction scale	*	*	*	*						*											
Job characteristics inventory	*			*																	
Size of employer (in 2002/2004, only at R's worksite)	*	*						*	*	*	*	*	*	*	*	*	*	*	*	*	*
Minority status of coworkers (1980 only), supervisor		*		*																	
Time R expects to stay at job	*	*	*	*																	
Participation in work-study program	*	*	*	*	*	*	*	*													
<b>B. Work experience since Jan. 1, 1978, or previous survey, or in past calendar year</b>																					
Weeks worked	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Hours usually worked per week	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Number of weeks, spells of unemployment	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Weeks out of labor force	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
<b>C. Characteristics of jobs since Jan. 1, 1978, or last survey, including current or most recent job (more than 10 to 20 hours per week and more than 9 weeks in duration if not CPS job)<sup>2</sup></b>																					
Occupation and industry (Census code)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Class of worker	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Start date and stop date	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Hours usually worked at home										*	*	*	*	*	*	*	*	*	*	*	*
Shift worked (only current/most recent job in 2002/2004)																	*	*	*	*	*
Rate of pay, hourly rate of pay	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Covered by collective bargaining	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

<sup>1</sup> Beginning in 1994, characteristics of the current or most recent job were collected in the first Employer Supplement loop, rather than in the CPS section. To maintain consistency, these questions are still included in this section of the table.

<sup>2</sup> Note that, beginning in 2002, the questionnaire includes separate sets of questions for self-employed respondents and respondents with nontraditional employment arrangements. The information collected is very similar to the regular employment questions, but wordings may vary to accommodate different situations. The three types of employer questions are not represented separately in the table.













**Table 3.6. Selected NLSY79 variables by survey year: Respondents ages 14 to 22 in 1979**

Variable	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	96	98	00	02	04
Future military plans	*	*	*	*	*	*	*														
Reason for entering and leaving military		*	*	*	*	*	*														
Contact with military recruiters	*	*	*	*	*	*	*														
Type of discharge		*	*																		
Enlistment or reenlistment bonuses received	*	*	*	*	*	*	*														
Civilian job offer at time of discharge		*	*	*	*	*	*														
Return to same employer after active duty		*	*	*	*	*	*														
<b>K. Educational and occupational aspirations and expectations</b>																					
Would R like more education or training; type	*																				
How much education desired and actually attained	*		*	*																	
Kind of work R would like to be doing at age 35	*			*																	
Expectation of achieving occupational goal	*			*																	
<b>L. Attitudes</b>																					
Knowledge of World of Work score	*																				
Would R work if had enough money to live on	*																				
Characteristics of job R is willing to take (R unemployed or out of labor force)	*	*	*	*	*	*	*	*	*												
Reaction to hypothetical job offers	*																				
Internal-External Locus of Control Scale (Rotter)	*																				
Mastery Scale (Pearlin)														*							
Attitude toward women working	*			*					*												*
Self-Esteem Scale (Rosenberg) (10 items)		*							*												
CES-Depression Scale (40-and-older health module in 1998–2004)														*		*		*	*	*	*
Person having most influence on R, his or her responses to various situations	*																				
<b>M. Retrospective evaluation of labor market experience</b>																					
Perception of age, race, and sex discrimination	*			*																	
Reason for problems in obtaining employment	*			*																	
<b>N. Delinquency, drugs, and alcohol use</b>																					
Activities within last year (20 items)		*																			
Income from illegal activities within last year		*																			
Alcohol consumption in last week or month				*	*	*	*		*	*			*	*						*	
Extent of cigarette use						*						*	*		*						
Age R first smoked and stopped smoking cigarettes												*	*		*						
Extent of marijuana use		*				*			*			*	*		*			*			
Age R first used marijuana						*			*			*	*		*			*			
Extent of cocaine use, age R first used						*			*			*	*		*			*			

**Table 3.6. Selected NLSY79 variables by survey year: Respondents ages 14 to 22 in 1979**

Variable	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	96	98	00	02	04
Extent of "crack" cocaine use, age R first used														*		*		*			
Ever used sedatives, barbiturates, and so forth						*								*		*		*			
Cigarette and alcohol use during pregnancy					*	*	*	*		*		*		*		*	*	*	*	*	*
Marijuana and cocaine use during pregnancy										*		*		*		*	*	*	*	*	*
<b>O. Reported police contacts</b>																					
Number of times stopped by police		*																			
Number of times booked or arrested		*																			
Number of convictions, charges		*																			
Number of times incarcerated; date of release		*																			
<b>P. Time use</b>																					
Use of time at various activities (school, work, watching TV, household chores, and so forth)			*																		
<b>III. ENVIRONMENTAL VARIABLES</b>																					
Region of residence	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Current residence urban or rural	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Current residence in metropolitan statistical area	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Unemployment rate of local labor market	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Changes in residence since January 1, 1978, or date of last interview (collected as a history)	*	*		*															*	*	*

NOTE: Excluded from this table are the detailed geographic, fertility, and child-specific variables that appear on separate NLSY79 constructed data files.