

Chapter 7.

National Longitudinal Surveys

Background

The National Longitudinal Surveys (NLS) are a set of surveys designed to gather information at multiple points in time on the labor market experiences (for example, employment histories and job characteristics) of six groups of the U. S. population. For more than three decades, NLS data have served as an important tool for economists, sociologists, government analysts, and other researchers who are interested in the study of employment behavior and human capital investments, among other topics.

The NLS were begun in the mid-1960s with the drawing of four samples: *Young men* were 14 to 24 years old as of December 31, 1965; *young women* were 14 to 24 years old as of December 31, 1967; *older men* were 45 to 59 years old as of December 31, 1965; and *mature women* were 30 to 44 years old as of December 31, 1966. Each sample originally had about 5,000 individuals, with an overrepresentation of blacks. In the early 1980s, the surveys of young men and older men were discontinued. (However, there was a followup of the older men or their widows or other family members in 1990.) The two women's surveys continue and are currently collected on a biennial cycle.

In 1979, a new cohort was begun with a sample of more than 12,000 young men and women who were 14 to 21 years old as of December 31, 1978. This sample included oversamples of blacks, Hispanics, economically disadvantaged nonblacks and non-Hispanics, and youths in the military. The military oversample was discontinued after the 1984 survey, and the economically disadvantaged nonblack and non-Hispanic oversample was discontinued in 1990. This survey is called the Youth 1979 data set, or NLSY79; its cohort was interviewed annually from 1979 to 1994 and on a biennial basis thereafter.

In 1997, data collection began for the sixth NLS cohort. The National Longitudinal Survey of Youth 1997 (NLSY97) documents the transition from school to work for a group of approximately 9,000 young people aged 12 to 16 as of December 31, 1996. This cohort is interviewed on an annual basis and includes oversamples of black and Hispanic youths.

Besides collecting information about labor market experiences, the surveys have gathered information regularly about a range of factors potentially affecting labor market attachment, including investments in education and training, local

IN THIS CHAPTER

Background	1
Sampling	1
Questionnaire Design	2
Collection Methods	3
Estimation	4
Uses	4
Presentation and Availability	5
Technical References	5

labor market conditions, parental influence, marital and fertility histories, income and assets, and participation in government assistance programs.

With the advent of the NLSY79 in the late 1970s, the content of the surveys expanded to reflect the interests of other government agencies besides the U.S. Department of Labor. For example, support from the U.S. Department of Defense made possible the administration of the *Armed Services Vocational Aptitude Battery (ASVAB)* to youths in the NLSY79 in 1980 and, later, to the NLSY97 cohort. The National Institute of Child Health and Human Development (NICHD) has provided funds for assessments and surveys of the children of female respondents in the NLSY79 and the development of the fertility and childcare components of the NLSY79. The 1989 National Longitudinal Survey of Mature Women pension-matching project received funding from other agencies in the Department of Labor, as well as from the Social Security Administration and the National Institute on Aging. The National School-to-Work Office (which is jointly funded by the U.S. Departments of Education and Labor) both sponsored the 1996 and 2000 school surveys of the NLSY97 and provided funding for the collection of high school transcript data for the cohort. In addition, sections of the NLSY97 are funded by the U.S. Department of Justice's Office of Juvenile Justice and Delinquency Prevention and by NICHD.

Sampling

Each of the four original NLS samples was designed to represent the civilian noninstitutionalized population of the United

States at the time of the initial survey. Each age-sex cohort is represented by a multistage probability sample drawn by the U.S. Census Bureau from 1,900 primary sampling units (PSUs) that had originally been selected from the Nation's counties and cities for the experimental *Monthly Labor Survey* conducted between early 1964 and late 1966. A PSU consists of Standard Metropolitan Statistical Areas (SMSAs), counties (or parishes in Louisiana), parts of counties (parishes), and independent cities. A total of 235 sample areas comprising 485 counties and independent cities was chosen to represent every State and the District of Columbia. From the sample areas, 235 strata consisting of one or more PSUs that were relatively homogeneous according to socioeconomic characteristics were created. Within each stratum, a single PSU was selected to represent the stratum. Finally, within each PSU, a probability sample of housing units was selected to represent the civilian noninstitutionalized population.

Because one of the survey requirements was to provide separate reliable statistics for blacks, households in predominantly black enumeration districts (EDs) were selected at a rate between 3 and 4 times that for the households in predominantly white EDs. The sample was designed to provide approximately 5,000 interviews for each of the four cohorts—about 1,500 blacks and 3,500 whites.

For the NLSY79 cohort, the following three independent probability samples, which were designed to represent the entire population of youths born between 1957 and 1964 and living in the United States in 1978, were drawn: (1) A cross-sectional sample designed to be representative of the noninstitutionalized civilian segment of young people aged 14 to 21 years as of December 31, 1978; (2) a supplemental sample designed to overrepresent civilian Hispanic, black, and economically disadvantaged non-Hispanic, nonblack youths; and (3) a military sample designed to represent the population aged 17 to 21 as of December 31, 1978, and serving in the military as of September 30, 1978.

All sample selection was done through a multistage stratified area probability sample of dwelling units and group-quarters units, except for individuals on active military duty. A screening interview was administered in approximately 75,000 dwellings and group quarters distributed among 1,818 sample segments in 202 PSUs (inclusive of most of the 50 States and the District of Columbia). As part of this screening interview, information was obtained that would allow the identification of persons eligible for membership in the sample.

Members on active military duty as of September 30, 1978, were sampled from rosters provided by the Department of Defense. Sample selection was accomplished in two stages. In the first stage, a sample of approximately 200 "military units" was selected, with probabilities proportional to the number of persons aged 17 to 21 years within the unit. Within selected units, persons aged 17 to 21 years were subsampled with probabilities inversely proportional to the first-stage selection probability. Females were oversampled at a rate approximately 6 times that of males, in order to produce ap-

proximately equal numbers of males and females. Within each sex, the sample was stratified on the basis of military service (Army, Navy, Air Force, and Marine Corps) and geographic location.

The NLSY97 cohort consists of two independently selected, stratified, multistage area probability samples designed to represent noninstitutionalized youths born between 1980 and 1984 and living in the United States in 1997: A cross-sectional sample and an oversample of black and Hispanic youths. For each probability sample, 100 PSUs were selected to represent all 50 States and the District of Columbia. For the cross-sectional sample, the probability of selection for each PSU was proportional to the 1990 census count of its housing units. For the supplemental sample, probabilities of selection for PSUs and segments were based on 1990 census counts of blacks and Hispanics aged 17 years and younger. In addition, areas that contain high percentages of black and Hispanic residents were oversampled. Only eligible black and Hispanic youths were screened into the NLSY97 sample from the supplemental areas.

Each of the four original cohorts was interviewed initially between 1966 and 1968. Respondents for the NLSY79 were first interviewed in 1979, and interviewing for the NLSY97 began in 1997. About 90 percent of the individuals designated for interviewing responded to each of the first-year interviews: 5,020 (91 percent) of the older men and 5,225 (92 percent) of the young men were interviewed in 1966; 5,083 (94 percent) of the mature women were interviewed in 1967; 5,159 (93 percent) of the designated young women were interviewed in 1968; 12,686 (90 percent) of the NLSY79 were interviewed in 1979; and, finally, 8,984 (92 percent) of the NLSY97 cohort were interviewed in 1997. Completion rates for the initial and latest survey years of each of the cohorts are summarized in table 1.

Questionnaire Design

The development of a questionnaire involves a review of earlier questions, the analysis of field notes from the previous round, and the identification of new topics and questions for inclusion in the current survey. The development of each survey instrument generally is begun at least 2 years prior to the fielding. Each of the ongoing surveys has its own design team made up of Bureau of Labor Statistics staff, NLS survey contractors, and outside experts. Advice from the NLS Technical Review Committee and other Government agencies is often sought regarding survey questionnaire initiatives and other survey-related issues.

The central focus of each cohort's survey is, to some extent, determined by the particular stage of labor market attachment that each of these six unique age-sex groups is experiencing at the time of the interview. Each survey instrument is organized around core sets of questions on the following topics: Employment, education, training, marital status, fertility, participation in government assistance programs, income, and assets.

Table 1. Completion rates by NLS cohort for the initial and latest survey years

NLS cohort	Designated for interviewing	Number and percent interviewed in initial survey year			Number and percent interviewed in last or latest survey year		
		Number	Percent	Year	Number	Percent	Year
Older men	5,518	5,020	91	1966	2,092	42	¹ 1990
Mature women	5,393	5,083	94	1967	2,318	46	2001
Young men	5,713	5,225	92	1966	3,398	65	1981
Young women	5,533	5,159	93	1968	2,806	54	2001
Youth 1979	14,574	² 12,686	87	1979	8,033	81	2000
Cross-sectional sample	6,812	6,111	90		4,949	81	
Supplemental sample	5,969	5,295	89		2,921	80	
Military sample	1,793	1,280	72		163	81	
Youth 1997	9,806	8,984	92	1997	7,883	88	2001
Cross-sectional sample	7,327	6,748	92		5,919	88	
Supplemental sample	2,479	2,236	90		1,964	88	

¹In 1990, interviews also were conducted with 2,206 widows or other family members of deceased respondents.

²After the majority of the military sample was dropped in 1985

and the economically disadvantaged nonblack, non-Hispanic oversamples were eliminated in 1991, the sample was left with 9,964 respondents eligible for interview.

Each cohort also is asked questions that are focused on its stage in the life cycle. For example, the surveys of older men have focused on plans for their future—specifically, retirement, pension plan participation, and health. Special topics for the mature women’s cohort have included questions on household and volunteer activities, childcare, care of parents, plans for retirement, and pension plan participation. The surveys of both the young men’s and young women’s cohorts have focused on educational goals, high school and college experiences, characteristics of their high schools, and future job plans. The young women’s cohort is ongoing, and, as the women are aging, the focus is shifting to retirement issues and transfers of time and money to and from their parents and their children.

The NLSY79 cohort has been interviewed periodically for over 20 years. When sample members were younger, special topics included their family background, knowledge of the world of work, educational and occupational aspirations, and expectations. As the respondents aged, special groups of questions were asked about pre- and postnatal health behaviors, job search, migration, and drug and alcohol use, among other topics. The changing life stage of the NLSY79 cohort prompted the NLS program to hold an NLSY79 Redesign Conference in 1998. The goal was to evaluate current questions in key areas of the survey, identify areas for improvement, and provide direction for future rounds as the NLSY79 passes into its next life stage. The NLS program commissioned papers by experts in key subject areas to provide this information and received many useful recommendations therefrom.

Eight development teams, each composed of recognized experts in critical content areas, contributed to the design of the initial NLSY97 survey. Lessons learned from the NLSY79

and other NLS cohorts were incorporated into the design of the NLSY97 survey instrument. The NLSY97 design team meets quarterly to discuss changes that might improve each round’s questionnaire to capture youths’ transitions from school to work and into adulthood. Among the many innovations in the NLSY97 are the collection of information on the work experience of the very young in freelance jobs, the gathering of data on delinquent and criminal behaviors, and the accumulation of extensive information on high school and college experiences.

Collection Methods

Respondents selected for interviewing each year, with the exceptions noted next, are those who had participated in the initial-year interviews and who were alive on the date of the interview.

For the original cohorts, subsequent to the first-year interview, those respondents who had refused to be interviewed were dropped from the sample. Beginning with the third interview year, respondents who had not been interviewed for any reason for 2 consecutive years also were eliminated from the sample. However, this noninterview exclusion was not applied to those members of the young men’s cohort who were subsequently inducted into the Armed Forces. No interviews were attempted with that group while its members were on active military duty. They were, however, retained in the sample and interviewed as soon as they left active military service. Beginning in the mid-1980s, the procedures for selecting respondents for interviewing were modified, and refusals and those not interviewed for 2 survey years were not dropped.

In the NLSY79 survey, attempts are made to reach all individuals within the active samples. No respondents have been routinely excluded from locator efforts, with the exception of respondents who have died. The permanent NLSY79 sample designated for interviewing during the 1979–84 interview years consisted of all civilian and military youths who were interviewed in the base year and who were alive on the survey date. In 1985, when interviewing of the full military sample ceased, the total NLSY79 sample size dropped from 12,686 to 11,607. Retained for interviewing were the original 11,406 civilian respondents, as well as 201 military respondents. After the 1990 survey, interviews with the economically disadvantaged nonblack, non-Hispanic oversample were discontinued, reducing the NLSY79 sample size to 9,964.

In the new youth cohort (the NLSY97), surveyors attempt to interview the full sample of 8,984 respondents in each round, regardless of whether they have been interviewed in previous rounds.

Although personal interviews have been the method used during most survey years, at times telephone interviews also were conducted. In addition, interviews by mail were conducted for the 1968 older men and mature women surveys. Telephone contact may occur in cases where the respondent resides in a remote area or the field staff determines that the respondent prefers to be interviewed by telephone.

Originally, the NLS interviews were conducted with paper and pencil. Interviewers filled in the relevant fields in long questionnaires, and, as a result, the interviewing process was slow and burdensome. With the advent of newer, faster computers, the NLS switched to computer-assisted personal interviews (CAPI), for which interviewers take laptop computers instead of questionnaire booklets into the field. A computer program automatically selects the next question, prevents interviewers from entering invalid values, and warns interviewers about implausible answers, all with an eye toward increasing the quality of the data collected. The first cohort to begin using CAPI was the NLSY79 in 1993. The women's surveys were switched over to CAPI in 1995. The newest NLS cohort, the NLSY97, has used only CAPI. In addition, it contains an audio computer-assisted self-interview (ACASI), in order to obtain potentially sensitive information from respondents, such as information on their drug and alcohol use and criminal behavior. The respondent is able to listen to the questions with earphones or read the questions on the screen, or both, and then enter his or her response.

Estimation

The NLS surveys are based upon stratified multistage random samples, with oversamples of blacks in all cohorts, oversamples of Hispanics in the NLSY79 and NLSY97, and additional oversamples of disadvantaged nonblack non-Hispanics and youths in the military in the NLSY79. Data from each interview year include a weight specific to that year. When this weight is applied, the number of sample cases is translated into the number of persons in the population that

those observations represent.

The assignment of individual respondent weights involves at least three stages. The first stage involves the reciprocal of the probability of selection at the baseline interview. Specifically, this probability of selection is a function of the probability of selection associated with the household in which the respondent was located, as well as the subsampling (if any) applied to individuals identified in screening. The second stage of weighting adjusts for differential response (cooperation) rates in the screening phase. Differential cooperation rates are computed (and adjusted) on the basis of geographic location and group membership, as well as by group subclassification. The third stage of weighting attempts to adjust for certain types of random variation associated with sampling, as well as sample "undercoverage." The estimated ratios are used to conform the sample to independently derived population totals.

Subsequent to the initial interview of each cohort, reductions in sample size have occurred due to noninterviews (the failure, for one reason or another, of the person to be interviewed). In order to compensate for these losses, the sampling weights of the individuals who were interviewed had to be revised. A revised weight for each respondent was calculated for each interview year, using the method just described.

In the event that one wishes to tabulate characteristics of the sample for a single interview year in order to describe the population being represented, it is necessary to weight the observations by using the weights provided. For example, to compute the average hours worked in 1987 by individuals in the NLSY79 (persons born in 1957–1964 and living in the United States in 1978), one simply weights the average hours worked by the 1987 sample weight. The weights are correct when used in this way.

Often, users confine their analyses to subsamples for which respondents provide valid answers to certain questions. Weighted means here will represent, not the entire population, but rather those persons in the population who would have given a valid response to the specified questions. Nonresponse to any item because of refusals or invalid skips is generally quite small, so the degree to which the weights are incorrect also is probably quite small. In these instances, although the population estimates may be moderately in error, the population distributions (including means, medians, and proportions) are reasonably accurate. Exceptions to this assumption might occur for data items that have relatively high nonresponse rates, such as family income.

Uses

NLS data are important tools for economists, sociologists, psychologists, and other researchers in the study of labor supply, earnings and income distribution, job search and separation, training, and other human capital investments. In addition, these data are used to study the effect of government policies and programs on labor market behaviors.

Several comprehensive reviews of NLS research (Bielby, Hawley, and Bils, 1979; Daymont and Andrisani, 1983) and the *NLS Annotated Bibliography* (which provides citations and abstracts of thousands of publications, dissertations, and other research papers that have used NLS data over the years) summarize much of the work that has been generated by the NLS since the mid-1960s. A recent article in the *Journal of Economic Perspectives* focuses on significant areas and corresponding research done with the NLSY79 (Pergamit and others, 2001). The continued relevance of the NLS for policymakers at the Federal, State, and local levels, as well as for the research community, is summarized in *The Future of the NLS: A Report from the NSF Conference on the Future of the NLS and the NLS Technical Review Committee* (Center for Human Resource Research, 1989). A summary of some uses of the NLS made by the Federal Government can be found in *How the Federal Government Uses Data from the National Longitudinal Surveys* (Pergamit, 1991).

The broad range of core NLS data, coupled with the extensive expansion of the youth surveys, the ongoing longitudinal nature of the data, and the replication of cohorts across time, helps to make the NLS a rich and yet-to-be fully

exploited source of data for the continued study of such issues as labor market behavior, life-cycle changes, the family, the aging process, retirement decisions, and geographic mobility, as well as a host of other topics and methodologies.

Presentation and Availability

Extensive information about the NLS can be found on the NLS Web site, <http://www.bls.gov/nls>, which houses BLS press releases on recent findings about employment from the NLSY79 and NLSY97 surveys and provides links to *Monthly Labor Review* articles that use NLS data, the *NLS Discussion Paper* series, and the *NLS Annotated Bibliography*. The site also provides links to the extensive NLS documentation available, such as the *NLS Handbook* and cohort-specific *NLS User's Guides*.

One of the great strengths of the NLS program is the ease with which its data can be accessed. NLS data are on cohort-specific compact discs, complete with user-friendly search-and-retrieval software. Check the NLS Web site for information on how to order NLS data files. In addition, data are now available for downloading from the NLS Web site.

Technical References

Bielby, William T., Clifford B. Hawley, and David Bils. "Research Uses of the National Longitudinal Surveys," *R&D Monograph 62* (U.S. Department of Labor, Employment and Training Administration, 1979).

Bureau of Labor Statistics. *NLS Handbook* (Bureau of Labor Statistics, 2003).

Center for Human Resource Research. *The Future of the NLS: A Report from the NSF Conference on the Future of the NLS and the NLS Technical Review Committee* (Columbus, The Ohio State University, 1989).

_____. *NLS Cohort-Specific User's Guides* (Columbus, The Ohio State University, 2002).

Daymont, Thomas N. and Paul J. Andrisani. "The Research Uses of the National Longitudinal Surveys: An Update," *Review of Public Data Use* (U.S. Department of Commerce, 1983).

Pergamit, Michael. *How the Federal Government Uses Data from the National Longitudinal Surveys*, Report NLS 92-1 (Bureau of Labor Statistics, 1991).

Pergamit, Michael R., Charles R. Pierret, Donna S. Rothstein, and Jonathan R. Veum. "Data Watch: The National Longitudinal Surveys," *Journal of Economic Perspectives*, vol. 15, no. 1, spring 2001, pp. 239–53.