

National Longitudinal Survey of Youth 1979
1220-0109
January 2018

**Information Collection Request for
The National Longitudinal Survey of Youth 1979
OMB # 1220-0109
Part B**

Submitted by the Bureau of Labor Statistics

TABLE OF CONTENTS

<i>B. Collections of Information Employing Statistical Methods</i>	3
1. Respondent Universe and Respondent Selection Method.....	3
2. Design and Procedures for the Information Collection.....	4
3. Maximizing Response Rates.....	6
4. Testing of Questionnaire Items.....	7
5. Statistical Consultant.....	8

B. Collections of Information Employing Statistical Methods

1. Respondent Universe and Respondent Selection Method

The initial NLSY79 sample was selected to represent (after appropriate weighting) the U.S. civilian and military population of 33,570,000 persons who were ages 14 to 21 as of December 31, 1978. The sample selection procedure included an overrepresentation of Hispanic and black youths so as to include sufficient sample cases to permit racial and ethnic analytical comparisons. Economically disadvantaged nonblack/non-Hispanic youths also were oversampled.

The NLSY79 originally included a supplemental sample of youths in the military. In 1985, the military supplemental sample was discontinued. In 1991, the economically disadvantaged nonblack/non-Hispanic oversample was discontinued. Appropriate weights have been developed so that the sample components can be combined in a manner to aggregate to the overall U.S. population born in the years 1957-64 and living in the United States when the sample was selected in 1978. The number of sample cases in 1979, excluding the discontinued military and nonblack/non-Hispanic samples, was 9,964. A breakdown by sex and race is depicted in table 5 below. We anticipate a response rate in Round 28 that is similar to the Round 27 experience.

Table 5. Interviews Completed in 1979 and 2016 (Preliminary) by Race and Sex

Race and Hispanic origin	1979			2016					
	Number of men	Number of women	Total number	Number of men	Retention rate for men	Number of women	Retention rate for women	Total number	Total retention rate
Nonblack, non-Hispanic	2,518	2,484	5,002	1,649	65.5	1,758	70.1	3,407	68.1
Hispanic	981	980	1,961	628	64.0	684	69.8	1,312	66.9
Black	1,524	1,477	3,001	1,048	68.8	1,144	77.5	2,192	73.0
Total	5,023	4,941	9,964	3,325	66.2	3,586	72.6	6,911	69.4

Retention rates for the NLSY79 are significantly affected by attrition due to death. Approximately 9.1% of the 9,964 NLSY79 respondents still eligible for interviewing were deceased after the 2016 survey; we are investigating ways to determine whether some respondents we cannot locate may be deceased. Table 6 provides information about retention (percent of all respondents interviewed) and response (percent of living respondents interviewed) rates for each year of the NLSY79.

Table 6. NLSY79 retention and response rates by sample type

Year	Number Interviewed	Retention Rate ¹	Number of Deceased Respondents	Response Rate ¹
1979	12,686	—	—	—
1980	12,141	95.7	9	95.8
1981	12,195	96.1	29	96.3
1982	12,123	95.6	44	95.9
1983	12,221	96.3	57	96.8
1984	12,069	95.1	67	95.6
1985	² 10,894	93.9	79	94.5
1986	10,655	91.8	95	92.6
1987	10,485	90.3	110	91.2
1988	10,465	90.2	127	91.2
1989	10,605	91.4	141	92.5
1990	10,436	89.9	152	91.1
1991	³ 9,018	90.5	144	91.8
1992	9,016	90.5	156	91.9
1993	9,011	90.4	177	92.1
1994	8,891	89.2	204	91.1
1996	8,636	86.7	243	88.8
1998	8,399	84.3	275	86.7
2000	8,033	80.6	313	83.2
2002	7,724	77.5	346	80.3
2004	7,661	76.9	399	80.1
2006	7,654	76.8	456	80.5
2008	7,757	77.9	503	82.0
2010	7,565	75.9	573	80.6
2012	7,301	73.3	689	78.7
2014	7,066	70.9	790	77.0
2016	⁴ 6,911	69.4	⁴ 909	76.3

¹ 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. 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.

² A total of 201 military respondents were 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.

³ 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.

⁴ Preliminary.

2. Design and Procedures for the Information Collection

The survey includes telephone or personal-visit interviews with all the respondents, regardless of their place of residence. At each interview, detailed information is gathered about relatives and friends who could be of assistance in locating the sample member if he or she cannot be readily

located in the subsequent survey round. Most interviews in Round 28 will be carried out between October 2018 and October 2019, with the field period extending later into 2019 if necessary. Every effort is made to locate respondents, as the attrition information above suggests. Interviewers are encouraged to attempt contacting respondents until they are located. There is no arbitrary limit on the number of callbacks. The success of NORC interviewers in this regard is indicated by a very low rate of attrition over the first 27 rounds of the survey. More than 76 percent of the living, in-scope, original respondents were surveyed in Round 27.

Preceding the data collection, interviewers are carefully trained, with particular emphasis placed on resolving sensitive issues. Most of the interviewers have lengthy experience in the field from participation in earlier NLSY79 interview rounds, as well as from involvement with other NORC surveys. Experienced interviewers receive self-study training consisting of over 8 hours spent on specially designed materials requiring study of the questionnaire and question-by-question and procedural specifications, with exercises on new or difficult sections and procedures. Experienced interviewers working on the Early Bird phone phase also receive several hours of remote training (by computer and conference call), mainly focused on using the call management system. All interviewers must successfully complete a practice interview with their supervisor before they are permitted to begin field work.

Efforts to assure quality data from the field are instigated at several points. The first 100 cases completed are reviewed, answer by answer, to determine whether there are any problems with the instrument. After this, every case identified by the interviewer as having a problem during the interview is reviewed in detail. Throughout the field period, individual cases are checked for problems, and rapid feedback is given to the interviewers so they can improve interviewing methods.

We will reduce burden by employing targeted validation. Cases that have unusual patterns in terms of length, time of day, break-offs, an incorrect entry to the question on the respondent's date of birth, or height and weight entries that are inconsistent with previous rounds will be validated. We also use review of recordings for validation purposes. If a case fails to validate, the entire caseload of the interviewer will be validated through review of recordings or validation re-interviews.

The questionnaires are prepared by professional staff at CHRR. When new materials are incorporated into the schedule, special assistance is generally sought from appropriate experts in the specific substantive area. The technical expertise of staff at NORC is also used in this regard.

Because sample selection took place in 1978 in preparation for the 1979 baseline interview, sample composition has remained unchanged except for the discontinuation of some of the oversamples as previously mentioned. A more detailed discussion of sampling methodology is available from the NLSY79 Technical Sampling Report at:
<http://www.nlsinfo.org/content/cohorts/nlsy79/other-documentation/technical-sampling-report>.

In an effort to reduce respondent burden while still providing a broad spectrum of variables for researchers and policymakers to use, certain topical modules are cycled in and out of the survey from one round to the next. Although the data from these modules are important, it is not necessary to collect data on all topics in every round. An example of such a topical module is the assets module, now asked every other round.

3. Maximizing Response Rates

A number of the procedures used to maximize the response rate already have been described in items 1 and 2 above. The success of the procedures is demonstrated by the low attrition rates indicated in tables 5 and 6. Hispanic attrition has been slightly higher than among blacks and nonblack/non-Hispanics, and attrition for men is higher than for women. Attachment 7 analyzes the patterns of non-response and its implications on the estimates of labor market outcomes. Clearly, it is more difficult to gain the cooperation of some respondents. Past non-response indicates a lower probability of participation in subsequent rounds. Still, there is little evidence that any selective response biases the analytical results. To the best of our knowledge, the NLSY79 has the best retention rate of any longitudinal survey in the U.S. We note, however, that interviewing becomes a little more difficult each round.

The other component of missing data is item nonresponse. The rate of item nonresponse due to the refusal of a respondent to answer a particular question in the NLSY79 is between 1 and 2 percent per question, depending on how questions are counted. The highest nonresponse rates occur for income and asset items.

One natural issue for longitudinal surveys is to determine whether the sample still represents its portion of the U.S. population. The NLSY79 originally was weighted to represent the 1978 population of 14-21 year-olds and closely matches the official statistics for that year. Sampling weights are prepared each year to adjust the remaining sample to representative proportions. These sampling weights are released with the other data on the public-use data file.

To investigate the issue of continued sample representation, table 7 compares numbers from the 2010 decennial census with NLSY79 population estimates. Census data are taken from the [American FactFinder](#) page of the Census Bureau website.^[1] The American FactFinder website provides the number of people living in the United States, who were ages 46 to 53 on April 1, 2010, which is the same age group that the NLSY79 sample represents. NLSY79 population estimates are from the weighted results of the Round 24 (year 2010) survey.

Table 7 shows the percentage of the NLSY79 sample and U.S. population by sex and race. Overall, the table has two significant features. First, the 2010 NLSY79 sample slightly overrepresents men, since there is a larger percentage of men in the NLSY79 sample than in the U.S. population. For comparison, the original NLSY79 sample in 1979 was composed of 50.8 percent men and 49.2 percent women, while the Census Bureau reported a 1979 population for the same age group that was 50.6 percent men and 49.4 percent women. (Source: U.S. Bureau of the Census, publication P25-917, *Preliminary Estimates of the Population of the United States, by Age, Sex, and Race: 1970 to 1981*; and 1979 NLSY79 data) Hence, the current male-female composition of the NLSY79 does not suggest any gender-biased attrition. Rather, the composition is similar to the sex ratio of the original panel. Nevertheless, the comparison with the recent Census Bureau estimates may suggest that the NLSY79 sample is experiencing lower male mortality than the overall U.S. population of the same age.

^[1] Three series were extracted from the American FactFinder website; PCT12, PCT12J, and PCT12H which track the U.S. population's age by sex and race/ethnicity. PCT12 tracks the total population, PCT12J tracks black non-Hispanics and PCT12H tracks the Hispanic population. Nonblack, non-Hispanic figures are computed by subtracting PCT12J and PCT12H from PCT12. All data come from the Census 2000 Summary File 1 (SF 1), which combines both short and long form answers and does not subsample any information.

Second, the NLSY79 sample underrepresents the current U.S. population of Hispanics. The NLSY79 sample does not include persons who entered the United States after 1978, and the rate of immigration among Hispanics has been very high since the NLSY79 sample was selected. (The differences between the NLSY79 sample and the recent U.S. population estimates in the percentages of nonblack/non-Hispanics and black non-Hispanics can be explained largely by the shortfall of Hispanics in the NLSY79 sample.) Comparing the NLSY79 sample with the U.S. population estimates for 1978, the NLSY79 sample correctly represents the Hispanic population on a weighted basis, and as described earlier in this document, the NLSY79 sample intentionally *overrepresents* the 1978 Hispanic population on an unweighted basis.

Overall, table 7 shows that, except for the Hispanic population, the NLSY79 sample still is similar to the U.S. population estimates for the same age group. If one accounts for the large amount of Hispanic immigration since the survey began, the remaining differences are not large. Moreover, the weights that are produced after each round compensate for the modestly different rates of attrition and mortality across demographic groups.

Table 7. NLSY79 Weighted Sample Composition in 2010 versus U.S. Census Data for Persons Ages 43 to 56 as of April 1, 2010

	2010 NLSY79	Census Data
Total	100.0%	100.0%
Men	50.9%	49.2%
Women	49.1%	50.8%
Nonblack, non-Hispanic	79.3%	75.9%
Men	40.2%	37.5%
Women	39.1%	38.4%
Black, non-Hispanic	14.2%	12.0%
Men	7.3%	5.7%
Women	6.9%	6.4%
Hispanic	6.5%	12.0%
Men	3.4%	6.0%
Women	3.1%	6.0%

4. Testing of Questionnaire Items

BLS is cautious about adding items to the NLSY79 questionnaire. Because the survey is longitudinal, poorly designed questions can result in flawed data and lost opportunities to capture contemporaneous information about important events in respondents' lives. Poorly designed questions also can cause respondents to react negatively, making their future cooperation less likely. Thus, the NLSY79 design process employs a multi-tiered approach to the testing and review of questionnaire items.

When new items are proposed for the NLSY79 questionnaire, we often adopt questions that have been used previously in probability sample surveys. We have favored questions from the other surveys in the BLS National Longitudinal Surveys program to facilitate intergenerational comparisons. We also have used items from the Current Population Survey, the Federal Reserve Board's Survey of Consumer Finances, the National Science Foundation-funded General Social Survey, the National Institute of Aging-funded Health and Retirement Study and Midlife in the United States, and other Federally funded surveys.

New questions are reviewed in their proposed NLSY79 context by survey methodologists who consider the appropriateness of questions (reference period, terms and definitions used, sensitivity,

National Longitudinal Survey of Youth 1979
1220-0109
January 2018

and so forth). Questions that are not well-tested with NLSY79-type respondents undergo cognitive testing with convenience samples of respondents similar to the NLSY79 sample members.

Existing questions are also reviewed each year. Respondents' age and their life circumstances change, as does the societal environment in which the survey is conducted. Reviews of the data help us to identify questions that may cause respondent confusion, require revised response categories, or generate questionable data. Sources of information for these reviews include the questionnaire response data themselves, comments made by interviewers or respondents during the course of the interview, interviewer remarks after the interview, interviewer inquiries or comments throughout the course of data collection, other-specify coding, and comparison of NLSY79 response data to other sources for external validation. We also watch carefully the "leading edge" respondents, who answer some questions before the bulk of the sample – for example, the first respondents to attend graduate school, to get a divorce, or to retire from the labor force. These respondents are often atypical, but their interviews can reveal problems in question functionality or comprehensibility.

Although further edits to questionnaire wording are extremely rare, we monitor the first several hundred interviews each round with particular care. Based on this monitoring, field interviewers receive supplemental training on how best to administer questions that seem to be causing difficulty in the field or generating unexpected discrepancies in the data.

5. Statistical Consultant

Kirk M. Wolter
Senior Fellow and Director
Center for Excellence in Survey Research
NORC