SUPPORTING STATEMENT FOR
National Longitudinal Survey of Youth 1979

OMB CONTROL NO. 1220-0109

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

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 oversampled Hispanic and Black youths to include sufficient sample cases to permit racial and ethnic analytical comparisons. Economically disadvantaged non-Black/non-Hispanic youths also were oversampled.

The NLSY79 originally included oversampled youths in the military. In 1985, the military oversample was discontinued. In 1991, the economically disadvantaged non-Black/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 interviewed in 1979, excluding the discontinued military and economically disadvantaged non-Black/non-Hispanic oversamples, was 9,964. A breakdown by sex and race is depicted in Table 10 below. NLS anticipates a response rate in Round 31 that is similar to the Round 30 experience.

Table 10. Interviews Completed in 1979 and 2022 (Preliminary) by Race and Sex

Race and Hispanic origin	1979			2022 (Round 30)					
	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
Non-Black, non-Hispanic	2,518	2,484	5,002	1,505	59.8	1,666	66.2	3,171	63.4
Hispanic	981	980	1,961	567	57.8	676	68.9	1,243	63.4
Black	1,524	1,477	3,001	939	61.6	1,053	69.1	1,992	66.4
Total	5,023	4,941	9,964	3,011	59.9	3,395	67.6	6,406	64.3

Retention rates for the NLSY79 are significantly affected by attrition due to death. Approximately 12.2% of the 9,964 NLSY79 respondents still eligible for interviewing were deceased after the 2020 survey. NLS currently is making refinements to this count through a match to administrative data, which NLS expects to have a minor impact. Table 11 provides

information about retention (percent of base year respondents interviewed) and response (percent of living base year respondents interviewed) rates for each year of the NLSY79.

Table 11. NLSY79 retention and response rates by sample type

Year	Number Interviewed	Retention Rate[1]	Number Deceased	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	6,912	69.4	915	76.4
2018	6,878	69.0	1,033	77.0
2020	6,535	65.6	1,185	74.4
2022	6,413	64.4	1,347	74.4

[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 denominator of 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 non-Black/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.

2. Describe the procedures for the collection of information including:

- **Statistical methodology for stratification and sample selection,**

- **Estimation procedure,**

- **Degree of accuracy needed for the purpose described in the justification,**

- **Unusual problems requiring specialized sampling procedures, and**

- **Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

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 31 will be carried out between September 2024 and September 2025, with the field period extending later into 2025 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 30 rounds of the survey. Despite the unique circumstances surrounding the coronavirus pandemic, approximately 75 percent of the living, in-scope, original respondents were surveyed in Round 30.

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.

NLS 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. NLS also uses 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. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

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 10 and 11. As shown in Table 10, attrition has been slightly lower among Black groups than among Hispanic groups and non-Black/non-Hispanic groups, and attrition for men is higher than for women. 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., although interviewing becomes a little more difficult each round. The NLS program has begun a project to measure the potential presence of non-response bias by matching a subset of its sample to data collected by the Census Bureau.

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, based on the original target population. These sampling weights are released with the other data on the public-use data file.

Although the NLSY79 weighting methodology adjusts for representativeness with respect to the original target population, based on U.S. residency in 1978, some data users may be interested in the extent to which the NLSY79 data are representative of the analogous age group currently living in the U.S. To investigate this issue of continued sample representation, Table 12 compares numbers from the 2020 decennial census with NLSY79 population estimates. Census data were taken from the website of the Census Bureau's Population Estimates Program. They provide the number of people living in the United States who were ages 55 to 63 on April 1, 2020, which is approximately the same age group that the NLSY79 sample represents. NLSY79 population estimates are from the weighted results of the Round 29 (year 2020) survey.

Table 12 shows the percentage of the NLSY79 sample and U.S. population by sex and race. Overall, the table has two significant features. First, the 2020 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). In the 2020 Census, 49.2 percent of the population in the relevant age range was male and 50.8 percent was female; a modest amount of differential mortality is evident in the change from 1979. Hence, as the NLSY79 continues to be weighted to represent its original composition it becomes slightly less analogous to the U.S. population in its age range.

Second, the NLSY79 sample underrepresents the current U.S. population of Hispanic people. The NLSY79 sample does not include persons who entered the United States after 1978, and the rate of immigration among Hispanic people 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 non-Black/non-Hispanic and Black non-Hispanic can be explained largely by the shortfall of Hispanic groups 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.

Table 12. NLSY79 Weighted Sample Composition in 2020 versus U.S. Census Data for Persons Ages 55 to 63 as of April 1, 2020

	2020 NLSY79	**Census Data**
Total	100.0%	100.0%
Men	50.8%	49.2%
Women	49.2%	50.8%
Non-Black, non-Hispanic	79.3%	75.7%
Men	40.2%	37.4%
Women	39.1%	38.3%
Black, non-Hispanic	14.2%	11.8%
Men	7.3%	5.5%
Women	6.9%	6.3%
Hispanic	6.5%	12.5%
Men	3.4%	6.2%
Women	3.1%	6.3%

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.

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, NLS often adopts questions that have been used previously in probability sample surveys. NLS has favored questions from the other surveys in the BLS National Longitudinal Surveys program to facilitate intergenerational comparisons. NLS also has 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, 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 NLS 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. NLS also watches 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, NLS monitors 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. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

Susan Paddock
Chief Statistician
NORC