

**SUPPORTING STATEMENT FOR
The National Longitudinal Survey of Youth 1997**

OMB CONTROL NO. 1220-0157

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.

This section summarizes the primary features of the sampling and statistical methods used to collect data and produce estimates for the NLSY97. Additional technical details are provided in the NLSY97 Technical Sampling Report, available online at <https://www.nlsinfo.org/content/cohorts/nlsy97/other-documentation/technical-sampling-report>. Chapter 2 of the report describes the design of the NLSY97 sample. Chapter 3 describes the sample-selection process. Chapter 4 describes the sample weighting process. Chapters 5 and 6 describe the accuracy and representativeness of the sample.

Additional information about statistical methods and survey procedures is available in the NLSY97 User's Guide at:
<https://www.nlsinfo.org/content/cohorts/NLSY97/>

The initial sample was selected to represent (after appropriate weighting) the total U.S. population (including military personnel) 12 to 16 years of age on December 31, 1996. The sample selection procedure included an overrepresentation of blacks and Hispanics to facilitate statistically reliable analyses of these racial and ethnic groups. Appropriate weights are developed after each round so that the sample components can be combined to aggregate to the overall U.S. population of the same ages. Weights are needed to adjust for differences in selection probabilities, subgroup differences in participation rates, random fluctuations from known population totals, and survey undercoverage. Computation of the weights begins with the base weight and then adjusts for household screener nonresponse, sub-sampling, individual nonresponse, and post-stratification of the nonresponse-adjusted weights. The number of sample cases in 1997, the first round, was 8,984. Retention rate information for subsequent rounds is shown in the table below.

NLSY97 Fielding Periods and Sample Retention Rates

Round	Months conducted	Total respondents	Retention rate	Number of deceased sample members	Retention rate excluding the deceased
1	February–October 1997 and March–May 1998	8,984	—	—	—
2	October 1998–April 1999	8,386	93.3	7	93.4
3	October 1999–April 2000	8,209	91.4	16	91.5
4	November 2000–May 2001	8,081	89.9	15	90.1
5	November 2001–May 2002	7,883	87.7	25	88.0
6	November 2002–May 2003	7,898	87.9	30	88.2
7	November 2003–July 2004	7,755	86.3	37	86.7
8	November 2004–July 2005	7,503	83.5	45	83.9
9	October 2005–July 2006	7,338	81.7	60	82.2
10	October 2006–May 2007	7,559	84.1	77	84.9
11	October 2007–June 2008	7,418	82.6	90	83.4
12	October 2008 – June 2009	7,490	83.4	103	84.3
13	September 2009 – April 2010	7,561	84.2	112	85.2
14	October 2010 – May 2011	7,420	82.6	118	83.7
15	September 2011 – June 2012	7,423	82.6	134	83.9
16	October 2013 – June 2014	7,141	79.5	151	80.8
17	September 2015 – May 2016	7,103	79.0	173	80.6
18	October 2017 – October 2018	6,734	75.0	207	76.7
19	October 2019–July 2020	6,948	77.3	231	79.4

Note 1: The retention rate is defined as the percentage of base year respondents who were interviewed in a given survey year.

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 NLSY97 includes personal interviews with all living Round 1 respondents, regardless of whether they subsequently become institutionalized, join the military, or move out of the United States. We employ a thorough and comprehensive strategy to contact and interview sample members. At each interview, detailed information is gathered about relatives and friends who could assist NORC field staff in locating respondents if they cannot readily be found in a subsequent survey round. Every effort is made to locate respondents. Interviewers are encouraged to attempt to contact respondents until they reach them. There is no arbitrary limit on the number of call-backs.

Preceding the data collection, the NORC interviewers are carefully trained, with particular emphasis placed on resolving sensitive issues that may have appeared in prior rounds. Most of the NORC interviewers have lengthy experience in the field from having participated in earlier NLSY97 rounds as well as from involvement with the NLSY79 and other NORC surveys. All new recruits are given one day of personal training on general interviewing techniques. All interviewers (whether having previous experience on the NLSY97 or NLSY79 or not) must complete a comprehensive and robust online training covering the questionnaire and its administration, data quality, study protocols and procedures and respondent confidentiality. Interviewers must attend training calls with their field managers and pass a certification test where they must demonstrate a full understanding of the questionnaire and how to administer it correctly. In addition to these trainings, the field staff received weekly memos throughout Round 19 which contained protocol reminders reinforcing proper procedures, tips for improving field work, and updates from the central office. This will be continued for Round 20.

Field interviewers are supervised by NORC Field Managers and their associates. All Field Managers complete the same online training that their interviewers will complete prior to the start of the study. NORC has divided the U.S. into 10 regions, each supervised by a Field Manager who is responsible for staffing and for the quality of field work in that region. A ratio of 1 supervisor to 15 interviewers is the standard arrangement. Field Managers are, in turn, supervised by one of the two Field Project Managers.

The interview content is prepared by professional staff at BLS, CHRR, and NORC. When new materials are incorporated into the questionnaire, assistance is generally sought from appropriate experts in the specific substantive area.

Because sample selection took place in 1997 in preparation for the baseline interview, sample composition will remain unchanged.

In Round 18, NLS converted the NLSY97 to a predominantly telephone survey, anticipating that approximately 75 percent of interviews will be completed by telephone, in contrast to a projected 26 percent in Round 17 and 15 percent in Round 16. Instead the telephone rates were higher with approximately 90% of interviews completed by telephone and an additional 30 percent of sample members requiring in-person outreach. In Round 19, the percentage of interviews completed by telephone was even higher (96%) while the in-person contact rate was lower (21%), in substantial part due to the Coronavirus pandemic, which prevented in-person contacts after mid-March 2020.

Within the survey research literature, both unit non-response and item non-response are documented to be higher in telephone administration than in in-person administration. (Safir and Goldenberg (2008) "Mode Effects in a Survey of Consumer Expenditures," Office of Survey Methods Research, Bureau of Labor Statistics retrieved from <http://www.bls.gov/osmr/abstract/st/st080200.htm>, Groves, Dillman, Eltinge and Little 2002 "Survey Nonresponse" New York: Wiley.) For the NLSY97, mode differences may come about from lack of coverage (respondents not having telephones), non-contact (telephone technology offers more ways for respondents to avoid interviewers), distraction (telephone respondents may be more likely to engage in other activities during the interview due to social norms governing in-person interactions), privacy (either increased perception of privacy from not being face-to-face with the interviewer or decreased privacy because the interviewer cannot ensure that the respondent is alone and out of the hearing range of others), or other factors.

To maintain connections that support response, the NLSY97 makes contact with its respondents between collection rounds. Such contacts may be through mailings or opportunities for respondents to update their contact information. Between Rounds 18 and 19, an effort was made to collect information on mode preferences and uses of technology from a randomly selected one half of the sample. Between Rounds 19 and 20, a supplemental survey of the whole sample was fielded to collect data on the rapidly changing impacts of the Coronavirus pandemic on respondents' employment and health. This ICR proposes to conduct an interim contact between Rounds 20 and 21 as done previously between prior rounds.

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 that are used to maximize response rates already have been discussed in items 1 and 2 above. The other component of missing data is item nonresponse. Nonresponse includes respondents refusing to answer or not knowing the answer to a question. Almost all items in the NLSY97 have low levels of nonresponse. For example, in prior rounds there was virtually no item nonresponse for basic questions like the type of residence

respondents lived in (YHHI-4400) or the highest grade of school respondents had ever attended (YSCH-2857).

Cognitively more difficult questions, such as “How many hours did you work per week?” (YEMP-23901) have low levels of nonresponse. In the hours per week example, 8 individuals out of 1,584 (0.5%) did not answer the question in Round 13.

Sensitive questions have the highest nonresponse. Table 4a and Table 4b present examples of Round 17 and Round 18 questionnaire items that are most sensitive or cognitively difficult. In Round 17, almost all respondents (0.6% nonresponse rate) were willing to reveal whether they had earned money from a job in the past year, but many did not know or refused to disclose exactly how much they had earned (13.6% nonresponse rate). Because high nonresponse rates were expected for the income amount question, individuals who did not provide an exact answer were asked to estimate their income from a set of predetermined ranges. This considerably reduces nonresponse on the income question. Only 8.6% of those who were asked to provide a range of income did not respond. These individuals represent about 1% (67/5770) of all individuals requested to provide income data in that round. The patterns for non-response to these items in Round 18 are similar with about 1% of individuals not providing income information for the previous year, though we see lower rates of nonresponse to the question on the exact amount of income they had earned from their jobs in the previous year (YINC-1700).

Table 4a. Examples of Nonresponse Rates for Some Round 17 Sensitive Questions

Q Name	Question	Number Asked	Number Refused	Number Don't Know	% Nonresponse
YINC-1400	Receive Work Income in 2014?	7103	37	9	0.6%
YINC-1700	How Much Income from All Jobs in 2014?	5770	78	713	13.6%
YINC-1800	Estimated Income from All Jobs in 2014? ¹	783	46	21	8.6%

¹Asked of respondents who were unable or unwilling to answer the previous question (YINC-1700).

Table 4b. Examples of Nonresponse Rates for Some Round 18 Sensitive Questions

Q Name	Question	Number Asked	Number Refused	Number Don't Know	% Nonresponse
YINC-1400	Receive Work Income in 2016?	6734	42	9	0.8%
YINC-1700	How Much Income from All Jobs in 2016?	5523	70	362	7.8%
YINC-1800	Estimated Income from All Jobs in 2016? ¹	432	48	14	14.4%

¹Asked of respondents who were unable or unwilling to answer the previous question (YINC-1700).

To reduce the proportion of “don't know” or “refused” responses to questions on income or assets (such as YINC-1700, shown in Table 4a and Table 4b), respondents who do not provide exact dollar answers are asked follow-up questions designed to elicit approximate information. For many income categories, the respondents are asked to select the applicable category from a predefined list of ranges. The approach for asset questions is slightly different: The initial question asks the respondent to provide an exact value, but if he or she is unable or unwilling to do so, interviewers are instructed to ask the respondent to define a range for the value using whatever values he or she feels are appropriate. If the respondent does not know or refuses to provide either an exact value or a range, a follow-up question asks him or her to select the appropriate range from a predefined list. This method provides researchers with some information on income, asset, and debt amounts when the respondent is reluctant or unable to furnish an exact figure.

Face-to-face interviewing has been found to lead to under-reporting of sensitive items relative to telephone interviewing. Thus, sensitive items that had previously been interviewer-administered in-person but will now be interviewer-administered by telephone may experience decreases in item non-response. Income and other financial items would be the chief examples of such items. (de Leeuw E.D., van der Zouwen J. (1988). “Data quality in telephone and face to face surveys: a comparative metaanalysis.” In: Groves RM, Biemer PP, Lyberg LE, Massey JT, Nicholls WL II, Waksberg J, eds. *Telephone Survey Methodology*. New York: Wiley: 273:99).

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 NLSY97 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 NLSY97 design process employs a multi-tiered approach to the testing and review of questionnaire items.

When new items are proposed for the NLSY97 questionnaire, we often adopt questions that have been used previously in probability sample surveys with respondents resembling the NLSY97 sample. NLS has favored questions from the other surveys in the BLS National Longitudinal Surveys program to facilitate intergenerational comparisons. NLS has also used items from the Current Population Survey, the Federal Reserve Board's Survey of Consumer Finances, the National Science Foundation-funded General Social Survey, and other federally funded surveys. In addition, NLS consults with experts to ensure the proper functioning of new questions, conducting evaluations of these questions as necessary.

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, recordings of items during the interviews, and comparison of NLSY97 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 or to get a divorce. 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. This review continues at a lower level throughout the field period, with interviewers receiving ongoing training until the end of the field period. All changes made between the Round 19 and Round 20 questions are listed in Attachment 5. Round 20 questions that have not appeared in previous rounds of the NLSY97 include:

Question on hours worked from home. NLS proposes to ask a follow-up question after the report of hours worked at a job to collect how many of those hours are usually worked from home. Such information may help classify jobs as “telework-able” or not, and help explain which jobs were able to continue during the Coronavirus pandemic.

Question on value of a job. NLS proposes to ask a question on the value of respondent’s main on-going job, “Imagine that your employer in this job offers you a one-time payment on the condition that you end your employment immediately. For what minimum dollar amount would you accept the offer to quit your job immediately?” A very similar question has been asked in the German Socio-economic panel and has performed well. This item would be useful in studies examining compensating differentials, fringe benefits, and job search behavior, and may serve as a measure of risk aversion. NLS will work with the Office of Survey Methods Research to evaluate this item prior to fielding.

Questions on labor market volatility due to Coronavirus pandemic. NLS proposes to add a question to the Round 20 questionnaire that asks about changes to hours/location/earnings due to Coronavirus pandemic for jobs worked after March 1, 2020. This item will help researchers evaluate job and earnings changes that occurred over this period. An identical item was included in NLSY79 Round 29, and it appears to be performing well in the field. A similar question was included in the NLSY97 Coronavirus pandemic supplement, capturing a summary measure of changes for *any* job. The Round 20 question aims to fill in the details as they pertain to *each* job and will be included as part of the event history on employment that is collected in the NLSY97 survey.

Questions on whether reasons for leaving a job, for not working, and for not looking for a job were related to the Coronavirus pandemic. Following items that ask the reason the respondent left an employer, was not working, or was not looking for work during period after March 1, 2020, NLS proposes to ask “Was that due to the coronavirus pandemic?” This item will help assess the labor market impacts of the Coronavirus pandemic on individuals by permitting researchers to classify job loss and non-work spells as related to the pandemic. NLS will evaluate adding an end date past which these questions are not asked once the Coronavirus pandemic has passed.

Questions on contracting Coronavirus. NLS proposes to add two questions asking whether individuals have contracted Coronavirus. These items may help us understand labor market activity over this period and long-term health outcomes as we continue to follow this sample. The items match those included in NLSY79 Round 29 and were recommended to NLS by NCHS in March 2020.

Questions on vaccines. NLS proposes to include two items about getting vaccines in Round 20. The first asks whether the respondent has received a Coronavirus vaccine. The second asks whether the respondent has received a flu vaccine in the last 24 months. The first question is new to the main round of NLSY97, but was included in NLSY79 Round 20 and the NLSY97 Coronavirus supplement. The latter question has been included in the NLSY79 for several rounds. The first item may help explain employment, risk aversion, and health outcomes, while the second may gauge general attitude toward vaccines and access and take-up of preventative

medicine.

Questions on application for Unemployment Compensation. NLS proposes to include three items asking whether respondents applied for unemployment insurance after March 1, 2020, how many times they applied for unemployment insurance, and what was the resolution of their initial application. NLS will continue to collect the event history on receipt of unemployment insurance. The new questions on unemployment insurance applications can provide a better picture of this cohort's experiences applying for unemployment compensation during the start of the Coronavirus pandemic.

Questions on criminal background checks. NLS proposes to include two questions that ask about criminal background checks. An initial question would ask how long it has been since the respondent applied for a job. The second question would ask when in the application process the potential employer asked if the respondent had been convicted of a crime. Including such items in the NLSY97 would permit researchers to access a collection of valuable life and employment history indicators that are not available in other surveys. This would extend NLS' understanding of implications of criminal records inquiries on labor market outcomes for applicants. These questions are new to the NLSY97. NLS will work with the Office of Survey Methods Research to evaluate these items prior to data collection.

Questions on internet access. NLS proposes to add four questions on internet availability at home and internet access. These questions have been asked in Pew surveys since 2015. Such items measure the "digital divide" and permit researchers to study the relationships between internet access, skills in the labor market, job search, and the ability to work from home, especially during the Coronavirus pandemic. These items might also inform future internet data collection for the NLSY97.

Questions on experience with Coronavirus supplemental survey and mode preference. To inform future internet data collection, NLS proposes to ask three items about respondent experiences with the NLSY97 supplemental Coronavirus survey fielded between Round 19 and 20. The items would vary depending on participation in the supplemental survey. For respondents who participated, NLS proposes to ask why they chose the mode that they did, whether they encountered any problems, and how likely they are to participate in internet collection if NLSY97 were to offer it in the future. For those who did not participate, NLS proposes to ask whether they were contacted about participating in the NLSY97 Coronavirus supplement, why they did not participate, and how likely they are to participate in internet collection if NLSY97 were to offer it in the future.

A list of all changes to the NLSY97 questionnaire from rounds 19 to 20 is contained in attachment 5.

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 person(s) who will actually collect and/or analyze the information for the agency.

Kirk Wolter
NORC
55 East Monroe Street, Suite 3000
Chicago, IL 60603

The sample design was conducted by NORC, which continues the interviewing fieldwork.