

SUPPORTING STATEMENT FOR THE NATIONAL LONGITUDINAL SURVEY OF YOUTH 1997 (NLSY97)

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.**

In data collection during its main rounds, 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. It employs 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.

The proposed supplement will also aim to reach all living Round 1 respondents and will also employ a concerted effort to locate, contact, and gain cooperation from them. However, it will be conducted in a short time frame (16 weeks) relative to the main rounds. Its interviews will be considerably shorter than those of main rounds, with a target of 12 minutes to complete. To minimize burden and cost, it will be conducted through a mixed mode approach. Sample members will be contacted via email and invited to complete the interview via an internet-based collection tool. Follow up will be conducted by phone, and a phone interview will be offered to respondents not completing the internet interview. NLS expects this approach to maximize the supplement's response rate given its short time frame. As an extensive interim engagement to be conducted between main rounds, and due to its focus on a topic of broad interest, the proposed supplement may cultivate respondent engagement in ensuing main rounds.

As in the collection of its main rounds, NLS will employ a well-trained and professional cadre of interviewers through its contractor CHRR and their subcontractor, NORC. Preceding the data collection, NORC interviewers are carefully trained, with particular emphasis placed on addressing sensitive issues that may have appeared in prior rounds. All interviewers employed for this interim supplemental collection will have worked on the NLS97 or NLSY97 surveys previously. All interviewers must complete a comprehensive and robust online training covering the questionnaire and its administration, data quality, study protocols and procedures and respondent confidentiality. In preparation for this supplemental collection, the project will continue training modules with content targeting the promotion of data quality in a mixed mode setting. Interviewers must attend training calls with their field managers and pass a certification where they must prove a full understanding of the questionnaire and how to correctly administer it. In addition to these trainings, the field staff received weekly memos during the data collection period which will contain protocol reminders reinforcing proper procedures, tips for improving field work, and updates from the central office.

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. The supplemental questionnaire focuses on content about the COVID pandemic. NLS has consulted with a number of other surveys, including the Panel Study of Income Dynamics, Census Household Pulse Survey, and the Understanding American Survey about content on these this. Additional assistance was sought from appropriate experts on labor market disruptions.

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

Because the supplement will employ multiple modes of collection, the collected data may exhibit mode differences. Such differences may come about from lack of coverage (respondents may not have reliable internet or telephone access), non-contact (different technologies offer different levels of convenience and different ways for respondents to avoid interviewers), distraction (internet and telephone respondents may be more likely to engage in other activities during the interview), privacy (perceptions of privacy may differ due to different levels of interaction with interviewers and the ability for others to overhear responses), or other factors.

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).

One potential advantage of the use of internet and telephone interviewing for the supplement is that it may be expected to incur less item non-response than face-to-face interviewing. Face-to-face interviewing has been found to lead to under-reporting of sensitive items relative to telephone interviewing. 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 has been 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. 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, and other federally funded surveys. We have followed this model in designing this supplement, while paying special attention to surveys such as the Household Pulse Survey that have used internet tools to collect information about the Coronavirus pandemic.

Existing questions are also reviewed each year. Respondents' ages 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, recordings of items during the interviews, and comparison of NLSY97 response data to other sources for external validation. For questions related to the Coronavirus pandemic, NLS will gain information from early experiences with similar questions that are administered in Round 29 of the NLSY79 during Fall 2020.

The content for the interim supplemental data collection is included in Attachment 5. Some items have appeared in previous rounds. Others are novel. Those that are new to NLSY97 are based on

questions in NLSY79 Round 29 and or Census' Household Pulse Survey. All general topics covered have appeared in previous rounds of the NLSY97. The content of the interim supplement is described below.

BLS conducted tests to assess if questions in the supplement work as intended and that respondents can understand and answer them both for themselves and for other members in their household. These tests used cognitive interviews to provide an in-depth understanding of the participant's thought processes and reactions to the questions. Test participants were asked to complete the survey and then answer debriefing questions to better understand their responses and reactions to the questions, both when reporting for themselves and when reporting information for other household members, if applicable. The interviews were conducted virtually on Microsoft Teams or over the phone, with participants from anywhere in the United States. The findings from all the cognitive interviews were evaluated qualitatively, and used to arrive at conclusions about the effectiveness of the proposed wording.

Questions on current employment: We propose to ask a series of question to understand the respondent's employment situation during the past 7 days. These items come from Round 10 of the NLSY97. Similar items are asked in the Household Pulse Survey.

Questions of current employment of spouse/partner: We propose to ask a series of question to understand the spouse/partner's employment situation during the past 7 days. These items are parallel to those asked for the respondent. Questions about spouse/partner's hours of work over the past 12 months have been asked in recent rounds of the NLSY97.

Questions on changes in work over the last 12 months due to COVID: We propose to ask a question about changes in work over the last 12 months due to COVID. This question is included in the NLSY79 Round 29 Questionnaire to be fielded beginning in fall 2020. The question attempts to capture changes in hours worked, work location, ending a job, and starting a new job.

Questions on school closure/distance learning: We propose to include questions on school enrollment, the use of distance learning, and the extent to which it caused difficulty for the respondent.. The questions are similar to those in the Household Pulse Survey and the Understanding America Study.

Questions on health: We propose to ask a general question on self-rated health, whether the respondent has been diagnosed with COVID, and a depression scale. All of these items appear in NLSY79 Round 29. With the exception of the questions that focus on COVID, all items were included in NLSY97 Round 19.

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.

Susan Paddock
Chief Statistician
55 East Monroe Street, Suite 3000
Chicago, IL 60603

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