

## SUPPORTING STATEMENT

National Longitudinal Survey of Youth 1979, 14–21 Years of Age on December 31, 1978  
23<sup>rd</sup> Round (2008 Survey) Rationale, Objectives, and Analysis of Content

### ***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 total 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. Another group that was oversampled was economically disadvantaged non-black/non-Hispanic youths. The NLSY79 also originally included a supplemental sample of youths in the military. In 1985, the military supplemental sample was discontinued, and 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 in 1979, excluding the discontinued military and non-black/non-Hispanic samples, was 9,964. A breakdown by sex and race is depicted in table 6 below. We anticipate a response rate in round 23 that is similar to the round 22 experience.

**Table 6. Civilian Sample Interviews Completed in 1979 and 2006 (Preliminary) by Race and Sex**

Race and Hispanic origin	1979			2006					
	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	2518	2484	5002	1827	72.56%	1956	78.74%	3783	75.63%
Hispanic	981	980	1961	737	75.13%	753	76.84%	1490	75.98%
Black	1524	1477	3001	1174	77.03%	1207	81.72%	2381	79.34%
Total	5023	4941	9964	3738	74.42%	3916	79.26%	7654	76.82%

Retention rates for the NLSY79 are significantly affected by attrition due to death. Approximately 4.8% of the 9,964 NLSY79 respondents still eligible for interviewing were deceased after the 2006 survey; we are investigating ways to determine whether some respondents we cannot locate may be deceased. Table 7 provides information about retention (percent of all respondents interviewed) and response (percent of living respondents interviewed) rates for each year of the NLSY79.

**Table 7. NLSY79 retention and response rates by sample type**

Year	Number Interviewed	Retention Rate <sup>1</sup>	Number of Deceased Respondents	Response Rate <sup>1</sup>
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	<sup>2</sup> 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	<sup>3</sup> 9,018	90.5	145	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	368	80.3
2004	7,661	76.9	421	80.3
2006	7,654	76.8	477	80.7

<sup>1</sup>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.

<sup>2</sup>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.

<sup>3</sup>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.

## **2. Design and Procedures for the Information Collection**

The survey includes personal-visit or telephone 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 respondent if he or she cannot be readily located in the subsequent survey round. Interviews in round 23 will be carried out between January and December 2008, with the field period extending into January 2009 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 put 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 22 rounds of the survey. Over 80 percent of the living, in-scope, original respondents were surveyed in 2006.

Preceding the data collection, NORC interviewers are carefully trained, with particular emphasis placed on resolving sensitive issues. Most of the NORC interviewers have lengthy experience in the field from participation in earlier NLSY79 interview rounds, as well as from involvement with other NORC surveys. All new recruits are given one day of personal training on general interviewing techniques, followed by three days of personal training on the questionnaire and field procedures. 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. 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. If anything suspicious is found as a result of a validation interview, the entire caseload of the interviewer will be validated.

NORC's Field Managers (FMs) supervise the field interviewers. For each round, NORC divides the continental U.S. into approximately 15 regions (depending on the locations of respondents), each supervised by a Field Manager who is responsible for staffing and for the quality of fieldwork in the region. A ratio of 1 supervisor to 15 interviewers is the standard arrangement. FMs are, in turn, supervised by one of three Divisional Field Managers.

The interview schedules are prepared by professional staff at the Center for Human Resource Research (CHRR) at The Ohio State University under contract with the U.S. Department of Labor. 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 NLS User Services at CHRR (phone: 614-442-7366, e-mail: [usersvc@postoffice.chrr.ohio-state.edu](mailto:usersvc@postoffice.chrr.ohio-state.edu)).

In an effort to reduce respondent burden while still providing a broad spectrum of variables for researchers and policy makers 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 series of questions on promotions and job hierarchies that was asked for the first time in 1990. A redesigned series of questions on promotions was asked again in 1996 and 1998 but was not included in subsequent surveys.

### 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 6 and 7. Hispanic attrition has been slightly higher than among blacks and non-black/non-Hispanics, and attrition for men is higher than for women. There is no evidence of any selective response problems that could bias analytical results due to the efforts of the high-quality NORC interviewers. 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 under 0.2 percent per question. 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 8 compares numbers from the 2000 decennial Census with NLSY79 population estimates. Census data, are taken from the FactFinder.Census.Gov website, which is the official website for disseminating Census 2000 data.<sup>1</sup> The FactFinder website provides the number of people living in the U.S. who were ages 35 to 43 on April 1, 2000, which is the same age group that the NLSY79 sample represents. NLSY79 population estimates are from the weighted results of the round 19 (year 2000) survey.

Table 8 shows the percentage of the NLSY79 sample and U.S. population by sex and race. Overall, the table has two significant features. First, the 2000 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 quite 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.

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

---

<sup>1</sup> Three series were extracted from the FactFinder website; PCT12, PCT12J, and PCT12H which track the U.S. population's age broken down by sex and race/ethnicity. PCT12 tracks the total population, PCT12J tracks black non-Hispanics and PCT12H track the Hispanic population. Non-black, 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.

percentages of non-black/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 8 shows that except for the Hispanic population the NLSY79 sample is still 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 8. NLSY79 Weighted Sample Composition in 2000 versus U.S. Census Data for Persons Ages 31 to 39 as of April 1, 2000**

	<b>2000 NLSY79</b>	<b>Census Data</b>
<b>Total</b>	100.0%	100.0%
Men	50.9%	49.8%
Women	49.1%	50.2%
<b>Non-black, non-Hispanic</b>	79.3%	76.4%
Men	40.2%	38.1%
Women	39.1%	38.3%
<b>Black, non-Hispanic</b>	14.2%	12.1%
Men	7.3%	5.7%
Women	6.9%	6.4%
<b>Hispanic</b>	6.5%	11.5%
Men	3.4%	6.0%
Women	3.1%	5.6%

#### 4. Testing of Questionnaire Items

A comprehensive pretest of the main NLSY79 questionnaire is carried out approximately 3 months preceding each round of the regular survey. Assuming that our separate proposal to augment the pretest is approved, this pretest will include about 100 respondents from different racial, ethnic, geographic, and socio-economic backgrounds. On the basis of this pretest, the various questionnaire items—particularly those being asked for the first time—are evaluated with respect to question sensitivity and validity. When necessary, problem items are deleted from the final survey questionnaire.

#### 5. Statistical Consultant

Dr. Kirk Wolter  
 NORC  
 1155 East 60th Street  
 Chicago, IL 60637  
 (312) 753-7500

A detailed statement of the sampling plan, prepared several years ago by Dr. Martin Frankel, is available from NLS User Services at CHRR. The sample design and the interviewing fieldwork are being carried out by NORC, 1155 East 60th Street, Chicago, IL, 60637.

As indicated earlier, analyses of the data collected in this survey round will be prepared for the contracting agencies by the Center for Human Resource Research, The Ohio State University, 921 Chatham Lane, Suite 100, Columbus, OH, 43221.