Timetable for Upcoming Data Releases

Do you need to know when the next dataset will be released? The following is the latest schedule for NLSY79 and NLSY97 data collections and data releases:

- The NLSY97 is an annual survey with 11 rounds of data now available to the public. Round-12 data are scheduled for release in the summer of 2010. Round-13 interviews began in September 2009.

- The NLSY79 main survey began as an annual survey in 1979 and became a biennial survey in 1996. Twenty-two rounds of data have been released to the public so far. Data collected during round 23 are scheduled for release in spring 2010. Round-24 fielding is scheduled to begin December 2009.

- Both the NLSY79 Child survey, which began in 1986, and the NLSY79 Young Adult survey, which became a separate survey in 1994, interview children of the main NLSY79 respondents and are fielded at the same time as the main NLSY79 survey. Child and Young Adult data from 1986 to 2006 are currently available, along with a preliminary release of the Young Adult 2008 data. The next full data release is scheduled for mid-2010 and will include 2008 data for both the Child and Young Adult surveys.

Variables Created to Help Match Mother-Daughter Pairs

New variables have been added to the NLS Women’s Survey (the women cohort respondents of the Original Cohorts) dataset to assist researchers in matching mother-daughter pairs within the series of questions about intergenerational transfers in the survey. Mother-daughter pairs are those pairs in which the mother was interviewed as a respondent in the Mature Women survey, and the daughter was a respondent in the Young Women survey. These pairings provide researchers a rich opportunity to compare two generations of the same family. These new variables are found in the transfer variable section, which includes questions about the transfer of money and time between a mother and her children. Examples of transfers include a mother giving financial assistance to her child through monetary loans and gifts, as well as time spent between a mother and child for child care, personal care, chores, and errands.

To link the Mature Women transfer data to the Young Women cohort, the variables R51513.01 through R51513.35 and R76242.01 through R76242.28 were added. These series of created variables include the identification codes of the daughters as well as “quality of match” variables, created to provide users with a rating of certainty regarding the accuracy of the match (that is, if it was an exact match, likely match, or probable match).

More details about these added variables as well as other variables used for matching mother-daughter pairs will be available in “Appendix 44: Variables Used For Matching Mother-Daughter Pairs in the NLS” in the Mature Women Codebook Supplement.

New Financial Condition Variables in Latest NLSY97 Dataset

Four questions in the NLSY97’s round-11 (2007) survey ask respondents about negative financial situations they may have experienced, thus providing researchers additional information on the severity of a respondent’s economic troubles. Respondents were asked if, in the 12 months preceding the interview, they or their spouse or partner had used a cash advance service using any of their credit cards, had obtained a payday loan, had been late in paying their rent or mortgage by more than 60 days, or had been pressured to pay bills by stores, creditors, or bill collectors. (See questions YINC-7950 through YINC-7980.)

Respondents also described their current financial situation by selecting the condition that best matches their state of affairs. Approximately 16 percent of them rated their situation as very comfortable and secure, 42 percent reported that they were able to make ends meet without much difficulty, 26 percent occasionally had some difficulty making ends meet, 13 percent responded that it was tough to make ends meet but they were keeping their heads above water, and 3 percent said they were in over their heads. (See YINC-7990.)

Note: These percentages are unweighted.

These financial condition questions can be found in the “Income” Area of Interest in the NLS Investigator (http://www.nlsinfo.org/investigator). The same questions will be repeated in round 12.
Online Tutorials Available for Users

Currently, four tutorials are available online for users who are seeking guidance on some of the more complex aspects of the NLSY datasets. Each tutorial provides step-by-step details as well as tips.

- “Linking Roster Items Across Rounds” explains how to find the same member of the respondent’s household across different NLSY79 survey rounds. The goal is to link household roster items across survey rounds using unique household member ID codes. Users learn how to find the household roster variables, extract selected roster data, and compare unique ID (UID) codes across rounds.

- “Matching Cohabitating Partners to their Characteristics in the NLSY97” assists users in linking a respondent’s cohabiting partner to the partner’s characteristics using information from the created event history arrays, household and non-resident rosters, and partner rosters. The tutorial takes users through the process of finding the age of the NLSY97 respondent’s first cohabiting partner. Users learn how to find and tag the relevant variables, extract selected variables, find the partner ID for the first partner, split the partner ID into round and loop number, locate the partner on the partner roster, and link the partner to his or her age.

- “Constructing Comparable Samples across the NLSY79 and NLSY97” walks users through the basic steps of constructing parallel samples for research projects that use both the NLSY79 and NLSY97 cohorts. The tutorial uses a specific example of constructing work status at age 20 for both samples. Users learn how to select the samples for analysis, determine the age and interview years needed for the analysis, create tagsets of variables to define work status at age 20 for both cohort datasets, and construct the work status variable for both samples.

- “Intergenerational Links: NLSY79 Mothers and Their Children” explains the general logic behind linking mothers and children of any age covered in the Children of the NLSY79 dataset. The tutorial also uses a specific example of mothers and young adult daughters to show how to create parallel variables indicating a first birth prior to age 18.

A fifth tutorial on how to conduct variable searches using the new NLS Investigator is in the works.

The first three tutorials listed above can be accessed individually at http://www.nlsinfo.org/nlsy97/nlsdocs/nlsy97/maintoc.html under “Other NLSY97 Documentation.” The fourth tutorial, on linking mothers and children, is available at http://www.bls.gov/nls/nlsy79ch.htm under “Children and Young Adult Data and Documentation.”

NLS Terms to Know: Valid Skip

A “valid skip” occurs when an NLS respondent is systematically skipped over a question or a section of a survey because the respondent is not part of the universe that is required to answer that question or question series. Respondents are not asked every question. For instance, some questions might apply to only one gender or to only a certain age range. Users should trace back skip patterns to determine whether a respondent was skipped out because a given topic was inapplicable to him or her or because the respondent answered similar questions along a different path.

The number of valid skips is provided on each variable’s codebook page in the NLS Investigator at http://www.nlsinfo.org/investigator. The codebook total includes the number of respondents who provided an answer for that particular question, those who answered “don’t know,” and those who refused to answer the question. It also provides the number of valid skips (in other words, the number of respondents who were skipped past that question for legitimate reasons) for that variable.

For example, in the NLSY97 round-11 (2007) survey, respondents were asked how often they had done any volunteer work during the 12 months preceding the interview. (See YSAQ-300V1.) Of the 7,418 respondents who were interviewed in the 2007 round, the codebook indicates that 2,508 respondents said they had participated at least once in some sort of volunteer work. The 4,910 remaining respondents said that they had not participated at all, they didn’t know whether they had or not, or they did not wish to answer the question. Only the 2,508 respondents who said they had participated in volunteer work were then asked the follow-up question about volunteerism: “Which of the following is the main reason you do volunteer work?” (See YSAQ-300V2.) The codebook page for that variable shows that 4,910 remaining respondents were valid skips, which are coded as “-4”.

“Invalid skips” occur when a respondent is inadvertently skipped over a section that he or she should have answered. This most often happened in the early years of the NLS surveys when the surveys were administered using paper and pencil, and the skip patterns were sometimes difficult to control. The computerization of the NLS surveys greatly reduced the number of invalid skips.

An examination of the questionnaires themselves is probably the best way to figure out complicated skip patterns. Users who have concerns that a particular variable does not have enough responses are encouraged to trace skip patterns back to determine the universe of respondents who were asked a question and those respondents who were skipped past that question or section. For a list of all NLS documentation, including the questionnaires, go to http://www.bls.gov/nls/nlsdoc.htm and search under “Questionnaires and Codebooks.”

Special Issue of Marriage & Family Review

Marriage & Family Review has released a special edition (Issue 45, 2-3, April 2009) featuring NLSY97 family process data. The special issue was made possible by the office of the U.S. Department of
Q1. I am currently using the restricted-use Geocode NLSY79 data and am having difficulty locating the variables that identified the country of origin for the respondent’s parents and the country of origin for the parent’s spouse or partner. I remember the variables being on an earlier round of restricted data.

A1. The variables for country of origin of parent and parent’s spouse or partner have been removed from the geocode release. After review, it was decided that releasing these variables posed too much of a risk of respondent disclosure.

Q2. Looking at NLSY79 income data in several survey years, I observe that the value “$839,078” is repeated for many observations. I don’t quite believe that “$839,078” is the exact amount of income for so many of the respondents. Why do many people share this maximum number?

A2. From 1989 to 1994, high income values were replaced with a group average to protect respondent confidentiality. For example, respondents who reported a wage income of more than $100,000 in 1992 do not have their actual responses listed in the dataset. Instead, those respondents’ income values are simply the group average of all people who reported an income of more than $100,000. This use of the group average resulted in many respondents having identical values.

Q3. I noticed that for the NLSY79 variable I’m using, the survey year is indexed as “XRND” rather than an actual survey year. What is the definition of XRND?

A3: The XRND denotation refers to a “cross round” variable. If a variable has the year listed as XRND, it means the information to create this variable came from the latest interview, whatever round that may be. For example, the NLSY79 created variable “CVC_HGC_EVER_2007” (the highest grade the respondent ever reported to have completed, as of the 2007 data collection) is an XRND variable because the highest grade completed information is present for each respondent regardless of whether they were interviewed in 2007.

Q4. How soon after they turn 40 do NLSY79 respondents receive the 40+ Health Module questions?

A4. The protocol was to administer the 40+ Health Module to respondents on one occasion after they turned 40 and to interview respondents as close to age 40 as possible. Because of the way birthdays and interview dates fall in a given year, a respondent wasn’t always interviewed at age 40. For instance, if a respondent who is interviewed in each round turned age 40 in 1998—but was interviewed prior to his birth date—then he would not take the 40+ Health Module until the next possible survey year. (In this case, the next survey year would be 2000 because NLSY79 surveys are biennial.) Therefore, the respondent would be either age 41 or 42, depending on the date of interview. If this same respondent was not interviewed in the 2000 interview, he would be interviewed in 2002, receiving the health module at that time. In that case, his age would be 43 or 44 at the date of interview.

The income section of the NLSY79 Users Guide, available at http://www.nlsinfo.org/nlsy79/docs/79html/79text/income.htm, has a more complete description of various topcoding algorithms the NLSY79 has used. Topcoding primarily affects seven of the NLSY79 income variables: income from respondent’s wages, respondent’s business, spouse’s wages, spouse’s business, partner’s wages, the wages of the rest of the family, and other sources such as rents, interest, and dividends.

Q5. Are the units in “INCOME” variables given in current dollars, or are they inflation-adjusted to the latest year of the survey?

A5. The income variables are in current dollars.

Q6. Where can I find the father’s race in the NLSY79 Child/Young Adult Survey?

A6. Father’s race is collected only after children become eligible for the Young
Completed NLS Research

The list that follows shows bibliographic information for recent research based on data from the NLS cohorts that has not appeared in its current form in a previous issue of the NLS News. (See the NLS Annotated Bibliography at http://www.nlsbibliography.org for a comprehensive listing of NLS-related research.)


Caputo, Richard K. and Susan E. Mason. “Role of Intact Family Childhood on Women’s Earnings Capacity: Implications for Evidence-Based Practices.” Journal of Evidence-Based Social Work 6, no. 3 (July 2009): 244-55. [Young Women]


Griffith, Amanda L. and Donna S. Rothstein. “Can’t Get There From Here: The Decision To Apply To A Selective College.” Economics of Education Review 28, no. 5 (October 2009): 620-628. [NLSY79]


Orth, Ulrich, Richard W. Robins, and


NLS Contact Information

NLS News is published quarterly by the U.S. Bureau of Labor Statistics. It is distributed both nationwide and abroad without charge to researchers using NLS data, as well as to other interested persons.

NLS User Services: Center for Human Resource Research
The Ohio State University
921 Chatham Lane, Suite 100
Columbus, Ohio 43221-2418
usersvc@chrr.osu.edu
(614) 442-7366
(614) 442-7329 (Fax)

NLS documentation, data, and data updates:

NLS Web site: www.bls.gov/nls

BLS-NLS publications: Jain_Rita@bls.gov
(202) 691-7405

NLS Program Office: National Longitudinal Surveys
2 Massachusetts Avenue, NE.
Room 4945
Washington, DC 20212-0001
Attention: Rita Jain
NLS_INFO@bls.gov
(202) 691-7405
(202) 691-6425 (Fax)

NLS News Editor: Donna S. Rothstein
Rothstein_D@bls.gov

NLS Program Director and Media Contact: Charles R. Pierret
Pierret_C@bls.gov
(202) 691-7519