



U.S. Department of Labor
U.S. Bureau of Labor Statistics

NLSY97 Round 7 Event History and Geocode Data Release

Event history and geocode data from round 7 of the National Longitudinal Survey of Youth 1997 (NLSY97) will be available in mid-October. Data from the second wave of the high school transcript survey and a newly created summary percentile score variable indicating results from the Armed Services Vocational Aptitude Battery (ASVAB) test also are included on this release. Round 1 interviews were conducted with 8,984 young adults aged 12 to 16 as of January 1, 1997; of these, 7,756 respondents (86.3 percent) were interviewed in round 7. This article describes the data and documentation available with this release.

Event history data

In addition to all variables on the main data file, including interview data, ASVAB scores, and transcript survey data, the event history data file contains created variables covering four major topics:

- employment status
- marital status
- program participation
- schooling experiences

Each topic is documented through a series of arrays that show the respondent's status or activities during each week, month, or year within a specific period.

Employment status. Employment status of each respondent (working for a specific employer, unemployed, out of the labor force, and so forth) is included for each week from the respondent's 14th birthday to the most recent interview date. Researchers are able to link the job in the array with associated characteristics using

the employer ID (YEMP_UID) that identifies the jobs held in a particular week in the status and the dual job arrays. This section also provides data on total hours worked at all civilian jobs each week and data on additional jobs held in the same week, where applicable. Finally, the section includes beginning and ending dates for each job and for gaps within jobs, allowing these dates to be easily linked to the employment arrays.

Also included are employment status "deny" variables. These variables flag employment status data that a respondent denies having reported in a previous survey round. For example, some respondents report working for a specific employer in one round and then later deny that they have ever worked for that employer.

Three types of job-specific variables, not in arrays, have been included since round 4 to show discrepancies in start dates of new jobs. Some respondents report a job with a start date prior to the date of the last interview that was not reported during that particular round. The event history arrays created at the previous interview date are not changed to include information about these new jobs. Instead, three variables are available to alert users to changes that would have resulted if the jobs had been correctly reported during the previous interview. The first variable, EMP_BK_WKS, tells how many weeks before the previous interview date the job started.

The second and third variables (with the titles EMP_BK_STATUS and EMP_BK_HOURS) show how the status and hours arrays would have been affected had the job beginning before the date of last interview been reported at the prior interview and included in the original array construction.

Employment status variables are found in the database under question names beginning with "EMP_."

Marital status. These variables cover the respondent's marital or cohabitation status during each month from his or her 14th birthday to the month of the most recent interview. Possible status labels include the following: Never married and not cohabiting, never married and cohabiting, married, legally separated, divorced, or widowed.

A second marital status variable combines the status with the total number of spouses/partners; a code of 100 indicates that the person living in the household during that month is a partner, while 200 denotes a spouse. The last digit of this variable corresponds to the total number of partners or spouses. For example, 102 would be two total partners, and 202 would denote the second spouse. This allows the data user to identify people who live with the same spouse or partner for two or more different periods separated by time spent living with a different spouse or partner. Researchers also can link the spouse/partner in the arrays with that individual's characteristics through the MAR_PARTNER_LINK variable and the PARTNERS_ID variable found in the PARTNER roster. There is also a deny variable. Marital status variables are found in the database under question names beginning with "MAR_."

Program participation status. For each month since the respondent's 14th birthday, these variables report the respondent's receipt of economic assistance. Program participation arrays are constructed individually for three need-based programs: Aid to Families with Dependent Children (AFDC), Food Stamps, and the

Women, Infants and Children (WIC) program. The AFDC array includes all Federal and State programs created under the Temporary Assistance to Needy Families (TANF) program or any government assistance program for low-income families that replaces AFDC. All other need-based programs, such as Social Security Income (SSI), are combined into a fourth program participation array entitled “Other.” In addition, arrays are available for two employment-based programs. Unemployment insurance is included in all rounds and workers’ compensation is included in rounds 1 through 3. A deny variable also exists here.

Program participation status variables can be located by searching for question names beginning with “WKCOMP_,” “UNEMP_,” “AFDC_,” “FDSTMP_,” “WIC_,” or “OTHER_.”

Schooling experiences. Unlike the other sections, this one presents some of the information on a yearly basis, beginning with each youth’s date of birth. For each year, the schooling variables provide data regarding:

- respondent’s grade in school
- number of times respondent changed schools
- number of months during which respondent did not attend school
- summer classes attended
- whether respondent repeated or skipped a grade
- number of times respondent was suspended

Monthly schooling event history variables, which provide information about the respondent’s educational status for each month from the round 2 interview to the current interview date, also are available. The three monthly arrays report the respondent’s enrollment status, the type of school attended during the month, and the ID code of the school. Because the same ID codes are used in the monthly arrays and on the NEWSCHOOL roster, users can link the monthly arrays with information collected in the schooling section of the interview. Finally, a “dual school” variable flags the small number of respondents who attended more than one school during the same month. There is only one dual school variable for the entire period; the specific

month of the overlap is not reported.

Schooling experience variables can be located by searching for question names beginning with “SCH_.”

Geocode data

These variables, which provide detailed statistics for each respondent’s county of residence, are not available in the main/event history database. Due to the confidential nature of these data, completion of a thorough application process and confidentiality agreement is required to obtain access to them. (See the end of this article for more information.) All of the variables described in this section have question names that begin with “GEO_.”

In addition to all main file and event history data, the geocode data set provides a list of the counties in which respondents lived between interviews. The unemployment rate for the respondent’s metropolitan area or State also is reported.

Economic and labor force characteristics are represented by geocode variables for the size of the county’s civilian labor force, the percent of the labor force employed in various industries, and the percent of workers aged 16 and older with jobs outside their county of residence. Income variables include per capita money income, per capita personal income, and median family money income for the respondent’s county. The unemployment rate for the respondent’s metropolitan area or State also is reported.

The final group of variables on the geocode CD focuses on colleges attended by the respondents. Survey staff use information from the Integrated Postsecondary Education Data System (IPEDS) to provide users with the identification code (UNITID) and State of each college attended by the respondent. For round 7, information on the colleges to which a respondent applied also is available. The identification codes can be used to associate the NLSY97 respondent’s college with various characteristics of the institution contained in the IPEDS database.

Additional variables: Transcript survey and ASVAB scores

This data release includes new high school transcript data collected in 2004 from the high schools of 4,815 youths who had graduated or who had reached age 18 and

were no longer attending high school. (This was the second wave of the NLSY97 transcript collection; the first wave took place in 1999-2000.) High schools provided a copy of the student’s transcript, a course catalog (if available), and information on the student’s enrollment in special programs (such as bilingual education, special education, or gifted/talented programs). The resulting variables offer researchers a more complete picture of a respondent’s school experiences.

Additionally, the latest data release contains a summary percentile score variable created from the computer-adaptive version of the ASVAB (CAT-ASVAB). Most round 1 respondents participated in the CAT-ASVAB, which measures knowledge and skills in 12 topical areas. Four of these areas—Mathematical Knowledge, Arithmetic Reasoning, Word Knowledge, and Paragraph Comprehension—were used in the creation of a single score variable.

Researchers should note that, while this score is similar to the Armed Forces Qualification Test (AFQT) score generated by the U.S. Department of Defense, this created variable reflects work done by the National Longitudinal Surveys (NLS) program staff and is neither generated nor endorsed by the Department of Defense.

Obtaining the event history and geocode data

NLSY97 event history data are available for free as a download at www.bls.gov/nls. Users also have the option of purchasing the public data on CD-ROM for \$20. The data set contains the record for each youth, including all information in the main file and the event history variables described above. The data file also includes Windows-based search and extraction software and complete codebook documentation on each variable. A downloaded data file contains exactly the same data that the CD contains.

To aid researchers in using the data, each data set is accompanied by the *NLSY97 User’s Guide*, which examines the data set in detailed topical sections; the codebook supplement; and an electronic copy of the round 7 questionnaire. Other supplemental documentation items, such as additional questionnaires, are available for purchase. Researchers can obtain NLSY97 event history CDs and documentation from NLS

User Services. Some documentation items also are available for download from www.bls.gov/nls or within NLS Web Investigator at the same Web site.

Because the NLSY97 geocode data set contains confidential data, researchers interested in obtaining the geocode CD must complete the accessing agreement procedure required by the U.S. Bureau of Labor Statistics (BLS). This process includes filling out an application and signing a confidentiality agreement. Allow 4 to 6 weeks for processing the application. For more information, or to receive an application for access, see the NLSY97 section of the NLS Web site or contact NLS User Services or Rita Jain at BLS. (See the back cover of this newsletter for contact numbers.)

Like the event history CD, the geocode disc is accompanied by the *NLSY97 User's Guide* and an electronic copy of the round 7 questionnaire. Other available documentation includes a geocode codebook supplement containing the codes for the various geographic areas. Geocode data files are not available for download, although a public-use version of the geocode codebook supplement will be available on the order page so that researchers can evaluate the potential uses of the data set. □

SSA Records Update Mortality Information for NLSW Respondents

In 2003, the U.S. Census Bureau, through the cooperation of the Social Security Administration (SSA), obtained SSA mortality data (variables R76151.00-R76153.00) for those respondents from the mature women and young women cohorts who had shared their Social Security numbers with the Census Bureau. A comparison of the SSA data with the NLS survey data showed a substantially higher number of deceased respondents than was originally reported. Therefore, the 'reason for noninterview' variables created in each NLS survey year needed to be revised to show the higher numbers in the 'deceased' category (as well as the consequently lower counts for other 'reason for noninterview' categories).

A 'Reason for Noninterview—Revised/Deceased' variable was created for each

survey year from 1970 on, using the SSA information (as an example, see the young women cohort variable R01453.20 in 1970). Coding changes for a deceased respondent were made for the first survey year following the respondent's reported year of death. A revised sampling weight for each year also was created. (See R01453.30 in 1970.) Researchers should use these updated variables rather than the original 'reason for noninterview' and 'sampling weight' variables.

As a general rule, whenever there is both an original version and a revised version of a variable, users should plan on using the revised variable in their analyses, as the revised variable usually has been created to reflect new information or a data correction. □

New Topics in the NLS Women Surveys

The 2003 NLS women surveys (the final collection year for the women surveys) debuted several new questions covering topics that might be of special interest to the research community.

Exercise frequency. Both the young women and mature women cohorts were asked, "During the last month, did you get at least 30 minutes of exercise three or more times a week?" Nearly 65 percent of the young women and 63 percent of the mature women reported that they had exercised.

Prescription expenses. A series of questions was introduced for both cohorts regarding prescription drug use and affordability. Respondents who reported taking prescription drugs were asked whether they took all the medication they were prescribed and, if not, the reason(s) why not. They also were queried on per month out-of-pocket expenses for their prescriptions. More than 86 percent of the mature women and nearly 71 percent of the young women reported taking prescription drugs.

Mental health medication. Respondents in both surveys were asked, "Do you now take antidepressants, tranquilizers, or pills for nerves?" In the young women survey,

16 percent reported taking these types of medication; close to 14 percent of the mature women reported usage.

Pet ownership. Respondents were asked a series of questions on pet ownership, including the number of and type of pets (dog, cat, fish, and so forth). The majority (53 percent) of young women respondents reported pet ownership, while fewer than one-third (32 percent) of the mature women reported any pets. "Dog" was listed as the most popular pet by both cohorts.

Income adequacy. Both women cohorts were asked their opinion of the amount of money needed to make ends meet and the amount of money spent for essentials. They also were asked to describe how they felt about their income on a seven-point scale from "extremely happy" to "extremely unhappy." The majority of the women questioned indicated that they were at least "somewhat happy" about their income.

Attitudes toward possible Social Security reforms. Respondents in the young women cohort were asked a series of questions regarding their feelings about hypothetical changes in the current Social Security program, including investment and payment options. (These questions were not asked of the mature women.) More than 50 percent of the young women indicated that, if they were a 30-year-old worker today, they would opt to replace part of their Social Security with a personal retirement plan. These Social Security questions are found in the "Attitudes" section of the 2003 young women questionnaire.

The young women and mature women data sets are available on CD or through Web Investigator at www.bls.gov/nls. □

A Question's Universe: Understanding Who Gets Asked What

As is the nature of questions in complex survey instruments, not every question in an NLS survey is asked of every respondent. Respondents skip questions or even, in many cases, entire sections. In addition, NLS data sets contain not only the ques-

tions respondents were asked but also a host of other variables that may be confusing to the untrained eye. Here are some helpful tips when “decoding” a question’s universe:

Study the questionnaire. In essence, the questionnaire is the “conversation” of the interview. Whenever possible, look at where a question fits into the actual questionnaire. Each question entry on Computer Assisted Personal Interview (CAPI) questionnaires will provide hyperlinks to those questions that lead into that entry and the question that follows next by default, along with any branching. If there is still some ambiguity about the question’s universe, trace the skip pattern back to the beginning of a section, because all respondents (as a general rule) pass through the first item of each section (and, thus, the first item in a section usually contains data for all valid interviews). Skip patterns may be straightforward, as when several questions are asked sequentially, or more complex, as when one or more questions in a sequence are not asked because they do not apply.

Two basic types of variables have been compressed out of the questionnaire to provide a more compact and readable version: Question records that are not necessary for understanding of the flow of the survey (such as internal function checks); and looped or repetitive questions (only the first loop appears). Questionnaires for all of the cohorts now are available electronically on Web Investigator at www.bls.gov/nls (through the individual cohorts links).

Understand the different types of variables. Within the NLS data sets are several types of variables. Some variables result from questions that are asked directly of the respondent, such as “Did you receive any income from service in the military?” or “When was the last time you visited a healthcare professional?”

A special type of “direct question” variable is the looped variable. A looped (or repetitive) variable results from a series of questions that gets asked multiple times. For example, a set of questions might be asked for each child, each household member, each job, or each school attended. There may also be nested loops (loops

within loops). In the NLSY97, many of these loops have been condensed into roster items, which can be found in Web Investigator by choosing “roster” as the “area of interest.” Codebook entries exist for each of the loops of variables; however, the questionnaire shows only the first pass through.

Other variables represent information provided by the interviewer (“In what language was this interview conducted?” for instance). Still other variables are “check items” that have been filled in by either the interviewer or the computer. These checks often occur at the beginning of a section or question series and rely on information provided in previous surveys or in an earlier part of the interview. For instance, a respondent’s reported marital status will be used as a check item at the beginning of some sections. Novice researchers will sometimes mistakenly choose a check item variable for analysis because check item titles may closely resemble other variable titles. Using a check item for analysis is not advisable because these items are designed primarily for questionnaire flow. Generally, a check item’s variable title will contain either “CHK,” “CK,” or “CHECK.”

Another type of variable, the created variable, is an edited or constructed variable that is generated after a survey has been fielded. A created variable will be labeled “created” or “CV,” depending on the cohort. A variable’s codebook entry will usually indicate whether it was created. “Number of weeks worked since last interview” is one example of a created variable. Users are encouraged to use created variables for their analyses wherever possible.

Eyeball the codebook frequencies. Common sense prevails here. If, for instance, a researcher selects a variable for which the total number of yes/no responses seems too low for the number of respondents interviewed that year, chances are that either an incorrect variable has been chosen for analysis or multiple instances of the question need to be combined across different universes to get a grand total. Researchers should pay special attention to the valid skip frequencies listed for each codebook entry.

Use the User’s Guide. Each cohort has a

cohort-specific *User’s Guide* with a large topical guide section. While specific skip patterns are not usually itemized here, information about the content of the different sections can prove beneficial in understanding a respondent’s path through different series of questions. As with the questionnaires, user’s guides are available within NLS Web Investigator at www.bls.gov/nls.

Contact NLS User Services: If a researcher still needs clarification about the nature and universe of a question, NLS User Services is always available to give guidance. NLS User Services can be reached at usersvc@chrr.ohio-state.edu or by calling (614)442-7366. □

Frequently Asked Questions

The staff of NLS User Services encourages researchers to contact them with questions and problems encountered while accessing and using NLS data or documentation. Every effort is made to answer these inquiries in a timely manner. Some recently asked questions that may be of general interest to NLS users are listed below with their answers.

Q1. I am trying to compute total costs of college, total borrowing, and total grants/scholarships for all years as of the 2002 interview in the NLSY97. I am assuming that each of the questions in a survey year regarding tuition (or loans or grants) for a particular college and term are relates only to the period since the date of the last interview. Is there any created variable that keeps a running total of these variables—or should I simply add them up?

A1. You will need to add up the variables, as there are no created variables that provide the total costs of college. A cautionary note: the college loop information is meant to be collected for all terms completed since the date of last interview. It is possible, however, for respondents to mistakenly provide information about a term that is still in progress. This term might then be re-reported in the following interview. So it is always a good idea to compare the actual start dates of the terms to make sure that information is not being duplicated.

Q2. For the NLSY79, what is the best way to determine if the respondent is living with a partner, but unmarried? There seem to be several possibilities for this.

A2. Probably the most consistent information on the presence of a partner in the household at the time of the interview comes from variables with the title RELATIONSHIP CODE OF CURRENT SPOUSE/PARTNER in the "Fertility and Relationship History" area of interest. These variables indicate whether it is a spouse or a partner being reported in the household. You may also want to look at the associated variables titled NUMBER OF SPOUSE/PARTNERS REPORTED. In this variable, a unique identification number is assigned to each spouse and partner. The identification number remains consistent throughout the survey years and allows researchers to trace the same spouses and partners over time.

Q3. Using the Mature Women and Young Women cohorts, I hope to determine how many job changes (on average) a woman has throughout her working career. Is this possible?

A3. You will find the employer record is complete beginning in 1992 for the Mature Women and in 1993 for the Young Women. In the early years, however, employer information was not collected in an event history format. The survey rounds in the early years are better at picking up occupation change within an employer than at picking up information on all employers. You can read more about this in the "Work Experience" section (Section 4.29) of the Mature Women user's guide, available through the BLS Web site at www.bls.org/nls under "Publications and Other Documentation." You also may want to search the NLS bibliography Web site (www.nlsbibliography.org) for references that have previously looked at job change issues in these cohorts. □

Completed NLS Research

The following is a list of 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 online at

www.nlsbibliography.org for a comprehensive listing of the NLS-related research.

Avellar, Sarah A. and Smock, Pamela J. "The Economic Consequences of the Dissolution of Cohabiting Unions." *Journal of Marriage and Family* 67, 2 (May 2005): 315-28. [NLSY79]

Bouffard, Leana Allen. "The Military as a Bridging Environment in Criminal Careers: Differential Outcomes of the Military Experience." *Armed Forces and Society* 31, 2 (Winter 2005): 273-95. [NLSY79]

Centeno, Mario. "The Match Quality Gains from Unemployment Insurance." *Journal of Human Resources* 39, 3 (Summer 2004): 839-63. [NLSY79]

Chapple, Constance L. "Self-Control, Peer Relations, and Delinquency." *Justice Quarterly* 22, 1 (March 2005): 89-106. [Children of the NLSY79, NLSY79 Young Adult]

Fairlie, Robert W. "Earnings Growth Among Young Less-Educated Business Owners." *Industrial Relations* 43, 3 (July 2004): 634-60. [NLSY79]

Grossman, Michael; Kaestner, Robert; and Markowitz, Sara. "Get High and Get Stupid: The Effect of Alcohol and Marijuana Use on Teen Sexual Behavior." *Review of Economics of the Household* 2, 4 (2004): 413-41. [NLSY97]

Heywood, John S. and O'Halloran, Patrick L. "Racial Earnings Differentials and Performance Pay." *Journal of Human Resources* 40, 2 (Spring 2005): 435-52. [NLSY79]

Kalil, Ariel and Ziol-Guest, Kathleen M. "Single Mothers' Employment Dynamics and Adolescent Well-Being." *Child Development* 76, 1 (Jan/Feb 2005): 196-212. [NLSY79, NLSY79 Young Adult]

Levine, Phillip B. and Zimmerman, David J. "Children's Welfare Exposure and Subsequent Development." *Journal of Public Economics* 89, 1 (January 2005): 31-56. [Children of the NLSY79, NLSY79]

Long, J. Scott and Pavalko, Eliza K. "The Life Course of Activity Limitations: Explor-

ing Indicators of Functional Limitations over Time." *Journal of Aging and Health* 16, 4 (August 2004): 490-517. [Mature Women, Young Women]

Lundquist, Jennifer Michelle Hickee and Smith, Herbert L. "Family Formation Among Women In The U.S. Military: Evidence From The NLSY." *Journal of Marriage and the Family* 67, 1 (February 2005): 1-13. [NLSY79]

Mimura, Yoko and Mauldin, Teresa A. "American Young Adults' Rural-to-Urban Migration and Timing of Exits from Poverty Spells." *Journal of Family and Economic Issues* 26, 1 (Spring 2005): 55-76. [NLSY79]

Park, Hye-Sook; Pearson, P. David; and Reckase, Mark D. "Assessing the Effect of Cohort, Gender, and Race on Differential Item Functioning (DIF) in an Adaptive Test Designed for Multi-Age Groups." *Reading Psychology* 26, 1 (January-March 2005): 81-101. [Children of the NLSY79]

Raffaelli, Marcela; Crockett, Lisa J.; and Shen, Yuh-Ling. "Developmental Stability and Change in Self-Regulation from Childhood to Adolescence." *Journal of Genetic Psychology* 166, 1 (March 2005): 54-75. [Children of the NLSY79]

Rowe, David C. "Under the Skin: On the Impartial Treatment of Genetic and Environmental Hypotheses." *American Psychologist* 60, 1 (January 2005): 60-70. [Children of the NLSY79]

Solberg, Eric J. "The Gender Pay Gap by Occupation: A Test of the Crowding Hypothesis." *Contemporary Economic Policy* 23, 1 (January 2005): 129-48. [NLSY79]

Steen, Todd P. "The Relationship Between Religion and Earnings: Recent Evidence from the NLS Youth Cohort." *International Journal of Social Economics* 31, 5/6 (2004): 572-82. [NLSY79]

Zietz, Joachim and Joshi, Prathibha. "Academic Choice Behavior of High School Students: Economic Rationale and Empirical Evidence." *Economics of Education Review* 24, 3 (June 2005): 297-308. [NLSY97] □

FIRST CLASS MAIL
Postage and Fees Paid
U.S. Department of Labor
Permit No. G-738

U.S. DEPARTMENT OF LABOR
U.S. Bureau of Labor Statistics
Postal Square Building, Rm 2850
2 Massachusetts Ave., NE.
Washington, DC 20212-0001

Official Business
Penalty for Private Use, \$300
Address Service Requested

NLS News



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@postoffice.chrr.ohio-state.edu (614) 442-7366 (614) 442-7329 (Fax)	NLS documentation, data, and data updates:	usersvc@postoffice.chrr.ohio- state.edu
		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)	<i>NLS News</i> Editor:	Donna S. Rothstein Rothstein_D@bls.gov
		NLS Program Director and Media Contact:	Charles R. Pierret Pierret_C@bls.gov (202) 691-7519