Transcript Data Outline—NLSY97 Respondent High School Experiences

In winter 1999-2000, the NLSY97 transcript survey sought specific educational information directly from high schools about all NLSY97 respondents who had graduated from high school or who were aged 18 or older and no longer enrolled. During this first wave of the transcript collection, the survey staff received 1,417 transcripts. Additional waves are planned as more respondents graduate from or leave high school.

Transcript survey data collection
A transcript request packet was mailed to each school at which an NLSY97 youth had received his or her high school diploma, or to the last school that the nonenrolled youth had attended. The packet contained informational materials about the NLSY97, a description of the NLSY97 transcript survey, and documentation of administrative permission from districts to contact schools. The packet also included a cover letter addressed to the school principal, a one-page cover sheet questionnaire designed to collect information on school-specific grading and transcript policies, a student request list identifying the sampled students in the school, and the signed permission forms for these students. Reproductions of the principal letter, one-page questionnaire, student request list, and permission forms may be viewed in appendix 11 of the NLSY97 Codebook Supplement.

Using course catalogs, transcript data, and clarification calls to school administrators, survey staff constructed histories of courses taken and term enrollment calendars for each of the 1,417 youths. Courses were coded using the Revised Secondary School Taxonomy (SST-R), a hierarchical framework for high school course offerings. For information about the coding of the SST-R variables, researchers should contact NLS User Services. (Contact information is provided on the back of this newsletter.) Data files also include any available information on absences, instances of tardiness, dates of enrollment, and high school graduation status, as well as indicators of participation in special education classes, gifted/talented programs, or bilingual education. If the files included scores on achievement tests, such as the ACT, PSAT, SAT, SAT II, or AP tests, those scores are reported.

Organization of transcript data
Transcript survey data for the initial 1,417 respondents are available on the round 3 NLSY97 event history CD-ROM. All of the variables have question names that begin with “TRANS_.” Similar transcript data will be added to the data set in future rounds as additional respondents leave high school.

The transcript data files on the CD-ROM are organized into four types of information—school, student, term, and course variables.

School-level variable. The variable TRANS_SCH-CAT indicates whether a course catalog was received from the school to aid in coding. This variable also functions as an identifying number of the school for the rest of the transcript variables, although the number does not link to any variables in the main data file. Because the largest number of schools reported for any respondent is 12, the variable is repeated 12 times. The school that provided the transcript is always listed as school #01. Because some transfer schools could not be identified from transcript records, a flag in the school file indicates “dummy” schools for which no additional data are available.

Student-level variables. Containing one observation for each respondent for whom a transcript was processed, these data include student-specific information not associated with a specific term. Examples include standardized test scores, absence and tardiness records, the student’s completion status, and dates of enrollment.

Term-level variables. A number of variables refer to the respondent’s terms of enrollment. For up to 18 terms, these items report the beginning and ending dates of the term (TRANS_TERM_START_DATE.xx, TRANS_TERM_END_DATE.xx), the way in which the school year is divided (TRANS_TERM_SEASON.xx), the academic year of the term (TRANS_TERM_YEAR.xx), the respondent’s grade level in that term (TRANS_TERM GRADE.xx), and the number of credits earned (TRANS_TERM_CREDIT.xx). A variable indicating the school that the respondent attended during that term, TRANS_TERM_SCH NU.xx, permits term-specific variables to be linked to the course catalog variable described in the school section above.

Course-level variables. The course data include information about each course appearing on a student’s high school transcript. Variables associated with each course include characteristics of the term in which the course was taken: Term start and end dates, the student’s grade level during the term, the school at which the student took the course, and the total credits earned by the student that term. Course-specific variables include the course code.
from the SST-R, the grade earned in the course, and the credit value of the course. Because schools use many different grading systems, the course grades were converted into a standard scale that can be compared across respondents. Credits earned, however, are listed in the units provided by the school and are not necessarily comparable across schools.

Transcript data on CD-ROM
NLSY97 transcript data are available to the public on the event history CD-ROMs for NLS97 rounds 1 through 3; the CDs cost $20 apiece. Each CD contains the data record for each youth, including all information in the main file, a set of event history variables, Armed Services Vocational Aptitude Battery (ASVAB) data, and the transcript variables described above. The data file also includes Windows-based search and extraction software and complete codebook documentation on each variable.

To aid researchers in using the data, each disc is accompanied by the NLSY97 User’s Guide, which examines the data set in detailed topical sections. Researchers using the transcript data also should obtain the round 3 NLSY97 Codebook Supplement. Appendix 11 in that document provides more details on the collection, coding, and data entry and processing procedures for the NLSY97 transcript survey. It also includes explanatory notes for specific transcript variables. Other supplemental documentation items, such as the questionnaires, are available at an additional charge. Researchers can obtain NLSY97 event history discs and documentation from NLS User Services. (Contact information is provided on the back of this newsletter.)

Transcript data in other NLS cohorts
Transcript surveys also have been conducted for the NLSY79. These surveys have gathered information on course subject matter, enrollment dates, and grades earned. Achievement test scores were collected during special transcript or school surveys for the NLSY79, young women, and young men cohorts. Scores were reported by the schools or coded from the respondent’s transcript. For more precise details about the content of each survey and availability of the data, users should consult the NLS Handbook or the appropriate cohort’s user’s guide.

Geographic Information and Other Restricted Data in the National Longitudinal Surveys

To protect respondent confidentiality, the National Longitudinal Surveys (NLS) public-use files do not include geographic variables such as State, county, metropolitan area, zip code, or census tract. Such variables, when combined with the rich longitudinal records of respondents’ labor market experiences and significant life events, would create an unacceptably high risk that someone could use the data to reidentify individual respondents and possibly violate their privacy. For that reason, geographic variables that are more detailed than the four census regions (Northeast, North Central, South, and West) are not included on the public-use files for the NLSY79 and NLSY97. In the public-use files for the original cohorts (older men, young men, mature women, and young women), only variables indicating South/non-South region of residence are available.

Although detailed geographic variables are not available on the NLS public-use files, such variables are available for researchers to use on a more restricted basis. NLSY79 and NLSY97 data fitting this description are available on geocode CDs or from the Bureau of Labor Statistics (BLS). Mature women and young women data of this type are available from the U.S. Census Bureau. The following sections describe how researchers can obtain access to detailed geographic variables, as well as other restricted-use data sets, in the NLS program.

State, county, and metropolitan area variables in the NLSY79 and NLSY97
For the NLSY79 and NLSY97, BLS has established a licensing system through which legitimate researchers at universities and other research organizations can use NLSY “geocode” data at their own facilities, provided that the research project and physical and electronic security measures are approved by reviewers at BLS. The geocode files for the NLSY79 and the NLSY97 include the State, county, and Metropolitan Statistical Area (MSA) of residence for each respondent in each survey year, as well as selected environmental variables from the Census Bureau’s County and City Data Books, and selected exact local unemployment rates from the BLS publication Employment and Earnings. The NLSY97 geocode file also includes data on migration histories of the respondents, and on institutions of higher education they have attended.

Researchers interested in obtaining NLSY geocode files must complete a BLS geocode application, in which they describe the objectives of the research project, the reason for needing access to the geocode variables, and the security measures that are in place at the researcher’s facilities. If the application is approved, a senior executive at the researcher’s organization must sign an agreement with BLS that legally obligates the researchers and the organization to adhere to the terms established by BLS to protect the confidentiality of survey respondents. The key provisions of the agreement are listed below:

- All work under the agreement must be performed at the institution’s facilities in a location specified in the agreement. No work may be performed at an individual’s home or other offsite locations or in any public area in which unauthorized persons could obtain access to the confidential data.
- The recipient agrees to allow employees or agents of BLS to have access to the recipient’s facilities, if requested, for the purpose of reviewing the recipient’s adherence to the confidentiality and security provisions of the agreement.
- The recipient agrees not to divulge, publish, reproduce, or otherwise disclose, orally or in writing, any confidential information to any individual other than authorized persons.
- The recipient agrees to store all documents, compact discs, diskettes, tapes, DVDs, laptop computers, or other storage media that contain confidential NLSY information in a locked receptacle that can be accessed only by the authorized persons named in the agreement.
The recipient agrees not to subcontract, by electronic or physical means, to the geocode data file and electronic outputs created from it. The geocode data file and these outputs may be copied to and stored on a network server, mainframe computer, desktop computer, laptop computer, compact disc, diskette, tape, DVD, or other storage media and may be accessed by modem or other electronic communication devices, provided that the information is protected by password or other secure means to prevent unauthorized access. The data cannot be accessible via the Internet.

The recipient agrees not to attempt to link the geocode data file with individually identifiable records from any other BLS or non-BLS data set, other than the NLSY public-use file. For example, linking the NLSY geocode file with a State’s list of licensed drivers would not be permitted. In contrast, linking the NLSY geocode file with a file showing variations across States in public-assistance programs would be permitted.

The recipient agrees not to use the NLSY geocode data file for the purpose of identifying specific respondents in any way. If the identity of a person is inadvertently discovered, the recipient will not use this knowledge and will hold the identity of the person in confidence.

The recipient agrees to notify the BLS project coordinator immediately upon discovering any breach or suspected breach of security or any disclosure of confidential information.

The recipient agrees to notify the BLS project coordinator immediately upon receipt of any legal, investigative, or other request for access to confidential information that the recipient has stored on a network server, mainframe computer, desktop computer, laptop computer, compact disc, diskette, tape, DVD, or other storage media must be returned to the BLS project coordinator or, with the BLS project coordinator’s permission, destroyed.

If the recipient needs to keep the confidential information beyond the expiration date of the agreement, the recipient’s project coordinator may request a 1-year extension from the BLS project coordinator. Such requests must be approved in advance and in writing by the BLS project coordinator.

All researchers involved in the project must read, sign, and return to BLS a copy of the BLS nondisclosure affidavit before they will be permitted to have access to the geocode data. Project coordinators also must read and sign a checklist that includes 10 statements related to the security of the NLS geocode data.

Although the NLSY geocode application process sounds time consuming, the application itself is fairly short, only 12 items altogether. The longest item on the form, a project description, rarely exceeds more than four or five paragraphs. BLS requires simply that the project description clearly states the research objectives, methodology, and need for the geocode variables. The other items in the application, which pertain to the physical and electronic security measures in place at the applicant’s institution, generally are brief and straightforward as well. The best way to ensure speedy approval of an NLSY geocode application is to formulate it clearly and concisely.

BLS tries to strike a balance between providing researchers with access to data and protecting the confidentiality of survey respondents. Maintaining this balance has become increasingly challenging in an age in which advances in computer technology make it easier to reidentify respondents. BLS is not aware of any cases in which NLS researchers have used restricted data intentionally for any improper purposes that would violate respondent confidentiality, and frankly, we doubt that that would ever happen. A greater concern is that NLS geocode recipients may be careless in their handling of the data, making it possible for unauthorized persons to gain access to the data and misuse it.

Staff at BLS, the Center for Human Resource Research (CHRR), and the National Opinion Research Center (NORC) have a legal obligation to ensure the confidentiality of NLSY79 and NLSY97 respondents; Census Bureau staff have a similar obligation to mature and young women respondents. All recipients of geocode data onsite or users of restricted data have the same legal responsibility to protect respondent confidentiality. Respondents voluntarily agree to participate in the NLS because of assurances that their confidentiality will be protected. Naturally, we do not want to give respondents a reason to doubt those assurances because of the potentially delerious effect on survey response rates. Failure to comply with all the terms of an NLSY geocode agreement or the restrictions associated with onsite access may jeopardize a recipient’s future access to any confidential BLS data.

BLS recently has adopted a policy not to enter into new NLSY geocode agreements, or extend existing agreements, with researchers and institutions outside of the United States. For NLSY geocode recipients in the United States, we have legal remedies to prevent willful misuse of the data, and we more readily can inspect the facilities of geocode recipients periodically to ensure that the data are protected adequately from unauthorized access. For researchers outside of the United States, we have no way to enforce U.S. privacy laws, and it simply is not feasible to inspect the facilities of researchers abroad. Although BLS has adopted this policy very reluctantly, it simply brings BLS in line with the longstanding policies of other U.S. statistical agencies, such as the Census Bureau and the National Center for Education Statistics.

To obtain an electronic copy of the NLSY geocode application or to ask questions about the application process, interested researchers should e-mail the
NLS program staff at BLS at NLS_Info@bls.gov.

Zip code and census tract variables and other confidential NLSY79 and NLSY97 data
Some research projects require data with more geographic detail than provided in the NLSY geocode files, such as information about census tracts and zip codes.

BLS has opportunities available on a limited basis for researchers from colleges and universities, government, and eligible nonprofit organizations to obtain access to confidential BLS data files for exclusively statistical purposes. The review process to obtain access to such data is considerably more lengthy than it is for the NLSY geocode program, and researchers can use the data only at BLS headquarters in Washington, DC.

College and university faculty and researchers in government and eligible nonprofit organizations can qualify for opportunities to use confidential NLSY data files for statistical research under terms of the Intergovernmental Personnel Act. Eligible graduate and undergraduate students may obtain access to confidential microdata for statistical research under the BLS Student Volunteer program. Persons wishing to inquire about more detailed eligibility requirements should contact BLS.

Researchers coming to BLS to perform statistical research will not be compensated by BLS for the work or any associated expenses, such as traveling or living expenses. In order for a researcher to participate in this program, the eligible institution with which he or she is affiliated must agree to enter into a written agreement with BLS. All published outputs from such research are in the public domain.

For additional information on eligibility requirements and procedures for obtaining access to zip code and census tract variables and other confidential data from the NLSY79 and NLSY97, e-mail the National Longitudinal Surveys program staff at BLS at NLS_Info@bls.gov.

Geographic variables in the NLS original cohorts
Data for the mature women and young women cohorts are collected for BLS by the Census Bureau. The Census Bureau also collected data for the older men and young men cohorts before those surveys were discontinued. All four survey groups are known collectively as the NLS original cohorts. The process for obtaining access to geographic variables that are not available on the public-use files for the original cohorts involves review at both the Census Bureau and BLS.

The public-use files for all of the original cohorts include variables indicating South/non-South residence. Information on State of residence also is available for researchers to use, but only at the Census Research Data Centers that have been established in several locations around the United States. In the near future, BLS and the Census Bureau hope to be able to provide researchers with access to variables on respondent county and metropolitan area, and possibly even finer geographic detail, such as census tracts. Again, researchers would be permitted to have access to these variables only at the Census Research Data Centers.

Geographic variables not currently on the public-use files may be requested for any of the original cohorts. Proposals in which researchers request access to such variables are reviewed on a case-by-case basis.

The approval process established by BLS and the Census Bureau for projects involving the original cohorts includes two main steps:

First, researchers must e-mail a description of the project to BLS at NLS_Info@bls.gov. The application should describe the research hypotheses and methodology, the geographic information needed for the project, and the way in which that geographic information will be used. BLS staff will review and comment on the proposal and possibly ask follow-up questions. When BLS staff has all the information it feels necessary, it will recommend whether to approve the project.

Second, if BLS approves the project, it will notify the researcher and the Census Bureau. The researcher then can submit a formal proposal to the Census Bureau’s Center for Economic Studies (CES), which oversees the Research Data Centers. Interested researchers should check the CES Web site at http://www.ces.census.gov/ for general proposal guidelines and information on the locations of Research Data Centers. Researchers also should work closely with CES staff and the administrator of the Research Data Center at which the researcher plans to work to develop the proposal adequately. CES staff can be contacted by e-mail at proposals.household@ces.census.gov. The names and addresses of Research Data Center administrators may be found on the CES Web page at http://www.ces.census.gov/ces.php/rdc#rdc.

Frequently Asked Questions
NLS User Services encourages researchers to contact them with questions and problems they have encountered while accessing and using NLS data and documentation. Every effort is made to answer these inquiries. Some recently asked questions that may be of general interest to NLS users are listed below with their answers.

Q1: How does the NLSY97 ascertain the race and/or ethnicity of respondents?
A1: In the initial screening of households selected for sampling for the NLSY97 survey, a household resident aged 18 or older was selected as the household informant. The household informant was asked questions that determined basic demographic information for all household members, including NLSY97-eligible youths. One of these questions asked the household informant whether each household member was Hispanic, Latino, or of Spanish origin.

The household informant also identified household members’ race. Informants could choose from the following categories: White; Black or African American; American Indian, Eskimo, or Aleut; Asian or Pacific Islander; or something else (specify). Although these were the only categories included in the survey questions, review of the verbatim responses in the “something else” category revealed that a large number of household residents were classified by the informant as either of mixed race or as Hispanic, with no other race specified. These two categories were added to the data set.

Users may prefer to use a variable, created by NLSY97 survey staff, that combines information on race and ethnicity for NLSY97 respondents (but not for other household members). This combined variable, KEYRACE_ETHNICITY, uses the race and ethnicity questions in the screener, information about the race and ethnicity of...
A2: The reason for the discrepancy is the source of the information on race and ethnicity. The race/ethnicity variable (R02147) was created by using information collected in the 1978 screener questionnaire. The person providing information about the household was the person who owned or rented that residence (or their spouse); this was often not the NLSY79 respondent. From this information, decision rules were implemented to designate each respondent’s race and ethnicity. In 1979, the respondents were asked for their ethnicity in variables R00096-R00102. Because of this, there is variability between information collected from the household in 1978 and information collected from the respondent in 1979. More information about the assignment of race and ethnicity is provided in the “Race, Ethnicity, and Nationality” section of the NLSY79 User’s Guide.

Q3: Does the NLSY79 ask any questions about primary or preferred languages spoken in the respondents’ households?

A3: Yes, this information is available for most NLSY79 respondents. There are several sources of information.

In round 1 of the NLSY79 a parent questionnaire was administered to one of the youth’s biological parents. If no biological parent lived in the respondent’s primary household, another adult household member was selected to answer the questionnaire. As part of the parent questionnaire, responding parents were asked, “Do you now speak any language other than English at home?” If they answered yes to this question, they were asked to select, from a list, all languages spoken at home. Almost 90 percent of respondents have completed parent interviews, but youths who do not have a completed parent questionnaire will not have this information available about languages spoken at home.

During the round 1 survey period, approximately 80 percent of NLSY97 respondents participated in the administration of the computer-adaptive version of the Armed Services Vocational Aptitude Battery. When they took this military enlistment test, one section included a questionnaire that collected background information about the respondents. This “online questionnaire” asked respondents to report whether English was their primary reading language and their primary spoken language, and whether it was spoken in their home during childhood. Respondents could then report up to two other languages spoken at home.

In addition, interviewers record whether each youth and parent interview was administered in English or Spanish. Interviewers also can report whether lack of English proficiency was a problem during the interview. While this interviewer-provided information does not directly address the issue of at-home languages, it may give researchers some idea of the English proficiency of the respondent.

Q4: In the NLS surveys, is the respondent ID number stable throughout each wave/year? Is respondent #1012, for example, respondent #1012 all the way through?

A4: The respondent ID numbers do not change. Respondent #1012 will always be respondent #1012 across all rounds.

Q5: I am using data for the mature and young women. The values for interview status recode are slightly different from those used for the reason for noninterview. Which variable should I use to identify which women died in a given survey round?

A5: Reason for noninterview is the better variable to use.

Completed NLS Research

The following is a listing 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 at www.nlsbibliography.org for a comprehensive listing of NLS-related research.


Cleveland, Hobart Harrington; Jacobson, Kristen C.; Lipinski, John J.; and Rowe, David C. “Genetic and Shared Environmental Contributions to the Relationship between the Home Environment and Child and Adolescent Achievement.” Intelligence 28, 1 (2000): 69-86. [NLSY79 Children]


Davey, Adam; Shanahan, Michael J.; and Schafer, Joseph L. “Correcting for Selective Nonresponse in the National Longitudinal Survey of Youth Using Multiple Imputation.” Journal of Human Resources 36, 3 (Summer 2001): 500-519. [NLSY79 Children]

Dooley, David; Praise, JoAnn; and Ham-Rowbottom, Kathleen A. “Underemployment and Depression: Longitudinal Relationships.” Journal of Health and Social Behavior 41, 4 (December 2000): 421-436. [NLSY79]


Gordon, Rachel A.; and Chase-Lansdale, P. Lindsay. “Availability of Child Care in the United States: A Description and Analysis of Data Sources.” Demography 38, 2 (May 2001): 299-316. [NLSY79 Children]


McCardle, John J. and Hamagami, Fumiaki.
Several NLS surveys ask questions about inheritances. The available data would allow researchers to answer questions such as:

- Are parents leaving inheritances to their children?
- Do parents have any money to leave?
- Do parents feel the need to leave anything to their children?

Researchers can compare inheritance information with data provided on retirement income sources to get a clearer picture of how attitudes about inheritances and the ability to leave something to children might have changed over the years. Some data on inheritances or retirement income sources are available for every NLS cohort except the young men. The first questions were asked beginning in 1971 with the older men survey. In 1997, 1999, and 2001, extensive sections on transfers between respondents and their parents or children were included in the mature women and young women surveys.
NLS Contact Information

NLS News is published quarterly by the 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 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 documentation, data, and data updates: usersvc@postoffice.chrr.ohio-state.edu

NLS web site: http://stats.bls.gov/nlshome.htm

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

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

NLS Program Manager and Media Contact: Jay Meisenheimer  
Meisenheimer_J@bls.gov  
(202) 691-7409