Chapter 3: Guide to the Mature Women Data

This chapter provides some practical information about how NLS variables are collected, created, and arranged on the CD-ROM. An explanation of the hard copy and electronic documentation items available is also included. The first section describes the different survey instruments used to collect the raw Mature Women data during the field period and explains how question numbers have been assigned during the various survey years. Next, the guide discusses the primary types of NLS variables and the process by which each is assigned a reference number and title that serve to identify it throughout the NLS documentation system. The third section reviews the codebook—that is, the information about each variable contained on the CD-ROM—and the accompanying paper documentation. This discussion will help users understand how to interpret the various pieces of data presented in the NLS documentation system. Finally, this chapter gives researchers some basic instruction on using the search functions on the CD-ROM to find variables relating to the area of interest.

3.1 Survey Instruments and Related Documentation

The primary variables found within the main data set of each NLS cohort are derived directly from one or more survey instruments (e.g., questionnaires, household interview forms, etc.). This section provides information on the conventions utilized within the NLS documentation system to identify questionnaire items from the primary survey instruments.

Certain other documents, namely *Interviewer Reference Manuals* and flowcharts, provide background information on how specific survey instruments were administered or offer the researcher additional tools for working with a questionnaire. While not actually survey instruments, these additional documents are described within this section.

Survey Instruments

A unique set of survey instruments is used during each survey year to collect information from respondents. The term "survey instrument" refers to: (1) the questionnaires, which serve as the primary source of data on a given respondent, and (2) documents such as the household interview forms or household record cards that collect information on members of each respondent's household.

Note that while the source of the majority of variables in the main NLS data sets is the questionnaire or one of the other survey instruments, certain NLS variables are created either from other NLS variables or from information found in an external data source.

Household Record Cards

NLS surveys include the collection, during each interview, of information on the members of each respondent's household. These data were collected primarily through the "Household Roster" section of the questionnaire, which in turn relies upon information provided by Census personnel and found on the

separate *Household Record Cards.* Respondents were selected on the basis of a screening of sample households. Both the instruments used for the household data collection and the household screening instruments that were used to draw the samples of respondents are described below. Starting with the 1995 survey, the information traditionally collected in the *Household Record Cards* became a part of the CAPI survey instrument.

Household Screener & Household Record Cards: Prior to most interviews, Census interviewers complete or update information found on a *Household Record Card.* Part of this information is transferred during the main interview to the "Household Roster" section of the questionnaire. The first *Household Record Card* (LGT-1, dated 2/23/1966) is the screening instrument used to select the Mature Women respondents for interview. Information for this first card was gathered from any available household member, while respondents provided comparable information in subsequent surveys. Each *Household Record Card* (1) enumerates all persons currently living in the household; (2) records for each person: name, relationship to respondent, whether this person is considered a household member (*Current Population Survey* definition), marital status, birth date, and sex; (3) summarizes changes since the last survey in household composition; and (4) provides information on the respondent's current and/or permanent address and telephone number at the time of interview, as well as the names of people who will know how to contact the respondent at the time of the next interview.

Five versions of the paper *Household Record Cards*, each covering approximately three surveys, have been used. While information from these cards does not, in general, appear as variables within any of the data files, certain information present on the cards detailing each respondent's current household composition is transferred to the "Household Roster" section of the questionnaires. In addition, certain demographic variables as of the initial survey year, notably age, birth date, race, and sex, were derived from the 1966 household screenings. Users should consult each survey's *Interviewer's Reference Manual* for the specific instructions and definitions used to complete each card.

Questionnaires

There are separate and distinctly different questionnaires each survey year. Each questionnaire is organized around a set of topical subjects, the titles of which usually appear on either the first page of each section of the questionnaire or as page headers.

User Notes: The questionnaires are critical elements of the NLS documentation system and should be utilized by each researcher to ascertain the wording of questions, coding categories, and the universe of respondents asked to respond to a given question.

Each questionnaire collects two general types of information: (1) information on the actual interview (e.g., interview dates, times, and contact methods) and (2) information supplied by the respondents on various topics related to their work and life experiences. Each survey instrument is organized around core sets of questions: current labor force status, retrospective work history, attitudes, health, marital history, household composition, assets, and income. In addition, the interview schedules contain special sets of questions on a variety of topics specific to the particular stage of life: child care and fertility questions were asked in the early survey years, while recent surveys have emphasized retirement and pension plans.

Information Sheet

Information Sheets (or flap items), located within the questionnaires, were usually designed in such a way that the interviewers could fold the sheet out to the side of the actual questionnaire and refer to the items on the flap during the interview. Various information items from previous interviews were clerically entered by Census and used by the interviewer during the survey. These included information such as name of previous employer, date of previous interview, and marital status and place of residence at the time of previous interview.

The interviewer also transcribed information recorded in the questionnaire during the current survey onto the *Information Sheet*. The only current survey year item that a user would need from the flap was "current marital status," transcribed from the *Household Record Card* in certain survey years. Items referenced frequently during the interview were more conveniently located when transferred to the flap.

With the advent of CAPI in 1995, *Information Sheet* items are inserted directly into the appropriate questions by the computer. During the interview, the computer prompts the interviewer with, for example, the name of an employer inserted as part of the question. The hard copy flap is therefore no longer used.

Questionnaire Item or Question Number: The questionnaire item or question number is the generic term referring to the printed source of data for a given variable. A questionnaire item may be a question, a check item, or an interviewer's reference item that appears within one of the survey instruments. Each questionnaire item has been assigned a number or a combination of numbers and letters to help the user link each variable to its location in a survey instrument. In general, questionnaire items found in the NLS documentation reflect the printed question numbers in the questionnaire.

Four different designations were used within the documentation system to identify varying types of questionnaire items in PAPI interview years (1967–92), as depicted in Table 3.1.1. The question

number appears to the right of each variable description within the codebook. CD-ROM users can access variable titles and codebook information via the "Accessing Data by Question Number" function.

Table 3.1.1 Question Numbering Conventions in PAPI Surveys

Question:	Question Number	112E; 59E
Interview Check:	Check Item (CH)	CH J3; CH AA
Interviewer Reference Item:	Interviewer Reference (R)	123R; R4
Unnumbered Questions:	Page Number	PG1

In the vast majority of cases, the reference is to a specific question item found in the survey (e.g., 22F or 3B). The convention "CH" is used to identify interviewer check items that occur within the survey (e.g., CH B). Their purpose is to direct the interviewer to the next appropriate question. The convention "R" denotes a reference item (e.g., R2 or 12R). Typically, reference items are grouped in a section of the survey instrument called the *Information Sheet,* which contains information that interviewers frequently refer to during the course of an interview. Items designated "R" in the survey instruments are also designated "R" in the documentation. Finally, when an item does not include a question number, only the page number ("PG") of the questionnaire on which a particular item appears is identified (e.g., PG 1). The first page of most questionnaires contains unnumbered interview status information and transcribed *Household Record Card* information.

Beginning with the first CAPI survey in 1995, questions are numbered using a different system. In general, all questions asked of respondents during each section of the questionnaire begin with a unique three-letter code. This letter code is followed by one or more numbers and letters indicating the placement of the question within that section. For example, HEA-11, a question on cigarette use, is followed by HEA-12A through HEA-12D, four questions on alcohol use. Questions that are part of a loop—that are asked about each different employer, for example—contain the additional letter group "ARR." Check items begin with "CK" followed by the three-letter code for that questionnaire section and a single letter (or letter and number) indicating the order in which the check was performed during the interview—for example, CK-HEA-A precedes CK-HEA-B. Table 3.1.2 lists the questionnaire sections and the corresponding codes. There are a few questions and check items within each section that may not follow this general pattern; users will be able to identify these questions by looking at the numbers assigned in the printed questionnaire.

Table 3.1.2 Questionnaire Sections and Sample Question Numbers in CAPI Surveys

Questionnaire Section	Abbreviation	Question Number	Check Item
Introduction and Household Record	HRC	HRC-3A, HRC-15	–
Respondent's Work History (CPS)	RWH	RWH-20, RWH-22B	CK-RWH-A, CK-RWH-K2
Respondent's On Jobs Supplement	OJS	OJS-1-ARR, OJS-5C-ARR	CK-OJS-E-ARR
Respondent's Employer Sort	RES	RES-1-ARR-01, RES-3B	CK-ES-C
Respondent's Employer Supplement	RSP	RSP-38-ARR, RSP-148B-ARR	CK-RSP-H-ARR
Respondent's Gaps in Employment	GAP	GAP-2-ARR, GAP-1B	CK-GAP-B-ARR
Husband's/Partner's Work History (CPS)	HWH	HWH-1, HWH-27A	CK-HWH-C, CK-HWH-L3
Husband's/Partner's On Jobs Supplement	HOJ	HOJ-3-ARR, HOJ-7	CK-HOJ-B, CK-HOJ-F-ARR
Husband's/Partner's Employer Sort	HES	HES-2-ARR, HES-3B	CK-HES-C
Husband's/Partner's Employer Supplement	HSP	HSP-34A-ARR, HSP-53-ARR	CK-HSP-G-ARR
Husband's/Partner's Gaps in Employment	HGP	HGP-1-ARR	CK-HGP-A
Health	HEA	HEA-2, HEA-19C	CK-HEA-H
Income	INC	INC-19E, INC-21	CK-INC-C
Other Family Background	OFB	OFB-8, OFB-20F-ARR	CK-OFB-P
Education and Training	EAT	EAT-29	CK-EAT-E
Parents and Transfers (1997)	PAR	PAR-12	CK-PAR-C
Interfamily Transfers (1999)	IFT	IFT-1B, IFT-36-ARR	IFT_CK_L
Mobility	MOB	MOB-4, MOB-5B	–
Attitudes and Contact Persons	ACP	ACP-4E, ACP-5	–
Interviewer Remarks	ASG	ASG-8, ASG-10A	–

User Notes: Most question numbers remain constant across CAPI surveys, so users can easily find the same question in different survey years. However, the order in which the questions are asked may vary for a number of reasons. Therefore, questions may appear in the CAPI questionnaire out of numerical order. Users should be able to locate them by following the skip pattern information provided in both the printed questionnaire and the codebook, or by using the reference numbers, which are assigned in numerical order.

In PAPI interviews, there is no question number for created variables. In CAPI surveys, however, created variables are identified using assigned question numbers similar to those included in the questionnaire. Each created variable is given a question number that indicates the content of the variable. For example, the constructed hourly rate of pay variables are assigned the question numbers HROP-ARR-01 through HROP-ARR-07, indicating the rates for jobs one through seven. Created

variables in CAPI survey years usually include the letters CV in the question name; in addition, all variables that have the word *KEY* in their title are created.

Interviewer's Reference Manual/Field Representative's Manual

Each survey instrument that goes into the field is accompanied by an *Interviewer's Reference Manual* (or *Field Representative's Manual*), which provides Census interviewers with background information on the NLS, respondent location instructions, and detailed question-by-question instructions for coding/completing the questionnaire and *Household Record Cards*. Note that *Interviewer's Reference Manuals* do not always include all the actual questions.

Flowcharts

The questionnaires are lengthy and often present the researcher with the complex task of determining the universe of respondents asked a specific question. To assist in this task, flowcharts have been developed that graphically depict the skip patterns for some questionnaires. Hard copy flowcharts are available for some post-1977 surveys; comparable information for earlier questionnaires appears within the codebook under the heading "Universe Information." Starting with the 1995 CAPI surveys, skip patterns are part of the codebook documentation.

3.2 Types of Variables

Four types of variables are present in the Mature Women data files. The type of variable affects the title or variable description which names each variable and the physical placement of the variable within the codebook. Types of variables include:

1. Direct raw responses from a questionnaire or other survey instrument.
2. Edited variables constructed from raw data according to consistent and detailed sets of procedures (e.g., occupational codings, *KEY* variables, etc.).
3. Constructed variables based on responses to more than one data item either cross-sectionally or longitudinally and edited for consistency where necessary.
4. Variables provided by the Census Bureau or another outside organization based on sources not directly available to the user (e.g., characteristics of respondents' geographical areas).

> *User Notes:* In general, CHRR does not impute missing values or perform internal consistency checks across waves. Data quality checks most often occur in the process of constructing cumulative and current status variables such as 'Highest Grade Completed.'

Reference Numbers

Every variable within the main NLS data set has been assigned an identifying number that determines its relative position within the data file and documentation system. **Persons contacting NLS User Services should be prepared to discuss their question or problem in relationship to the reference number(s) of the variable(s) in question.**

Reference numbers, once assigned, remain constant through subsequent revisions of the files. Reference numbers are assigned sequentially, with variables from the first survey year having a lower reference number than those variables specific to the second year, and so forth. Occasionally, variables are created sometime after the year in which the data were actually collected. These variables are frequently given a reference number that reflects the year in which the actual data were gathered rather than the year the created variable was constructed.

Table 3.2.1 lists reference numbers for each survey year since 1967 for the Mature Women.

Table 3.2.1 Mature Women Reference Numbers by Survey Year

Survey Year	Reference Numbers	Survey Year	Reference Numbers
1967	R00001.–R00813.	1982	R05270.–R06640.
1968	R00833.01–R00868.	1984	R06650.–R07192.
1969	R00868.01–R01333.01	1986	R07203.–R07809.
1971	R01334.–R02047.	1987	R07820.–R08863.
1972	R02048.–R02876.	1989	R08865.–R10077.
1974	R02878.–R03067.	1992	R10080.–R13027.
1976	R03080.–R03273.50	1995	R16007.–R34923.
1977	R03280.–R04540.	1997	R34950.–R42503.
1979	R04542.–R04892.	1999	R42527.–R54394.
1981	R04900.–R05267.		

Variable Titles

Every variable within NLS main file data sets has been assigned an 80 character summary title that serves as the verbal representation of that variable throughout the hard copy and electronic documentation system. Variable titles are assigned by CHRR archivists who endeavor, within the limitations described below, to capture the core content of each variable and to incorporate within the title (1) common words that facilitate easy identification of comparable variables; (2) UNIVERSE IDENTIFIERS that specify the subset of respondents for which each variable is relevant; and (3) for

some variables, REFERENCE PERIODS that indicate the period of time (e.g., survey year or calendar year) to which these data refer. Universe identifiers and reference periods are discussed below.

Universe Identifiers: If two ostensibly identical variables differ only in that they refer to different universes, the variable title will include a reference to the applicable universe.

> Example 1:　　'Reason for Being OLF, 77 (Not Empld, Have Worked)'
> 　　　　　　　'Reason for Being OLF, 77 (Not Empld, Not Worked)'

Reference Periods: Variable descriptions may include a phrase indicating the time period to which these data refer. The following general conventions apply:

Survey Year: When the variable title includes either the phrase XX INT (82 INT) or the year (e.g., 67) without the year being preceded by the preposition "IN," this indicates the survey year in which that variable was measured, not necessarily the year to which it applies.

> Example 2:　　'Move to Current Residence - Prior SMSA, 82 INT' refers to a residential move occurring in the period before the 1982 interview.
> Example 3:　　'Number of Weeks Worked in Past Year, 67' refers to the weeks worked in the 12-month period preceding the 1967 survey.

Calendar Year: When a date follows a verbal description of a variable and is part of the prepositional phrase "in XX," the date identifies the calendar year for which the relevant information was collected. The title in Example 4 refers to payments received in calendar year 1988 with data collected during the 1989 survey.

> Example 4:　　'Income from Social Security Payments Based on R's Work Record in 88? 89.'

> *User Notes:* All searches for NLS variables are essentially searches for variable descriptions or titles. Electronic searches of NLS variables via the NLS CD-ROM accessing methods ultimately produce listings of variables by their reference number and variable description or title.

Flexibility in variable title assignment for raw data items is restricted by (1) the actual wording of the question as it appears within the survey instrument; (2) precedent, i.e., how that type of variable has been titled in previous survey years; and (3) the maximum allowable length for variable titles. An attempt is also made to include key phrases in variable titles so that large groups of variables with similar or related subject matter can be easily identified.

Users should be careful not to assume that two variables with the same or similar titles necessarily have the same (1) universe of respondents **or** (2) coding categories **or** (3) time reference period. While the universe identifier and reference period conventions discussed above have been utilized, users are urged to consult the questionnaires for skip patterns and exact time periods for a given variable and to factor in the relevant fielding period(s).

Variables with similar content (e.g., information on respondents' labor force status) may have completely different titles, depending on the type of variable (raw versus created).

> Example 5:   'Employment Status Recode' (ESR) is the created or reconstructed version of the 'Activity Most of Survey Week' raw variable. The 'Activity' variable is derived from the first item of the full series of questions used by the Department of Labor (DOL) to obtain employment status; the title reflects questionnaire content. ESR, on the other hand, reflects the procedure used to recode the 'Activity' variable. This produces a constructed variable for all NLS respondents based upon responses to the 'Activity' question and all other questions used by the DOL to obtain employment status. These other questions serve to qualify and refine employment status beyond the answer to the initial 'Activity' question. (Note that ESR has been replaced by a similar variable, MLR, beginning in 1995; see the "Labor Force Status" section of this guide for details.)

Finally, different archivists over a period of three decades have performed the task of assigning variable descriptions to data from the NLS cohorts. While every effort has been made to maintain consistency, users may find some differences in variable titles. Two primary sources of variation exist in Original Cohort variable title assignment. The first is systematic error in which identical questions may have the same question wording across the four Original Cohorts but slightly different variable titles. The rule before 1995 was to make title consistency *within* a cohort of highest priority. Starting in 1995, joint fielding forced the archivists to choose one title and cross-reference the other cohort's title in the archivist notes. The second variation is attributed to random error due to spacing or punctuation errors. The sorting process that produces variable title listings usually places these variables near if not next to the series of interest.

> *User Notes:*   There are important differences between the content of telephone and personal interviews. In the late 1960s and early 1970s, most of the interviews were conducted in person, usually at the respondent's home. There was one attempt at a mail survey for the Older Men and Mature Women cohorts in 1968; however, the low response rate led to dropping that type of contact. After the first five years, the decision was made to conduct a major survey every five years and two telephone surveys during the five-year span so that problems of recall could be avoided and contact could be maintained with the respondents.

There are several different ways of identifying whether a survey is a personal or telephone interview. Users can (1) refer to Table 2.4.1 in the "Interview Schedule & Fielding Periods" section of Chapter 2, which depicts the type of interview by survey year, or (2) examine variable titles assigned to questions of similar content. Differences in what appear to be comparable variables reflect variations in the wording of the question or the fact that the reference period for an identically worded question may be different in a personal versus a telephone interview. Questions that refer to the last five years were usually found in a personal (or five-year) interview. This difference means that some questions were only asked in the five-year surveys and some were asked only in the telephone surveys. Users conducting longitudinal analysis need to change their variable creation procedures to account for the differences in data collection between the early years of uninterrupted personal interviews and subsequent survey years when telephone interviews were used.

Starting with the 1989 survey, the collection pattern was altered again; a decision was made to conduct a personal interview every other year and collect data going back to the date of the last interview. However, the scheduled 1991 survey was delayed until 1992 due to the demands of the 1990 decennial census and the decision to interview the Young Women first, in 1991. The scheduled 1994 survey was then delayed until 1995 so that the women's cohorts could be interviewed at the same time using the same CAPI instrument. Biennial surveys have been conducted in 1997 and 1999.

When analyzing data, users should remember that not all surveys were conducted during the same season of each survey year. Responses to labor force status questions, for example, may differ significantly if fielding occurred during the summer versus winter months. See the discussion of fielding periods in chapter 2 of this guide for more information.

3.3 Mature Women Codebook System

All variables present on a main file data set are documented via: (1) a cohort-specific codebook, (2) an accompanying codebook supplement, and (3) error updates. This section describes these three primary components of the codebook system and discusses the important types of information found within each.

Codebooks

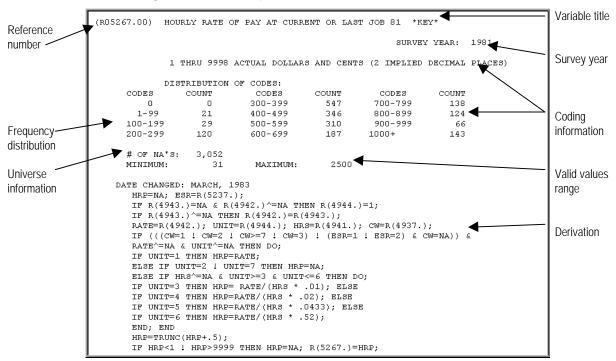The codebook is the principal element of the documentation system and contains information intended to be complete and self-explanatory for each variable in a data file. Codebook information can be viewed using the search software or printed from the CD-ROM. This feature enables researchers to customize their documentation for their particular research needs and to select and print information for the variables of interest.

Every variable is presented within the documentation as a block of information called a "codeblock." Codeblock entries depict the following information: a reference number, variable title, coding information, frequency distribution, reference to the questionnaire item or source of the variable, and information on the derivation for created variables. The codeblocks of many variables include special notes containing additional information designed to assist in the accurate use of data from that variable. Users will find that codeblocks from CAPI surveys present greater detail on each variable, including question text, universe totals, universe skip patterns, and more range of acceptable values information. Each of these terms is described more completely below.

Codebooks are arranged by reference number. Variables are first grouped according to survey year. Within each survey year, those variables related to the interview (e.g., interview method, interview date, reason for noninterview, sampling weight, etc.) appear first, followed by variables picked up directly from the questionnaire and *Information Sheet*. In general, created and edited variables appear last, although the created environmental variables are grouped with variables relating to the interview in the early survey years.

User Notes: NLS codebooks are not a substitute for the questionnaires. Although these two pieces of documentation contain similar information, the questionnaires should be used to determine precise universe information and question wording.

Codebook Item Descriptions

The following common types of information for each variable within a codeblock will be described in this section: coding information, multiple responses, missing responses, derivations, frequency distribution, questionnaire items (question numbers), universe information, valid values range, and verbatims. The sample codeblocks in Figures 3.3.1 and 3.3.2 provide visual examples of this information. The first codeblock is a created variable and the second is a question actually asked of respondents during the interview, allowing users to identify the differences between codeblocks for these two types of variables.

Figure 3.3.1 Sample Created Variable Codeblock


```
(R05267.00)   HOURLY RATE OF PAY AT CURRENT OR LAST JOB 81  *KEY*

                                          SURVEY YEAR:  1981

             1 THRU 9998 ACTUAL DOLLARS AND CENTS (2 IMPLIED DECIMAL PLACES)

            DISTRIBUTION OF CODES:
     CODES       COUNT        CODES     COUNT        CODES      COUNT
       0           0         300-399     547        700-799      138
      1-99         21        400-499     346        800-899      124
     100-199       29        500-599     310        900-999       66
     200-299      120        600-699     187        1000+        143

   # OF NA'S:   3,052
     MINIMUM:      31     MAXIMUM:      2500

   DATE CHANGED: MARCH, 1983
     HRP=NA; ESR=R(5237.);
     IF R(4943.)=NA & R(4942.)^=NA THEN R(4944.)=1;
     IF R(4943.)^=NA THEN R(4942.)=R(4943.);
     RATE=R(4942.); UNIT=R(4944.); HRS=R(4941.); CW=R(4937.);
     IF (((CW=1 | CW=2 | CW>=7 | CW=3) | (ESR=1 | ESR=2) & CW=NA)) &
     RATE^=NA & UNIT^=NA THEN DO;
     IF UNIT=1 THEN HRP=RATE;
     ELSE IF UNIT=2 | UNIT=7 THEN HRP=NA;
     ELSE IF HRS^=NA & UNIT>=3 & UNIT<=6 THEN DO;
     IF UNIT=3 THEN HRP= RATE/(HRS * .01); ELSE
     IF UNIT=4 THEN HRP=RATE/(HRS * .02); ELSE
     IF UNIT=5 THEN HRP=RATE/(HRS * .0433); ELSE
     IF UNIT=6 THEN HRP=RATE/(HRS * .52);
     END; END
     HRP=TRUNC(HRP+.5);
     IF HRP<1 | HRP>9999 THEN HRP=NA; R(5267.)=HRP;
```

Reference number · Variable title · Survey year · Coding information · Frequency distribution · Universe information · Valid values range · Derivation

Figure 3.3.2 Sample Interview Question Codeblock


```
R19282.       [RMP-81I-ARR-01] HOURLY RATE_OF_PAY EXCLUDING OVERTIME, 95 JOB #01
                                              Survey Year: 1995

(If B1B = 1 (Yes), fill parenthetical, "Excluding overtime pay...")

[Excluding overtime pay, tips, and commissions,] What [(is)(was)] your hourly
rate of pay on this job?

0-999999999 actual rate  with two implied decimal places
NOTE:  SEE SECTION ON
INCOME TOPCODING IN APPENDIX 3B

       0      0:      None  (Go To R19306.)
     269      1 To 999:       1 -    999  (Go To R19306.)
      84   1000 To 1999:    1000 - 1999  (Go To R19306.)
       9   2000 To 2999:    2000 - 2999  (Go To R19306.)
       1   3000 To 3999:    3000 - 3999  (Go To R19306.)
       0   4000 To 4999:    4000 - 4999  (Go To R19306.)
       0   5000 To 5999:    5000 - 5999  (Go To R19306.)
       0   6000 To 6999:    6000 - 6999  (Go To R19306.)
       0   7000 To 7999:    7000 - 7999  (Go To R19306.)
       0   8000 To 8999:    8000 - 8999  (Go To R19306.)
       0   9000 To 9999:    9000 - 9999  (Go To R19306.)
       0  10000 To 999999999: 10000 - and over  (Go To R19306.)
    -------
     363

Refusal(-1)         6 (Go To R20038.)
Don't Know(-2)      4
TOTAL =========>   373  Valid Skip(-4)  2338    Non-Interview(-5)  2372

Min:        100    Mean: 835.77
Max:       3500

Lead In: R19270.[Default]
Default Next Question: R19294.
```

Reference number · Question number · Verbatim · Variable title and survey year · Coding information · Frequency distribution · Coding conventions · Universe information · Valid values range

Coding Information

Each codeblock entry presents the set of legitimate codes that a variable may assume along with a text entry describing the codes. *Users should note that coding information for a given variable in the NLS codeblock is not necessarily consistent with the codes found within the questionnaire or for the same variable across years. Use only the codebook coding information for analysis.* The following types of code entries occur in NLS codeblocks:

Dichotomous or yes/no variables are uniformly coded "Yes" = 1, "No" = 0. Other dichotomous variables have frequently been reformulated to permit this convention to be followed.

Discrete (Categorical), as in the case of 'Activity Most of Survey Week 92.'
1	Working	5	Keeping house
2	With a job, not at work	6	Unable to work
3	Looking for work	7	Retired
4	Going to school	8	Other

Continuous (Quantitative), as in the case of 'Hourly Rate of Pay at Current or Last Job 81 *KEY*.' These variables have continuous data, but the codebook presents a frequency distribution as in the sample codeblocks above for ease of use.

Combined Quantitative-Qualitative Variables, i.e., variables which are ostensibly quantitative but which may have several nonquantitative responses, utilize positive integers equaling the actual values for the quantitative responses and extreme positive numbers for the qualitative (categorical) responses. For example, 'Expected Age of Retirement' is coded as follows:

> 45 thru 99 actual age
> 999 already retired
> 998 never plan to retire

Multiple Responses: Response categories to multiple entry questions found in certain job search, child care, discrimination, or health questions have been coded in a geometric progression. For example, more than one response to the question "Method of seeking employment to be used in next year" was possible. The response categories to that question were each assigned a value as follows:

Checked with public employment agency	1
Checked with private employment agency	2
Checked with employer directly	4
Checked with friends or relatives	8
Placed or answered ads	16
Other method	32

Multiple responses are then coded for each respondent by adding the individual codes, which yields a unique value for each combination. Such multiple entry variables are identified by an asterisk (*) next to the answer categories in the questionnaire. If a multiple entry has only a few unique combinations, the codebook will specify the exact combinations; those with many combinations need to be unpacked.

Methods of unpacking such multiple entry variables are presented in Appendix C at the end of this guide.

> *User Notes:* After the 1989 survey, the practice of coding multiple entry variables in a geometric progression was discontinued and all responses were coded as yes/no. In this system, the question above would have six corresponding variables in the codebook, one for each response category. Codes of 1 and 0 would indicate whether the respondent answered positively for each category. Respondents who do not know or refuse to answer the question receive the appropriate missing value for all the variables that correspond to that question. Respondents who do not know or refuse to respond to just one category receive the appropriate missing value for the corresponding variable.
>
> The system for coding missing values in multiple response questions changed slightly in 1999. There are still separate variables for each response category, and respondents who do not know or refuse to respond to just one category are coded with the correct missing value for the corresponding variable. The difference is that, at the end of the series of variables, a new variable indicates that it is the final record for the series. In this variable, respondents who answered any or all of the category questions receive either a –8 or a 0 code, depending on the series, to indicate that they are done selecting response categories. In this variable, respondents who replied "don't know" to the entire series are coded as –2 and those who refused to answer the entire series are coded as –1. For some series, this final variable may have other options in addition to those described above.

Missing Responses: Negative numbers are used to indicate that a respondent does not have a valid value for a particular variable. Different numbers indicate different reasons for nonresponse:

"Refusal" indicates that the respondent refused to answer a given question. These respondents are assigned a value of -1. This code is only used for CAPI interviews (1995–99).

"Don't know" indicates that the respondent did not know the answer to a given question. These respondents are assigned a value of -2. This code is used for all interviews of this cohort.

"Invalid skip" indicates that the respondent was not asked a question that she should have answered, usually due to programming or interviewer error. These respondents are assigned a value of -3. This code is only used for CAPI interviews (1995–99).

"Valid skip" indicates that the respondent was skipped past the question intentionally, because she was not in the universe of respondents to whom that question applied. These respondents are assigned a value of -4. This code is only used for CAPI interviews (1995–99).

Finally, a value of -5 has been assigned for slightly different reasons in different years. In PAPI surveys (1967–92), a -5 code indicates the absence of a valid response,

because (1) the respondent is not in the applicable universe, (2) the respondent refused to respond, (3) interviewer, coding, transcribing, or data entry error occurred, or (4) the respondent was not interviewed in that year survey. Beginning with the first CAPI survey in 1995, the -5 code is reserved for respondents not interviewed in a given year. Because computer interviewing permits more exact determination of the reason for nonresponse, the other reasons for the absence of valid data are described by the expanded missing value codes listed above.

User Notes: The missing value codes described above are accurate for the 1999 data release. In previous years, a more complicated system was used to indicate missing data in the PAPI interviews. Beginning in 1999, the missing values were reassigned using a standardized system that matches the Mature Women's CAPI data as well as the other NLS cohorts. This standardization should make it easier to use the data in analysis. However, researchers using programs written for a previous release of the Mature Women data may need to change the parts of their programming code related to missing values. Users who need more information about the codes previously used in order to make these adjustments should contact NLS User Services.

Three additional negative codes are used only with the NLS women's cohorts for particular types of nonresponse.

In questions dealing with usual hours per week worked, if the respondent reported that her hours varied, she was assigned a code of -6.

Women who have been widowed since the last survey are asked a series of questions regarding their husband's care and their financial situation since his death. A code of -7 was assigned to women whom the interviewer judged to be emotionally unable to answer these questions.

Some variables in multiple response question series include codes of -8, indicating that the respondent was done with the series. A more detailed description is provided under "Multiple Responses" above.

User Notes: In CAPI surveys, respondents are initially assigned a default code of -4 (valid skip) for all questions in the interview. As the interview proceeds, the -4 codes are replaced by valid data. The -3 (invalid skip) codes must be inserted into the data as hand-edits when data archivists uncover skip pattern errors during the data cleaning process. Therefore, some respondents classified as valid skips may actually have skipped a question incorrectly. If researchers need to know the exact reason a question was not answered, they can examine the skip patterns and universes in the questionnaire to determine whether any additional respondents should have been identified as invalid skips.

Derivations: The decision rules employed in the creation of constructed variables have been included, whenever possible, in the codebook under the title "DERIVATIONS." This information is designed to enable researchers to determine whether available constructs are appropriate for their needs. In the 'Hourly Rate of Pay at Current or Last Job 81 *KEY*' example (Figure 3.3.1), the derivation describes in detail the items of the interview schedule used to create the variable. If the derivation is too lengthy to include in the codebook, the codeblock will instead refer users to the supplemental documentation item that contains variable creation information.

Frequency Distribution: In the case of discrete (categorical) variables, frequency counts are normally shown in the first column to the left of the code categories. In the case of continuous (quantitative) variables, a distribution of the variable is presented using a convenient class interval. The format of these distributions varies. In the case of the illustrative variables in Figures 3.3.1 and 3.3.2, the frequency count is straightforward. For example, in Figure 3.3.1 there are twelve categories; the maximum category shown is 1000 and above (since two decimal places are implied, 1000 represents $10.00), for which there is a frequency count of 143.

Questionnaire Item: "Questionnaire item" is a generic term identifying the source of data for a given variable. A questionnaire item may be a question, a check item, or an interviewer's reference item appearing within one of the survey instruments. Questionnaire item identifications are located in the extreme right hand column of the codebook. The question number, when available, is copied exactly from the questionnaire. The question numbering system is described in the questionnaire section earlier in this chapter. In Figure 3.3.2, the question number is RSP-81I-ARR-01.

During PAPI interview years, the absence of a question entry in the codeblock (as in Figure 3.3.1) indicates that a variable was not taken directly from the questionnaire and is therefore a created variable. Created variables in CAPI survey years usually include the letters CV in the question name and usually have the word *KEY* in their title.

Valid Values Range: Depicted below the frequency distribution are the maximum and minimum fields, which define the range of valid values (the upper and lower limits) for a given question. "MINIMUM" indicates the smallest recorded value exclusive of nonresponse codes; "MAXIMUM" indicates the largest recorded value. In the case of the 'Hourly Rate of Pay at Current or Last Job' example (Figure 3.3.1), the maximum, or highest value recorded, is 2500 with two implied decimal places, or $25.00.

Topcoding Income and Asset Values: Confidentiality issues restrict release of all income and asset values. To ensure respondent confidentiality, income variables exceeding particular limits are truncated each survey year so that values exceeding the upper limits are converted to a set maximum value.

These upper limits vary by year, as do the set maximum values. From 1967 through 1972, upper limit dollar amounts were set to 999999. From 1974 through 1980, upper limit variables were set to maximum values of 50000, and from 1981 to 1984 the set maximum value was 50001. Beginning in 1986, income amounts exceeding $100,000 were converted to a set maximum value of 100001.

From the cohort's inception, asset variables exceeding upper limits were truncated to 999999. Beginning in 1977, assets exceeding one million were converted to a set maximum value of 999997. Starting in 1992, the Census Bureau also topcoded selected asset items if it considered that the release of the absolute value might aid in the identification of a respondent. This topcoding was conducted on a case-by-case basis with the mean of the top three values substituted for each respondent who reported such amounts.

Codebook Supplements

Variable creation procedures and supplemental coding information are provided within the *Codebook Supplement*. Information provided in these documents is not available in the electronic documentation files on the NLS CD-ROMs. The following attachments and appendices are included in the *Codebook Supplement*, which is available in hard copy form.

> *Attachment 2: 1960, 1980 & 1990 Census of Population Industry and Occupational Classification Codes* provides the occupation-industry coding assignments made by Census Bureau personnel from the verbal descriptions obtained in the interviews. Users should refer to the "Industries" or "Occupations & Occupational Prestige Indices" sections in chapter 4 of this guide for information about which coding schemes were used in various survey years. This attachment also contains a copy of the Duncan Socioeconomic Index, an ordinal prestige scale assigning a rank of 0–97 to each of the three digit 1960 Census occupations.

> *Attachment 4: Bose Index* provides a mean occupational prestige score for each of the three-digit 1960 occupation codes for respondents of the cohort.

> *Attachment 5: Employment Status Recodes* describes the methodology used by Census to calculate each respondent's employment status from the CPS questions that are asked in each NLS survey. This document provides (1) definitions of 'working,' 'with a job but not at work,' 'unemployed,' and 'not in the labor force'; (2) the decision rules used to assign or recode respondents to a particular labor force status; and (3) Census methodology for dealing with exceptions to the rules.

> *Appendix 3:   State Names and State Codes by Census Division Listing*
> *Appendix 4:   Derivations for R02847. (Reason Left Current Job 1967)*
> *Appendix 5:   Source for Occupational Atypicality Codes*
> *Appendix 6:   Derivations for R02872.50 (Occupational Training 1967–1972)*
> *Appendix 7:   Derivations for R00792.50 (Training prior to 1972)*

Error Updates

Prior to working with an NLS data file, users should make every effort to acquire current information on data and/or documentation errors. A variety of methods are used to notify users of errors in the data files and/or documentation and to provide those persons who acquired an NLS data set from CHRR with corrected information. Errors discovered after the release of a data file are distributed in hard copy form to current disc purchasers along with the data set. Error notices also appear, along with information on how to acquire the corrected data and/or documentation, in *NLS News*, the quarterly newsletter.

3.4 CD-ROM Search Functions

Variables can be accessed through search and extraction software available on the NLS CD-ROMs. This software provides users with bridging information to the codebook and/or survey instruments. The search indexes and lists described below can be used individually or combined to produce a more refined list of variables. This section provides only a cursory overview of the search and extraction software on

the NLS CD-ROMs. Researchers who need more information should refer to the quick reference guide in Appendix A of this document and to the *CD-ROM User's Guide* distributed with the data.

Any word search.   The "Any Word in Context" function (also called "contextual search") on the CD-ROM software allows the user to search for and select those variables whose titles contain any single word or combination of words found in the entire documentation database. This function allows users to easily access variables on a variety of topics but is still dependent on the wording of each variable title. For more information on the naming of variables, see the "Variable Titles" discussion in section 3.2 above, especially the User Notes.

Area of interest.   Areas of interest contain lists of variables grouped by topic. For example, questions on a respondent's health and medical insurance are grouped in the "Health" area of interest. Researchers should be aware that an individual question can be linked to only one area of interest, so questions that apply to a common research topic may appear in different areas of interest.

Question number list.   The CD-ROM contains a searchable list of the question numbers for every Mature Women variable. By accessing this list, users who know the question number for the item of interest can locate the variable without performing an any word search. Researchers can also browse through the questions for a given questionnaire section.

Reference number list.   The disc also contains a searchable list of all Mature Women reference numbers. By using this list, researchers can locate variables for which they already know the reference number or browse through questions of interest, which are generally arranged in the order in which they were asked during the interview.

Year index.   Finally, the CD-ROM includes an option by which users can limit their searches to a single year of interest. Researchers can browse through all the variables relating to a given survey year or can combine the year with an any word search to locate specific variables of interest.