

## INTRODUCTION

---

This document describes the question content of the 2003 interview of the National Longitudinal Survey of Women (NLSW). The survey was conducted entirely as a computer-assisted interview, and although a few respondents were interviewed over the telephone, the dominant interview mode was the in-person interview. Therefore, the data collection method for the 2003 NLSW is referred to simply as CAPI, or computer-assisted personal interview. This is different from the way the survey was conducted prior to 1995. From 1967 to 1993 these cohorts were surveyed using paper and pencil (PAPI) methods. One important difference between the PAPI and CAPI survey data collection methods is that the paper-and-pencil interviews made it easy for researchers to determine the content of the survey by examining the printed questionnaire and other interviewing materials. In a CAPI survey, however, there is no real printed questionnaire since the survey exists as a group of programs stored on every interviewer's laptop computer. As a result, the questionnaire gets generated from the CAPI survey and is longer, more complex, and somewhat more difficult to use than the PAPI questionnaire.

Our goal is to present the content of the 2003 CAPI interview into a clear, complete, and compact questionnaire that accurately summarizes the “conversation” between the interviewer and respondent. We use the term “conversation” to emphasize the difference between the questionnaire and the codebook. The codebook contains the data released as a result of the interview, but some information collected during the interview is not present in the codebook. For example, the questionnaire includes questions that collect names of household members, employers, etc. Such items are not released, so the questions that collect them are not present in the codebook. Conversely, the codebook contains some information not directly collected in the interview. For example, while created variables are documented in the codebook, there is usually no reference to them in the questionnaire as they are not part of the conversation between interviewer and respondent.

The next section of the introduction describes the format of the typical question blocks and defines key terms such as “Question name,” “Distribution Code Block,” and “Lead in Questions”. The final section describes in more detail the differences in the content of the questionnaire and the codebook.

## TYPES AND FORMATS OF QUESTION BLOCKS

Question records are the building blocks of the CAPI questionnaire, with various types of records performing different data collection and/or manipulation functions. Loosely grouped, CAPI question records will either: (a) provide administrative information that assists the interviewer or provides an introduction to the next question; (b) provide question text and control for data entry of the response; or (c) perform internal operations and calculations to guide the interview and maintain the internal data storage during the interview. All questions essential to understanding either the content of the survey or the flow of respondents through the interview will appear within the 2003 NLSW questionnaire. However, some of the elements not essential to understanding the conversational flow of the interview have been removed, making each question block less cluttered and easier to read.

The 2003 NLSW questionnaire contains some elements found in the 2003 NLSW codebook, such as the question names, skip instructions, and lead-in and default next question information. Figure 1 shows all of the fields that appear in a question block in the questionnaire. A brief description of each of the numbered fields is provided below.

**FIGURE 1: Example**

---

∂ [R47484.00]     ●HWH-42

÷ (IF SIGOTHER = 1 OR 2 FILL PARENTHETICAL ( ) WITH HUSBAND’S NAME)  
(IF SIGOTHER = 3 OR 4 FILL PARENTHETICAL ( ) WITH PARTNER’S NAME)

≠ INTERVIEWER: MARK 3, 4, OR 5 ONLY IF R VOLUNTEERS INFORMATION.

≡ Does (H-FIRST) currently want a job, either full or part time?

≈

	(a)	(b)	(c)
1	YES OR MAYBE, IT DEPENDS...(Go to CK-HWH-M5)		
0	NO		
3	RETIRED...(Go to CK-HWH-M4)		
4	DISABLED..(Go to CK-HWH-M3)		
5	UNABLE		

... If Answer >=-2 and Answer <=-1 Then Go To: CK-HWH-M6

↵ Design Note: See Appendix 37 for more detail

| DefNext: CK-HWH-O

— Lead-In: CK-HWH-M2 [4:4]

---

### 1. Reference Number

The reference number is the variable name assigned to each question in the codebook and CHRR Investigator and extraction software. Every data item in the survey has a unique reference number, so that no two items have the same reference number. Also, reference numbers are assigned only to questions for which data exists. In other words, there are no reference numbers for confidential items, or items that no respondents reached. Thus, questions for which no data exists are still included in the questionnaire, but instead of showing a reference number in brackets, the question block is blank “[ ]”. Some questions have more than one reference number listed. This occurs when a single question, such as a date, contains two or more data fields. For example, question OJS-5A-ARR is represented as a single question in the questionnaire, but actually consists of 3 data items in the codebook and data, once each for Month (R43546.00), Day (R43547.00) and Year (R43548.00). Finally, not all reference numbers pertaining to every question appear in the questionnaire. This is because many questions are part of a roster, such that the question is asked once about every employer (as in OJS-5A-ARR), or every person in the household (as in HRC-6-ARR). For example, the start date collected in question OJS-5A-ARR actually generated 30 separate data items in the data and codebook: Day, Month, and Year for up to 10 employers (3 date entries x 10 employers = 30). To conserve space and simplify the questionnaire we list only the reference numbers pertaining to the items in the first roster (i.e., OJS-5A-ARR-01D, OJS-5A-ARR-01M, and OJS-5A-ARR-01Y).

### 2. Question Name

Question names are the identifiers assigned to identical questions across CAPI survey years. Question names provide information on: (a) the section of the instrument from which each question was derived—e.g., question names beginning with HEA- come from the “Health” section of the questionnaire, while

those starting with OJS- come from the respondent's "on jobs" section.; (b) the order in which the question was administered within each section.; e.g., HEA-5 precedes HEA-6; and (c) whether a question was part of a "looped" or rostered series of questions. A question ending with "-ARR" could have been asked more than once, and will appear as many times in the codebook as there are data items. For example, questions in the RSP section were asked once for each employer name the respondent had listed in her employer roster. However, since some respondents have more than one employer, and the list of employers does not always begin with the first "column" in the roster (the "-01" column), many questions in the RSP section generated numerous data items. For example, question RSP-1-ARR ("Were you employed by government, by a private company, a non-profit organization...") generated 8 data items, listed in the codebook as RSP-1-ARR-01 through RSP-1-ARR-08 (with corresponding reference numbers of R44401, R44624, R44824, R45024, R45216, R45405, R45584, R45756).

### **3. Text fill instructions**

Text appearing within parentheses does not generally appear on screen during the interview. This is true for the text-fill instructions, the skip instructions (i.e., "Go to..."), and the "-2" and "-1" that follow "Don't Know" and "Refuse", respectively. In the current example, field ÷ shows the instructions the CAPI program used to automatically fill in the appropriate name in the text of the question. These instructions do not appear on-screen during the interview so they are not read to the respondent, nor does the interviewer see the words "[fill name]".

### **4. Interviewer instructions**

Text in uppercase is either an instruction or clarifying comment for the interviewer. These comments appear on screen during interview but the interviewer does not read them to the respondent.

### **5. Question Text**

This field contains the text of the question that the interviewer asked the respondent. Generally, text that is asked of the respondent appears in the questionnaire in a conventional mixture of upper and lowercase letters.

### **6. Distribution code block**

The distribution code block contains the list of response categories to each question. Each distribution code has a value, a description of the category, and an optional skip instruction.

- (a) Column (a) contains the numeric or alphabetic value associated with each response category. All data in the 2003 NLSW data file are numeric, so each possible item that could be selected is associated with an integer.
- (b) This column contains text that describes each response category of the distribution code block. This text is sometimes referred to as "value labels" in statistical programs such as SPSS.
- (c) Column (c) contains the skip instructions for the response category. Skip instructions are needed only when the response is branched to somewhere other than the default next question.

### **7. 'Don't Know' and 'Refuse'**

Responses of 'Don't Know' and 'Refuse' are acceptable responses to nearly every question, but these answers do not appear in every question in the questionnaire. Specifically, "Don't Know" and "Refuse" do not appear in the distribution codeblock when these responses are skipped to the default next question (DefNext). Conversely, when the 'Don't Know' or 'Refuse' responses are skipped to a question other than the DefNext question they are listed in the codeblock, as shown in this example.

## 8. Default Next Question

This field specifies the question to be asked next unless another skip specification in the distribution code block applies. In Figure 1, the responses categories “Unable” is skipped by default to CK-HWH-O, as indicated in the Default Next Question instructions.

## 9. Lead in Questions

When using PAPI questionnaires, researchers needed to study the instrument in detail to determine under what circumstances a particular question was asked. The CAPI questionnaire simplifies this process somewhat by including a “Lead-In(s)” field in the question codeblock. This field lists the question or questions that lead into each particular item. By following the skip pattern backward, one may determine the universe of respondents asked a particular question.

Examples:

RSP-1-ARR [Default]	This means that the default path from question RSP-1-ARR leads to the current question.
CK-HWH-M2 [4:4]	This means that all cases with a value of 4 in CK-HWH-M2 are skipped to the current question.
CK-HWH-M2 [1:4]	This indicates that values of 1 through 4 (inclusive) in CK-HWH-M2 are skipped to the current question record.

## 10. CHRR archivist comments

This field contains comments added by the CHRR data archivist. These comments are not part of the original question text or the interviewer instructions. They are included in the questionnaire to provide users with additional information about the question, the interview protocol, and the data. The archivist comments usually appear at the bottom of the question, i.e., after the question text and the response options.

## DIFFERENCES BETWEEN THE CODEBOOK AND QUESTIONNAIRE

One of the differences between the content of the codebook and questionnaire is that the codebook contains questions that are not included in the questionnaire. In addition, there are some questions in the questionnaire that are not included in codebook. Another difference is that some types of questions are split into separate records in the codebook and stored as separate data items in the data file but are combined into a single question in the questionnaire (e.g., dates in the D/M/Y format and Code-All-That-Apply questions). Finally, for some items, the skip patterns shown in the questionnaire are different from those shown in the codebook.

### A. Codebook Items not in the Questionnaire

Rostered, or “looped”, questions appear only once in the questionnaire. For example, the CAPI instrument asks questions about the education, employment, etc., of everyone in the respondent’s household. To take an example, the question “Did [fill name of HH member] do any work for pay last week?” (question HRC-11-ARR) was asked once for each person over the age of 14 listed on the household roster. This resulted in the creation of 11 data items for this question in the Young Women data and codebook: HRC-11-ARR-02 through HRC-11-ARR-11, and HRC-11-ARR-13 (the 01<sup>st</sup>, 12<sup>th</sup> and 14<sup>th</sup> through 18<sup>th</sup> iterations were blank and so are not included in the data.) Users can readily identify rostered questions because the question names end with “ARR”.

Also, the 2003 NLSW questionnaire does not include certain “key” variables created by CHRR after the interviews were conducted. For example, neither net total family assets (reference number R51311.00) nor the number of weeks worked since the last interview (R54382.00) are included in the questionnaire.

## **B. Questionnaire Items not in the Codebook**

Questions that contain confidential information such as names of household members, addresses, and names of employers are not included in the codebook nor are they included in the data but they are retained in the questionnaire. For example, in question HRC-5A of the household record section the names of additional members of the household are collected. This question appears in the questionnaire but has no corresponding data element and is not included in the codebook.

Another group of questions that appear in the questionnaire but not the codebook are questions for which no data was collected. For example, questions OJS-4A-ARR through OJS-4E-ARR appear in the questionnaire but are not found in the codebook. These particular questions lie in a path in the survey that no respondents entered and were therefore left blank. In order to conserve space these empty items are omitted from the codebook and data file. (Note that reference numbers are not assigned to these items).

## **C. Combining Questions in the Questionnaire**

There are two types of questions that appear as separate items in the codebook but are combined into one question in the questionnaire. The first are questions collecting information about dates, such as when the respondent started or stopped working for an employer (e.g., OJS-5A-ARR). In the codebook these types of date questions are split into three variables: one for day, one for month, and one for the year. In the questionnaire these three data items are combined into one question, which is the form these questions took during the interview. For example, the question “When did you last stop working for [EMPLOYER NAME]?” is shown as one question in the questionnaire (OJS-5A-ARR), while it is divided into 3 separate questions in the codebook (OJS-5A-ARR-01D, OJS-5A-ARR-01M, and OJS-5A-ARR-01Y).

The second type of question in which separate codebook items are combined into a single questionnaire item is the “Mark-all-that-apply” question. This happens when the respondent can mark as many categories that apply from a list of response options. For example, questionnaire item MOB-12 asked the respondent to select the reasons she moved from her previous address to the current one. While this is presented as one item in the questionnaire there are 15 data items in the codebook and data file, one question for each of the reasons included in the list, MOB-01 through MOB-12@KEY. (Note in the questionnaire that the reference numbers for all 16 data items are listed: R51486.00 through R51501.00).

## **D. Skip Patterns**

The skip patterns shown in the questionnaire do not always match the skips shown in the codebook because items for which no data exists are not printed in the codebook. However, since all questions appear in the questionnaire the skip patterns have not been adjusted. For example, since none of the Young Women respondents reached questions OJS-4B and OJS-4E, these items are not included in the Young Women’s data nor in the codebook. The deletion of these questions from the codebook affected the branching instructions in the lead-in question, CK-INC-F. According to the questionnaire, when CK-OJS-D equals 1 the instrument should skip to OJS-4C, and when it equals 0 the skip is to OJS-4A. However, in the codebook the skip instructions for CK-OJS-D were adjusted to reflect the deletion of both OJS-4A and OJS-4B. Thus, in the codebook the skip instructions indicate that the default next question for CK-OJS-D is CK-OJS-F.