# 2001 NATIONAL HOUSEHOLD SURVEY ON DRUG ABUSE

## GENERAL PRINCIPLES AND PROCEDURES FOR EDITING DRUG USE DATA IN THE 2001 NHSDA COMPUTER-ASSISTED INTERVIEW

Prepared for the 2001 Methodological Resource Book

RTI Project No. 7190
Contract No. 283-98-9008

Deliverable No. 28

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, MD 20857

Prepared by:

RTI International
Research Triangle Park, NC 27709

June 2003

# 2001 NATIONAL HOUSEHOLD SURVEY ON DRUG ABUSE

## GENERAL PRINCIPLES AND PROCEDURES FOR EDITING DRUG USE DATA IN THE 2001 NHSDA COMPUTER-ASSISTED INTERVIEW

Prepared for the 2001 Methodological Resource Book

RTI Project No. 7190
Contract No. 283-98-9008

Deliverable No. 28

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, MD 20857

Prepared by:

RTI International
Research Triangle Park, NC 27709

June 2003

## Table of Contents

**Section** **Page**

## List of Exhibits

# 1.0 Background

The 1999 National Household Survey on Drug Abuse (NHSDA) marked the transition from data collection based on paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI). In the 2001 NHSDA, data were collected using CAI. As in PAPI, the CAI instrument allowed a private mode of data collection for respondents to answer questions pertaining to drug use and other sensitive topics. In CAI, this self-administration was accomplished through use of audio computer-assisted self-interviewing (ACASI) in which respondents could read the questions on the computer screen and enter their responses directly into the laptop computer. All respondents also were encouraged to listen to an audio recording of the questions on headphones and then enter their answers into the computer. This prevented interviewers (or others in the household) from knowing what questions the respondents were being asked and how they were answering. This feature of ACASI was especially useful for respondents with limited reading ability because they could listen to the questions instead of having to read them. For demographic questions, computer-assisted personal interviewing (CAPI) was used; interviewers read the questions and respondents gave their answers aloud to the interviewers, who then entered the responses into the computer.

The CAI instrument was divided into core and noncore sections. Core sections, such as key demographic characteristics and drug use prevalence questions, were designed to stay relatively constant from 1 year to the next in order to permit measurement of trends in drug use. In contrast, the content of noncore sections could change considerably across years to measure new topics of interest or to rotate certain topics in or out of the interview. In noncore sections, therefore, questions or entire modules could be added or deleted, or the wording of existing questions could change from 1 year to the next.

This report is designed to provide documentation on how the basic drug use prevalence data were edited from the 2001 CAI instrument. The overall purpose of any editing of the 2001 NHSDA CAI data was to provide the most accurate information possible about drug use and related issues among U.S. residents aged 12 or older.

## 1.1 Improvements in Data Quality with CAI

Conversion of the NHSDA interview to a CAI format greatly reduced (or in some cases, eliminated) the following data quality problems that can occur in the PAPI data:

- illegible responses, multiple marks, or out-of-range values;
- item nonresponse (i.e., missing data);
- incorrectly executed skip patterns; and
- inconsistencies among related variables.

In particular, multiple marks do not occur in the CAI because the computer program will permit entry of only one response at a time per item. Similarly, the instrument has been programmed not to allow out-of-range values for certain items, such as frequency-of-use items, thereby reducing the amount and types of out-of-range values that would otherwise need to be addressed through machine editing.

The skip logic in the CAI instrument also was designed to reduce the amount of missing data by skipping respondents past questions that did not apply to them, instead of requiring respondents to repeatedly mark a "does not apply" response category. (In PAPI, users and nonusers had to answer all questions on several drug use answer sheets, even though the questions did not apply to nonusers.) Because the computer program controlled the skip logic, respondents or field interviewers (FIs) also did not have the opportunity to enter a branch when they answered a lead item in a manner that would suggest that they should have skipped. Further, the skip logic was designed to reduce the occurrence of inconsistent data by not giving respondents the *opportunity* to provide inconsistent answers. The occurrence of inconsistent data was further reduced through the use of consistency checks built into the CAI program that prompted respondents to resolve inconsistencies that occurred between related items.

## 1.2 Data Issues Needing To Be Addressed with CAI

Despite the potential for improvements in data quality through a CAI instrument, we recognized that conversion to CAI would not completely eliminate data problems. We also recognized that in some situations, conversion to CAI could introduce new data quality issues. For example, missing data were not completely eliminated because CAI respondents still had the option of entering a response of "don't know" or "refused" when answering a given item. In addition, items that were unanswered because of a breakoff still had a missing value.

Similarly, even though skip patterns and consistency checks were designed to reduce inconsistent reporting, opportunities could still exist for inconsistent data to remain. For example, if a respondent did not resolve an inconsistent report when given the opportunity to do so, the interview simply proceeded with the inconsistency in place. In addition, the CAI program was not equipped to address every possible inconsistent report that a respondent could make—especially in noncore sections of the interview.

Section 2 of this report discusses general principles that were applied in editing the 2001 CAI data. Section 3 describes the initial steps in the editing and coding process, and Section 4 discusses edits that were implemented for the self-administered core drug use questions. Edit procedures for noncore ACASI modules and for the interviewer-administered CAPI sections are described in one or more companion documents.

# 2.0 General Principles of Editing

This section discusses the following issues and general principles of data processing and editing that were applied throughout the 2001 CAI data:

- assignment of standard NHSDA missing data codes, and
- assignment of relevant "not applicable" codes.

## 2.1 Assignment of Standard NHSDA Missing Data Codes

In 2001, the procedures for assigning standard NHSDA missing data codes were conducted as part of the regular processing of transmitted data (see Section 3.2). The following standard codes for missing data that were used in prior NHSDAs also were relevant to the 2001 CAI data:

94 (or 994 or 9994, etc.) = DON'T KNOW (DK),

97 (or 997 or 9997, etc.) = REFUSED (REF), and

98 (or 998 or 9998, etc.) = BLANK (i.e., nonresponse [NR]).

However, the program that was used to conduct the interview employed codes of 8 (or 98 or 998, etc.) to denote responses of "refused" and codes of 9 (or 99 or 999, etc.) to denote responses of "don't know." NHSDA project leadership within the Office of Applied Studies (OAS) in the Substance Abuse and Mental Health Services Administration (SAMHSA) approved the recommendation to replace these program codes with the standard NHDSA codes for "don't know" (DK) and "refused" (REF) that had been used in the past.

Notable advantages of assigning these standard NHSDA missing data codes to the CAI variables were that (1) use of values to signify missing data that have been used in prior NHSDAs would provide some continuity for users of the NHSDA data, despite the change to a new instrument; and (2) use of common missing data values between the CAI and PAPI data could facilitate analyses comparing data between the two instrument versions, such as amounts of and reasons for missing data. In particular, use of the standard NHSDA codes would reduce the amount of readjustment required by longtime users of the NHSDA data who might be accustomed to the missing data codes from prior rounds of the NHSDA. Moreover, some of the missing data codes supplied by the program matched codes that had very different meanings in prior NHSDAs. Specifically, codes of 98 (or 998, etc.) that meant "refused" in the program meant "blank, no answer" in prior NHSDA data. Similarly, codes of 99 (or 999, etc.) that meant "don't know" in the program meant "legitimate skip" in prior NHSDA data.

In addition to the edits for assigning standard NHSDA missing data codes, data were sometimes identified that were inconsistent with other data in a respondent's record. For example, if a respondent reported first using a drug at an age greater than his or her current age, the CAI program indicated to the respondent that this age at first use was inconsistent. The respondent was prompted to revise the age at first use, his or her current age, or both, to make the data consistent. As noted above, however, respondents did not always resolve these types of inconsistencies. If the age at first use was still inconsistent with the respondent's age despite the opportunity that the respondent had to resolve the inconsistency, we assigned a "bad data" code to the age-at-first-use variable to indicate that the data were inconsistent with other data. As was the case with prior NHSDAs, we assigned the following codes to denote "bad data" (i.e., usually inconsistent with other data):

85 (or 985, or 9985, etc.) = BAD DATA Logically assigned.

Other situations where bad data codes were assigned are discussed below in connection with specific steps in the machine-editing process.

In addition, blanks in the data file created from the cleaned transmitted interview data were denoted by codes of "." ("dot"). These were converted to codes of 98 (or 998, etc.).

## 2.2 Assignment of Relevant "Not Applicable" Codes

As noted above, the CAI logic skipped respondents out of questions that did not apply to them. Therefore, the following codes were assigned when respondents were skipped out of a given question *and it could be determined unambiguously* that the question did not apply, based on the answer to a previous question:

91 (or 991, or 9991, etc.) = NEVER USED [DRUG(s) OF INTEREST],

93 (or 993, or 9993, etc.) = USED [DRUG] BUT NOT IN THE PERIOD OF INTEREST, and

99 (or 999, or 9999, etc.) = LEGITIMATE SKIP.

Strictly speaking, codes of 91 and 93 in the CAI data could be considered variants of the more generic legitimate skip code. Their use was designed to provide analysts with more information about the reason that respondents were skipped out of a question.

Codes of 91 and 93 were most often used in the core drug sections of the interview (i.e., tobacco, alcohol, marijuana, cocaine, crack cocaine, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives). For example, codes of 91 (or 991, etc.) in the marijuana section denote the pattern where respondents were skipped out of all remaining marijuana questions because they answered "no" to the lifetime marijuana question MJ01 (i.e., "Have you ever, even once, used marijuana or

hashish?"). Similarly, codes of 93 (or 993, etc.) in the marijuana section denote situations where respondents were lifetime marijuana users but were definitely not users in the past 30 days or past 12 months (or both).

These codes of 91 and 93 also were used to a limited extent in noncore sections of the interview (i.e., following the sedatives section) because the CAI logic took into account respondents' prior answers to core drug use questions to determine whether particular noncore questions applied. For example, the substance dependence and abuse module was relevant to respondents who used cigarettes or specialty cigarettes in the 30 days prior to the interview or who used other drugs in the 12 months prior to the interview. Thus, for example, if a respondent last used cocaine more than 12 months prior to the interview, codes of 93 in the substance dependence and abuse variables pertaining to cocaine would signify to an analyst why the CAI program skipped the respondent out of these questions. Similarly, the substance treatment section was relevant only to respondents who used alcohol or some other drug (other than cigarettes) at some point in their lives. Consequently, codes of 91 in the substance treatment variables would signify to an analyst that the respondent was skipped out of the substance treatment section because he/she had never used alcohol or drugs.

Legitimate skip codes of 99 were used most often in the noncore sections of the interview. For example, the youth experiences module was intended to be administered only to respondents aged 12 to 17. Consequently, if a respondent was 18 or older, codes of 99 were assigned in the machine-editing process to the skipped youth experiences variables. Similarly, if a respondent had used alcohol or some other drug at least once in his or her lifetime but answered the lifetime substance treatment question TX01 as "no," the CAI program skipped the respondent out of all remaining substance treatment questions. Codes of 99 were assigned to the skipped substance treatment variables in this situation to signify that the respondent had used alcohol or drugs at least once but had never received substance abuse treatment.

The following analogous codes also were assigned through machine editing:

> 81 (or 981, or 9981, etc.) = NEVER USED [DRUG(s)] Logically assigned,
>
> 83 (or 983, or 9983, etc.) = USED [DRUG(s)] BUT NOT IN THE PERIOD OF INTEREST Logically assigned, and
>
> 89 (or 989, or 9989, etc.) = LEGITIMATE SKIP Logically assigned.

These codes were given values in the 80s to signify that existing values were overwritten during machine editing. For example, if a respondent was somehow routed into the youth experiences module but that respondent was subsequently classified as being 18 or older, any answers that the respondent gave in the youth experiences module were overwritten with codes of 89 (or 989, etc.). These codes signify that this adult respondent logically was not eligible to be asked the youth experiences questions. Other situations where these logically assigned codes apply are discussed below in connection with specific editing steps.

The preceding discussion in this section applies *only* to situations where there was total certainty that a respondent should have skipped a question. For example, if a respondent reported in question MJ01 that he or she has never used marijuana, it was absolutely clear that subsequent questions about marijuana use (e.g., age at first use of marijuana) did not apply.

However, the CAI skip logic often treated responses of "don't know" or "refused" to lead questions as equivalent to a negative response (important exceptions for refusals noted below). For example, if a respondent in 2001 was uncertain whether he or she had ever used marijuana (and answered question MJ01 as "don't know"), the CAI program skipped the respondent out of all remaining questions about marijuana use, as though the respondent had never used it. From the standpoint of respondent burden, there is little value in routing these respondents into questions that imply an affirmative response to a preceding question. If, for example, a respondent did not know whether he or she had ever used marijuana, it would make no sense to ask, "How old were you the first time you used marijuana or hashish?" This also would imply a belief that the respondent really has used but is reluctant to admit it.

The 2001 interview included follow-up probes for respondents who initially refused to report whether they had ever used a drug of interest. Follow-up probes were not included for respondents who reported that they did not know whether they had ever used a drug. As noted above, respondents who gave an answer of "don't know" were treated the same as respondents who had given an answer of "no."

Respondents who initially refused to report whether they had ever used a drug (or refused all questions about specific drugs within a category) were routed next to a follow-up question that attempted to persuade them to reconsider answering the question they had refused. Similarly, in the inhalants, pain relievers, tranquilizers, stimulants, and sedatives modules in 2001, respondents who refused to answer all questions about their use of specific substances in that category were asked whether they had ever used any type of drug in that category (e.g., any inhalant). If respondents changed their initial refusal (or initial series of refusals) to a response of "yes," they were routed into subsequent detailed questions about use of that drug, just as if they had answered affirmatively at the outset. Similarly, if respondents changed their initial refusal to a response of "no," it could be unambiguously inferred that the respondent was a nonuser, and subsequent questions about use of that drug did not apply. If respondents changed their initial refusal to a response of "don't know," or they again refused on follow-up, the CAI program routed them in the same manner as if they had given a negative response.

Although the CAI program skipped respondents out of questions if they answered a lead question as "don't know" or "refused" (or gave similar answers on follow-up), these types of responses to a lead question that governs a skip pattern are ambiguous; they do not provide an analyst with conclusive information one way or the other about the behavior or event of interest. Consequently, such responses could be thought of as *potentially* affirmative responses, as opposed to inferring that they are negative responses. In particular, as noted above, respondents who initially refused to answer a question about

their lifetime use or nonuse of a drug were given a second opportunity to answer the question as "yes" or "no." Similarly, if a respondent who initially did not know whether he or she had ever used marijuana had thought about the issue further, the respondent may have recalled a time when he or she in fact had used it—and more detailed questions about marijuana use would have been relevant for this respondent. Alternatively, if the respondent gave more thought to the issue and decided that he or she really should answer the lifetime marijuana use question as "no," an analyst would have a solid basis for determining that subsequent questions did not apply.

Further, the procedures for statistically imputing missing data did not automatically infer lifetime nonuse when respondents provided ambiguous information about whether they had ever used a given drug (see Section 4.1). Rather, such respondents were eligible to be statistically imputed to be lifetime users or nonusers.

For these reasons, variables retained missing values in the machine-editing procedures when questions were skipped due to respondents answering a lead question as "don't know" or "refused" (or answering in a similar manner in response to a follow-up probe). If respondents refused a lead question (or refused all questions and again refused on follow-up, when respondents were requested to answer multiple lead questions, such as in the Inhalants section), refusal codes were assigned to all of the subsequently skipped items as part of the machine edits (i.e., the lead refusal was propagated). That is, it was logically inferred that the respondents were globally refusing to answer *any* questions on that topic.

In situations where respondents answered a lead question (or questions) as "don't know," values of "blank (no answer)" were retained in the skipped questions. Unlike the above situation for responses of "refused," it does not logically follow that a response of "don't know" to a lead question would imply that the respondent would answer "don't know" to all subsequent questions on that topic. Furthermore, respondents who answered a lead question as "don't know" were unsure whether the question applied at all. For example, if a respondent answered the lifetime marijuana question as "don't know," assigning a "don't know" code to the age-at-first-use variable (corresponding to question MJ02) would imply that the respondent was a lifetime user but did not recall the age when he or she first used.

# 3.0 Initial Editing and Coding Steps

Procedures for initially processing, cleaning, and editing the 2001 NHSDA data encompassed the following activities:

- coding of "OTHER, Specify" data,
- creation of "edit-ready" raw variables (done as part of the daily processing of transmitted data in 2001),
- initial processing of age-related variables,
- identification of "usable" cases,
- investigation of response patterns in records that otherwise met the minimum data requirements of the "usable case" criteria, and
- edits of "date-dependent" variables when the interview date was judged to be questionable.

A separate validity study also was conducted in 2001 for respondents aged 12 to 25. This study assessed the validity of self-reported drug use by collecting hair and urine samples from consenting respondents. Because the content of the core drug modules was almost identical in both instruments, initial editing and coding steps were conducted for combined data from the validity study ($n$ = 2,123) and the main study ($n$ = 68,929) for 2001 (total $N$ = 71,052). Where reference is made in this document to numbers of respondents affected by a particular edit or pattern in the questionnaire, these numbers could include validity study cases as well as main study cases for 2001.

## 3.1 Coding of "OTHER, Specify" Data

This activity took alpha answers that respondents had typed (e.g., specific other drugs used, specific other offenses for which respondents were arrested and booked in the past 12 months besides those listed, specific other payment sources of treatment) and converted them into numeric codes suitable for further analysis. In the remainder of this section, we refer to these alpha answers (or the numeric codes resulting from them) as "OTHER, Specify" data.

Coding of the "OTHER, Specify" variables was accomplished through computer-assisted procedures. For types of "OTHER, Specify" data where we had prior experience from 2000 or previous survey years, we based the 2001 coding procedures on our experience with these data. New coding classifications were developed for types of "OTHER, Specify" data that were introduced for the first time as part of the 2001 CAI instrument (e.g., other ways that respondents used heroin).

In implementing the coding procedures, "OTHER, Specify" responses were first converted to ALL CAPS because respondents could type essentially the same thing but use different combinations of upper- or lowercase characters. If an exact match was found between what the respondent keyed and an entry in the system (e.g., "ALCOHOL"), the computer-assisted procedures assigned the appropriate numeric code (in this example, 807 for alcohol). The system also included commonly encountered misspellings for drugs (e.g., "ALCHOHOL"). Answers that the respondent typed that did not match an existing entry were reviewed by analysts to determine whether an existing code should be assigned to the response or a new code should be assigned. Analysts also reviewed the codes assigned through the computer-assisted procedures to verify that codes were being assigned correctly.

The remainder of the discussion in this section focuses on issues related to coding of the "OTHER, Specify" data pertaining to drug use. Further details about coding of "OTHER, Specify" data pertaining to tobacco brands are presented in Section 4.4 in connection with the discussion on edits of the tobacco brand data. Similarly, further details about coding of "OTHER, Specify" data in noncore sections of the interview are presented in accompanying documents.

In situations where respondents reported use of (or treatment for) "some other drug," respondents could enter up to five responses. Respondents hit the "ENTER" key to move to the next available field. For the most part, respondents in 2001 specified only one "other" drug or specified only one drug at a time, if they specified use of more than one other drug. In some situations, however, respondents specified the same drug more than once or entered different responses for which the same numeric code applied (e.g., specific alcoholic beverages all would get a code of 807 for alcohol). When these occurred, we assigned codes in a manner that preserved the respondents' answers; we did not edit the data to delete the duplicate responses. (A special exception to this general rule is discussed at the end of this section.)

Within a single field, respondents also could type responses for which more than one drug code could be assigned, such as if two drugs were used in combination with one another. In these situations, we assigned the second code to the next available slot, as explained below. For example, if a respondent specified both "Drug A" and "Drug B" in the first field, and did not specify anything else, then the code for "Drug B" was assigned to the second specify field. In this situation, the order of the codes would reflect the order in which the respondent specified the drugs. Suppose, however, that a respondent specified "Drug A" and "Drug B" in the first field and specified "Drug C" in the second field. In this situation, we considered the second field to be already occupied by the response for "Drug C." For simplicity, we assigned the code for "Drug A" to the first field, kept the code for "Drug C" in the second field, and assigned the code for "Drug B" to the third field, which was the next available slot; we did not edit the data to move "Drug C" to the third slot in order to open up the second slot for "Drug B." Consequently, in situations where more than two "OTHER, Specify" drug codes were assigned, the order in which the responses appear may not always reflect the order in which a respondent keyed his or her answers. For this reason, the descriptions associated with "OTHER, Specify" variables in the 2001 CAI

codebook use the terminology "OTHER [DRUG] - SPECIFY 1," "OTHER [DRUG] - SPECIFY 2," and so on, as opposed to using "OTHER [DRUG], FIRST MENTION, "OTHER [DRUG], SECOND MENTION," and so on, as had been done in 1998 and earlier.

If respondents specified more information about "other" drugs than could be captured within five final "OTHER, Specify" variables, priority was given to assigning as many unique codes as possible. In particular, if respondents specified more information than could be captured in five variables but they made multiple mentions of the same drug, the redundant information for that drug was not coded. If more than five unique mentions of drugs occurred (i.e., after any redundant mentions had been eliminated), we did not code mentions of drugs that respondents had reported using in a previous question. For example, if a respondent reported lifetime use of LSD in question LS01a and then specified LSD as "some other hallucinogen," we would not keep the duplicate LSD mention in the "OTHER, Specify" data. However, if the respondent had *not* reported lifetime LSD use in question LS01a but specified use of LSD as "some other hallucinogen," we would retain this mention, so that the respondent would subsequently be inferred to be a lifetime LSD user (see Section 4.1). Further priority in retaining responses in the final drug codes was given to (a) mentions of illegal or prescription-type drugs, as opposed to drugs that are legally available without a prescription, and (b) mentions that were relevant to the category of interest (e.g., in hallucinogens, giving priority to mentions of hallucinogens over mentions of drugs that were not classified as hallucinogens).

## 3.2 Creation of "Edit-Ready" Raw Variables

The collected interview data were transmitted from the field to RTI International[1] as ASCII files, and daily SAS datasets were created. As noted in Section 2.1, this daily processing included assignment of standard NHSDA missing data codes.

In addition, daily processing of the transmitted data included processing of data from questions in the interview that allowed respondents to choose more than one response from a list. Respondents who wanted to choose more than one response did so by pressing the space bar between each number they typed. In the remainder of the section, these are referred to as "enter all that apply" questions because respondents were allowed to enter as many responses from the list as applied. For the pain relievers question PR04a, for example, respondents could indicate use of more than one of the list of pain relievers shown below the red line on the show card for pain relievers.

These "enter all that apply" data were captured in the order that respondents keyed their answers, which was not necessarily the order in which response categories appeared on show cards or on the computer screen. Although these data could have been kept in the order that the respondent reported

[1]RTI International is a trade name of Research Triangle Institute.

them, it often is useful analytically for all responses of a particular type to be grouped together. For example, if an analyst wanted to know how often people reported lifetime use of Oxycontin, it would be more straightforward for this information to be captured in a single variable. In contrast, if the miscellaneous lifetime pain reliever variables reflected whatever order the respondent reported use of different pain relievers, an analyst would have needed to check 18 different variables in the 2001 data to identify all of the times when the respondent reported use of Oxycontin.

For this reason, the daily processing of transmitted data remapped "enter all that apply" variables to correspond to the order in which response categories appeared to the respondent on a computer screen or show card. In addition, respondents could use a function key to answer "don't know" or "refused" as their first response to these "enter all that apply" types of variables. The CAI program then routed the respondent to the next appropriate question. For example, if a respondent reported nonmedical use of one or more prescription pain relievers shown below the thick red line on Pill Card A (i.e., question PR04 answered as "yes"), the respondent was asked in question PR04a to indicate which of the specific pain relievers he or she had used nonmedically. If the respondent answered "don't know," the program exited question PR04a and routed the respondent to question PR05 (i.e., nonmedical use of any other prescription pain reliever besides the ones shown on Pill Card A).

In situations where respondents answered "don't know" to an "enter all that apply" type of question, it would be reasonable to infer that the respondent did not know whether each particular item on the list applied. That is, we could infer that a response of "don't know" applied globally to every item on the list. Consequently, a code of "don't know" was assigned to each of the recoded "enter all that apply" variables as part of the daily processing of the transmitted data (i.e., we propagated the code of "don't know"). For example, if a respondent answered question PR04a as "don't know," we inferred that the respondent did not know whether he or she had ever abused codeine, Demerol, Dilaudid, and so on, through Ultram.

Similarly, if a respondent refused to answer an "enter all that apply" question, a refusal code was assigned to all of the recoded variables on that list as part of the daily processing of the transmitted data (i.e., we propagated the refusal code). For example, if a respondent refused to indicate which of the specific pain relievers he or she had used nonmedically in question PR04a, we inferred that the respondent was refusing to indicate whether he or she had ever abused codeine through Ultram.

Following daily processing of the data, each day's SAS dataset was merged with the cumulative transmission data until a complete data file was produced that contained transmitted cases as of the end of the quarter. Each quarterly data file underwent initial cleaning and processing (not done as part of the machine editing) to modify or correct field errors, such as erroneous ID entries by the field staff. The cleaned-up SAS dataset from each quarter served as the starting point for subsequent machine edits.

## 3.3 Initial Processing of Age-Related Variables

This step in the editing procedures had the following principal objectives:

- to establish a preliminary (or best available) age, based on what the respondent reported in the interview, for use by the statisticians in assigning a final age for each respondent; and
- to flag situations for further inspection when some question existed about the age data reported by the respondent.

A key aim of establishing the preliminary age (called BESTAGE) was to have an available age for test runs of other machine edit programs prior to availability of a final age from the statisticians. Having BESTAGE was particularly relevant for testing edits that were age specific. As noted above, for example, some subsequent edits assigned "bad data" codes if respondents persisted in reporting that they first used a drug at an age greater than their current age, despite being alerted that these data were inconsistent. Therefore, the BESTAGE variable could be used initially in other programs to check if edits for setting age-at-first-use variables to "bad data" were functioning as expected.

In addition, the BESTAGE variable was used by the statisticians as the starting point for finalizing each respondent's age. To aid the statisticians in identifying cases for further examination, various indicator or flag variables were created. These variables were designed to indicate the following:

- if respondents revised their age at any point in the interview;
- if respondents revised their age in such a manner as to cross a boundary between sampling strata (e.g., in going from the 12 to 17 age group to the 18 to 25 age group);
- if respondents revised their age in the Tobacco or Alcohol sections to move from being underage to being of legal age for possession or use of these products (e.g., changing one's age in the Alcohol module from 20 to 21);
- if respondents revised their age by more than 1 year during the course of the interview;
- the number of times that respondents revised their age (where applicable); and
- situations where there was a problem with the interview date (e.g., missing data, interview year other than 2001), because the interview date was used by the CAI program in conjunction with the respondent's date of birth to calculate an age.

## 3.4 Identification of "Usable" Cases

Another step in the editing procedures established the minimum item response requirements in order for cases to be retained for weighting and further analysis (i.e., "usable" cases). These procedures were designed to eliminate cases with unacceptable levels of item nonresponse (i.e., missing data),

thereby retaining cases with lower levels of missing data and reducing the amount of statistical imputation that would be needed for any given record. As discussed below, however, the usable case rules that were implemented were not the sole determinants of whether a case was counted as a final respondent.

Respondents were asked more detailed questions about different drugs only if they reported lifetime use of that drug (or lifetime use of one or more drugs within a broader category, such as hallucinogens).[2] Consequently, whether a respondent was a lifetime user or nonuser of drugs of interest could be readily determined from review of respondents' answers to the raw lifetime question(s) for that drug (or category of drugs).

A number of different options were investigated in the 1999 NHSDA for establishing usable case rules for the new CAI data. Several of these are discussed in greater detail in a report on editing and imputation of the 1999 CAI data.[3] Options for defining cases as usable also were reviewed by NHSDA expert consultants and discussed in February 2000.

The requirements for the final usable case rule that was decided upon are noted below.

1. The lifetime cigarette gate question CG01 had to have been answered as "yes" or "no." This requirement was set so that lifetime use or nonuse would be fully defined for at least one substance. Consequently, data about lifetime use or nonuse of cigarettes could be used in subsequent statistical imputations for other drugs where lifetime use/nonuse was undefined.

2. At least nine (9) of the following additional gates had to have answers of "yes" or "no": (a) chewing tobacco, (b) snuff, (c) cigars, (d) alcohol, (e) marijuana, (f) cocaine (in any form), (g) heroin, (h) hallucinogens, (i) inhalants; (j) pain relievers, (k) tranquilizers, (l) stimulants, and (m) sedatives. Crack cocaine was not included in the usable case rule because the logic for asking about crack cocaine was dependent upon the respondent having answered the lifetime cocaine question as "yes." Although the CAI instrument also asked about pipe tobacco, this was not included in the usable case rule because there was only one other question about pipe tobacco in addition to the gate question.

---

[2]In all modules except the Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives modules, the logic for asking more detailed questions about use of that drug was based on the answer to a single "yes/no" question (e.g., "Have you ever, even once, used marijuana or hashish?"). In the Hallucinogens through Sedatives modules, the logic for asking more detailed questions about use of that category of drugs was based on respondents' answers to multiple "yes/no" questions about the lifetime use or nonuse of specific drugs within that category (e.g., lifetime use or nonuse of the specific hallucinogens LSD, PCP, peyote, mescaline, psilocybin/mushrooms, Ecstasy, or "any other" hallucinogen).

[3]Chromy, J.R., Bowman, K.R., Giacoletti, K.E., Grau, E.A., Kroutil, L.A., Marsden, M.E., Myers, L.E., Packer, L.E., & Singh, A.C. (2000, January). *Editing and imputation diagnostics report for the 1999 NHSDA 6-month analysis* (report version prepared for February 1, 2000, meeting of expert consultants; RTI/07190.189). Research Triangle Park, NC: Research Triangle Institute.

Beginning in 2000, the CAI instrument included follow-up probes for respondents who initially refused to answer a gate question. Follow-up probes were included for the following modules that were relevant to the usable case rule: Cigarettes; Chewing Tobacco; Snuff; Cigars; Alcohol; Marijuana; Cocaine; Heroin; the specific hallucinogens LSD, PCP, and Ecstasy; the specific stimulant methamphetamine; and any use of inhalants, pain relievers, tranquilizers, stimulants, or sedatives (if respondents refused all lifetime use questions about these latter five drug categories). If respondents changed their initial refusal to a response of "yes" or "no," they were considered to have provided usable data to that drug's gate information. In particular, in the Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives sections, if respondents initially refused to answer all gate questions for a given drug but subsequently answered that drug's follow-up probe as "yes" or "no," they were considered to have satisfied the reporting requirement for that drug.

For the multiple gate drugs (i.e., hallucinogens through the sedatives), we considered the gate to have been answered if at least one lead lifetime question in the series was answered as "yes" or "no" (e.g., if at least one question in the series LS01a through LS01h was answered as "yes" or "no" for hallucinogens). We did not require all questions within a multiple gate series to have been answered as "yes" or "no" because the CAI logic considered respondents to be lifetime users of a given category of drugs if they answered at least one question in the series as "yes," even if they had missing data (i.e., responses of "don't know" or "refused") for other questions in the series. The CAI program subsequently routed such respondents to more detailed questions about use of that category of drugs, such as the age they first used and when they last used that drug. Setting a requirement that all questions within a multiple gate series had to have been answered as "yes" or "no" could have resulted in data from these follow-up questions being disregarded when the respondent had already reported that he or she was a user of that category of drugs.

Data for cases that did or did not meet the usable case criteria were passed on to the statisticians to determine the final status of interview records. Cases that did not meet the usable case criteria because they were ineligible were assigned a final status to indicate this.

However, the usable case rule was a necessary, but not sufficient, requirement for a case to be considered a final respondent. For example, cases that had sufficient data to meet the usable case criteria could still be dropped from further analysis if their interview data suggested potential response pattern problems, as discussed below.

## 3.5 Investigation of Response Patterns in "Usable" Records

Although conversion to CAI reduced or eliminated some data quality problems that could occur in a PAPI format, we also recognized that the CAI environment could encourage some respondents to enter nonsensical patterns of answers if they were not paying attention to questions or were not taking the

interview seriously for other reasons. Thus, even if a respondent had sufficient data to meet the usable case criteria, certain patterns of answers could call into question the overall validity of the respondent's data.

Therefore, we developed a data diagnostics program to screen for the following patterns of responses that might cause us to question the validity of the interview as a whole:

- high numbers of "yes" responses to lifetime use of specific hallucinogens, inhalants, or psychotherapeutics (i.e., pain relievers, tranquilizers, stimulants, or sedatives), which might indicate that respondents were indiscriminately keying data without paying attention to what they were entering;
- alternating "yes/no" responses to questions about lifetime use of specific hallucinogens, inhalants, or psychotherapeutics (or alternating patterns of "response entered/not entered" in the Psychotherapeutics sections), which might indicate some type of pattern-making;
- high numbers of illegal drugs that respondents reported using every day or just about every day in the past year or past month (where applicable), in which case one might question either the validity of the answers or the respondent's competence to complete the interview;
- high numbers of substances that respondents reported first using at age 1 or 2, which might indicate indiscriminate keying of 1s or 2s, especially given that the age-at-first-use questions followed gate questions where a response of 1 denoted "yes" and a response of 2 denoted "no"; and
- consistent keying of the same code (either 1 or 2) throughout one or more modules, which would suggest a pattern of indiscriminate answering.[4]

These patterns of responses were examined on a case-by-case basis in order to determine whether a case should be retained as a final respondent or dropped.

A total of eight cases in 2001 met the usable case criteria but were treated as nonrespondents because their responses were suspect based on one or more of the patterns described above. All of these were youths or young adults.

An additional nine cases were recommended for retention as respondents, but their data from one or more modules were wiped out (i.e., original responses replaced with "bad data" codes). This process included setting responses to bad data that would indicate they were lifetime users of a given drug. Unlike the cases described above, not all of these respondents were aged 25 or younger. Data for these respondents were typically set to bad data because the respondents keyed values of 1 or 2 to every

[4]An important change to the CAI instrument in 2001 was that response categories for certain consistency checks were changed from 1=yes/2=no to 4=yes/6=no. For example, if a respondent reported first using marijuana at age 1 or 2, the respondent could not use a response of "1" or "2" in question MJCC05 to verify that this age-at-first-use was correct. This change to the CAI instrument was designed to stop respondents if they might have been engaged in a pattern of keying responses of "1" or "2."

question asked in a module, beginning with the age-at-first-use question. Data for the lifetime variables for these cases were set to bad data as part of the edits for the lead lifetime drug use variables (see Section 4.1). In addition, flags were set in the editing program for the lead lifetime drug variables to wipe out data for related variables as part of the edits of "date-dependent" variables (see Section 3.6). For example, if a case was identified that had "bad" stimulants data, the lifetime stimulants data corresponding to responses in questions ST01 through ST05 were set to bad data as part of the lifetime drug use edits, and a flag was set to indicate that data subsequently needed to be set to bad data for related variables pertaining to nonmedical use of prescription-type stimulants (both core and noncore variables pertaining to stimulants).

In addition, three cases had data selectively wiped out for the ages when they started smoking cigarettes every day or when they first used the hallucinogen Ecstasy. This was done because other related incidence data for cigarettes or hallucinogens suggested that respondents had miskeyed answers for the ages in question, such as by entering only one digit of a two-digit age. These data also were set to bad data as part of the edit program for wiping out "date-dependent" data.

## 3.6 Edits of "Date-Dependent" Variables and Other General "Bad Data" Edits

The CAI instrument used the interview date information that was stored by the system to anchor the periods of reference for questions pertaining to the "past 30 days" and "past 12 months." Specifically, the past 30 day period was calculated as exactly 30 days prior to the date stored in the system.[5] Similarly, the past 12 month period was calculated as exactly 12 months prior to the date stored in the system. Thus, in the recency-of-use questions that asked respondents when they last used the drug of interest, the response category "within the past 30 days" included a "date fill" to remind respondents when the past 30 day reference period began for them. Similarly, introductions to specific questions about frequency of use of a particular drug in the past 12 months and past 30 days included date fills to remind respondents of the period they should be thinking about when answering these questions.

For 14 cases in 2001 (approximately 0.02 percent of the final respondents), the interview date that was originally entered was sufficiently problematic to call into question the respondents' answers to date-dependent questions in self-administered sections of the interview. For 11 of these 14 cases, the interview date had already been manually revised prior to preparation of a raw data file, based on investigation by the field operations staff.

In these situations where interview dates were problematic, we also considered subsequent data in the self-administered modules to be problematic when the data were related to or affected by the

[5]If a respondent broke off and then resumed the interview at a later date, the past 30 day and past 12 month reference periods were updated when the interview resumed. This was done because 30 days and 12 months from the date when an interview resumed might be later than the 30-day and 12-month periods when a respondent had previously been asked questions.

interview dates. For example, if a respondent's interview date was incorrect for some reason, the CAI program would calculate a 30-day reference period based on this incorrect interview date. Consequently, answers that the respondent gave on the number of days that he or she used different drugs in the past 30 days could reflect the number of days that the respondent used these drugs in a period other than the intended 30-day reference period.

However, this interview date issue did not present a problem if respondents never used a particular drug in their entire lives (or never engaged in other behaviors). If respondents never engaged in a behavior prior to when they were being interviewed, that presumably would be true regardless of the value stored for the interview date. Further, when respondents reported never having engaged in a particular behavior, the CAI program skipped them out of questions where the interview date would be important for establishing reference periods. For these reasons, cases where there was some question about the interview date were still retained as final respondents.

Instead of cases being dropped, we set the following types of self-administered questions to bad data if a respondent was routed to them:

- questions pertaining to behaviors in the past 30 days,
- questions pertaining to behaviors in the past 12 months,
- questions pertaining to the most recent time that an event occurred (e.g., when a respondent last used a drug of interest), and
- questions pertaining to the respondent's age when an event occurred (e.g., the age when the respondent first used a drug of interest).

Self-administered questions about age at first use and other ages when an event occurred were not directly related to the interview date but were indirectly related via respondents' ages. That is, respondents' ages were calculated by comparing the date of birth with the interview date. In turn, age-at-first-use and other age data in the self-administered modules were compared for consistency with the respondents' ages.

In these situations, respondents' answers were set to "bad data" in the raw variables before any further editing was done. By setting the responses to bad data in the raw variables, we could distinguish between situations where the data were deemed to be bad on input and situations where we might set a variable to bad data in subsequent edit steps because of inconsistencies with other data in the respondent's record.

This edit did not apply to gate questions that asked whether a respondent had ever engaged in a behavior (e.g., "Have you ever smoked part or all of a cigarette?"). As noted above, whether a respondent had ever engaged in a behavior prior to being interviewed would presumably not be dependent on the

value stored for the interview date. In situations where respondents reported lifetime use of a particular drug or other lifetime behaviors, this edit also did not apply to questions within a module that asked whether a more detailed behavior of interest was *ever* true for respondents (e.g., "Have you smoked at least 100 cigarettes in your entire life?").

This edit also did not apply to sections of the interview that were administered directly by the field interviewers (FIs). If an interview date was incorrect, FIs could exercise greater control over the reference period provided to respondents in the interviewer-administered questions.

As noted above, edits in this step of the procedures also wiped out data in situations in which data had been flagged based on potentially patterned responses that had previously been identified (see Section 3.5). Thus, for example, if a case had been flagged as having patterned responses in the stimulants data, data (other than lifetime stimulants variables) from both core and noncore modules were wiped out if they pertained to stimulants, including variables pertaining to symptoms of stimulant dependence or treatment for abuse of prescription stimulants.

# 4.0 Editing of the Self-Administered Core Drug Use Variables

The preceding section discusses issues that were relevant to both core and noncore sections of the interview. As noted in Section 1.0, the core sections of the interview included interviewer-administered demographic questions (e.g., date of birth, Hispanic origin, race, marital status) and self-administered questions on use of tobacco (i.e., cigarettes, chewing tobacco, snuff, cigars, and pipe tobacco), alcohol, marijuana, cocaine in any form, crack cocaine, heroin, hallucinogens, inhalants, and nonmedical use of prescription-type pain relievers, tranquilizers, stimulants, and sedatives. This section discusses edits of the core self-administered drug use variables. As noted previously, edits of core demographic variables that were intended to be read aloud by FIs are discussed in a companion document.

Edits of the core self-administered drug use variables encompassed the following key activities:

- edits of the lead lifetime use variables (i.e., gate questions), where respondents indicated whether they have ever used the drug of interest;
- edits of the recency-of-use variables, where respondents who indicated lifetime use of the drug indicated when they last used that drug;
- edits of the 12-month and 30-day frequency variables, where respondents who indicated use of a drug in the 12 months or 30 days prior to the interview indicated the number of days they used that drug in the period of interest; and
- edits of remaining variables in a module.

In connection with each of these edits, the discussion focuses on relevant issues or inconsistencies in the data that needed to be addressed through machine editing. As noted previously, however, the skip logic in the CAI instrument limited the chances for respondents to be routed to questions where they could give answers that were inconsistent with their answers to previous questions. For example, if respondents reported never using marijuana in question MJ01, the CAI program did not ask additional questions that would presume use of marijuana at least once. Consequently, most processing of the CAI data was relatively straightforward, and the issues discussed below were not widespread relative to the total number of 71,052 main and validity study respondents in 2001.

## 4.1 Edits of Lead Lifetime Use Variables

As discussed in connection with the usable case criteria (Section 3.4), CAI respondents were first asked whether they had ever used a drug of interest. For hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives, respondents were asked a series of questions to establish whether they had ever used one or more of specific types of drugs within that category (e.g., LSD, PCP, peyote). Only those respondents who indicated lifetime use of that drug (or lifetime use of one or more specific

hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, or sedatives) were asked more detailed questions about that drug (including situations in which respondents initially refused to answer a question about their lifetime use of a drug but then changed their answer to "yes" on follow-up). Hence, these lifetime questions also can be thought of as gate questions to the further questions about that drug. Similarly, the multiple questions about use of specific hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives can be thought of as multiple gate questions because for these drugs respondents had multiple opportunities to indicate that they used at least one drug within the category.

Processing of these gate variables established whether (a) respondents had used a drug of interest at least once, (b) they had never used the drug, or (c) lifetime use or nonuse of the drug could not be determined. In particular, respondents could answer these gate questions as "don't know" or "refused" in addition to answering them as "yes" or "no." As discussed in Section 2.2, final responses of "don't know" and "refused" to these gate questions were treated by the CAI skip logic as equivalent to situations where respondents never used the drug of interest. For the multiple gate drugs of hallucinogens through the sedatives, the CAI skip logic treated combinations of "no," "don't know," and "refused" in the absence of any affirmative response the same as if the respondent had answered all of the gate questions negatively. In these situations where a gate question was answered as "don't know" or "refused," we treated the respondent's lifetime use or nonuse status as unknown because these responses did not provide conclusive information one way or the other.[6] Cases with unknown lifetime use/nonuse status were subsequently imputed through statistical procedures to be lifetime users or nonusers.

For the multiple gate variables in the Hallucinogens through Sedatives sections, if respondents answered all questions in the series as "no," the edit procedures assigned codes of 91 to the entire series of gate questions.[7] This was done to indicate that the respondents had never used any of the drugs in that category.

For inhalants through sedatives, summary variables also were created to indicate whether respondents had ever used one or more drugs within that category (e.g., INHEVER for inhalants). If respondents answered "yes" to at least one gate question in the series for a given drug, the corresponding summary variable was coded as 1 (i.e., "yes").

---

[6]For multiple gate drugs, this included situations where respondents answered one or more of the gate questions as "don't know" or "refused" but did not answer any of the other gate questions affirmatively.

[7]In the Hallucinogens module, this included situations in which respondents initially refused to report whether they had ever used LSD, PCP, or Ecstasy but then changed their answer(s) to "no" on follow-up. Similarly, in the Stimulants module, this included situations in which respondents initially refused to report whether they had ever used methamphetamine but changed their answers to "no" on follow-up.

If respondents initially refused to answer all gate questions within one of these sections and then continued on follow-up to refuse to indicate whether they had ever used a drug within that category, the summary variable was assigned a code of 97 (i.e., refused). The individual gate questions within that module also retained codes of 97.

If respondents initially refused to answer all gate questions within one of these sections and then reported on follow-up that they had used some drug in that category at least once, the summary variable also was coded as 1. For example, if at least one affirmative answer appeared in questions about lifetime use of inhalants (including the follow-up question INREF), INHEVER was coded as 1.

If respondents were routed to a follow-up question because they refused all of that drug's gate questions and then they reported that they never used that type of drug, a code of 91 was assigned to the relevant summary variable. Codes of 91 also were assigned to all of the individual gate variables within that section. For example, if respondents initially refused to answer all questions about lifetime use of specific inhalants, they were asked whether they had ever used any type of inhalant. If respondents answered this follow-up question as "no," the summary variable INHEVER was assigned a code of 91, and all individual lifetime inhalant variables were assigned codes of 91 as well.

In contrast, a code of 2 (i.e., "no") in specific multiple gate questions had the following meaning:

- the respondent was a user of at least one drug in the category but had never used the specific drug of interest (e.g., if a respondent was a lifetime user of LSD but not PCP), or
- the respondent reported never using the particular drug of interest but answered other questions in the series as "don't know" or "refused" (e.g., if a respondent did not know whether he or she had ever used LSD but definitely knew that he or she had never used PCP).

If respondents had never used any of the drugs in a series, they also would not have been routed to questions where they could specify the use of some other drug in that overall category. In this situation, blank values in the raw "OTHER, Specify" drug variables were replaced with codes of 9991 to indicate that the questions were skipped because the respondents had never used that class of drugs.

Similarly, if respondents reported that they used at least one drug in the category but they never used some other drug besides the ones asked about, they were legitimately skipped out of the questions that asked them to specify what "other" drug they had used. In this situation, blank values in the raw "OTHER, Specify" drug variables were replaced with codes of 9999 (i.e., legitimate skip).

Functionally, the codes of 9991 and 9999 both indicate that respondents legitimately skipped out of the "OTHER, Specify" questions in that drug's section. However, the 9991 codes provide for analysts

the extra level of detail that the respondents were legitimately skipped out of these questions because they had never used *anything* within that category of drugs.

A similar edit was applied in the Psychotherapeutics sections (i.e., pain relievers, tranquilizers, stimulants, and sedatives) where respondents were asked if they had ever used any of the medications below the red line on that drug's pill card. If respondents answered "yes," they were asked to indicate which of the drugs they had used. If the respondents answered "no," they were skipped out of these follow-up questions. Therefore, if respondents had never used *any* prescription-type psychotherapeutic medications in that category (i.e., in addition to never having used any of the medications below the red line on that drug's pill card), the edits assigned codes of 91 to all of the specific drugs that were skipped (e.g., codeine through Ultram in the PR04a series). In contrast, if respondents reported never using any of the medications below the red line on the pill card but they reported use of at least one other drug (or they answered at least one other gate question as "don't know" or "refused"), the skipped drug questions were assigned codes of 99 (i.e., legitimate skip).

Exhibit 1 describes additional issues that were relevant to the processing of gate variables. All of these issues pertain to the content of respondents' answers when they specified use of some other drug. The exhibit also indicates what (if anything) was done in the editing procedures to address each of these issues.

If any edits were done to specific gate variables (e.g., the variable DARVTLYC, corresponding to question PR01 in the Pain Relievers module and referring to nonmedical use of Darvocet, Darvon, or Tylenol with codeine), the following codes were assigned:

| | | |
|---|---|---|
| 3 | = | Yes LOGICALLY ASSIGNED, |
| 4 | = | No LOGICALLY ASSIGNED, |
| 81 | = | NEVER USED [DRUG] Logically assigned. |

As an example of a situation where a code of 3 might be assigned to a gate variable, respondents could report in the Psychotherapeutics sections that they never used any of the medications below the red line on that drug's pill card but then specify something that would indicate that they had used at least one of the drugs in that group (e.g., specifying morphine as some other pain reliever). In this situation, we inferred that the corresponding lead question (e.g., ANLCARD in the Pain Relievers section) should have been answered as "yes." In the example given here, we also inferred that the particular drug that was specified should have been reported in that series (e.g., inferring a value of response entered in the variable MORPHINE to indicate that the respondent should have reported nonmedical use of morphine in connection with question PR04a).

**Exhibit 1. Data Issues Involving the Gate Questions**

| Issue | Edits Implemented |
|---|---|
| Respondents (Rs) could specify something that corresponded to a drug in the list. For example, if an R specified use of "marijuana laced with PCP" as "some other hallucinogen," this response would indicate PCP use—even if such use had not previously been reported. | If the corresponding drug in the list was not answered as "yes," then "yes" was logically inferred, and the "specify" response was retained to indicate to analysts where the logical inference came from. This was done because Rs may provide a final answer of "don't know" (DK) or "refused" (REF) to a list question but specify something to indicate that they have used the drug. This was especially relevant to LSD, PCP, and methamphetamine where lifetime use or nonuse was imputed for these specific drugs. Consequently, there would be no need to impute for these drugs if an R did not know whether he or she had used these drugs or refused to report their use when asked directly but subsequently specified something that indicated their use. |
| Rs could use street names or slang terms when specifying their use of some other drug besides the ones they previously had been asked about. Use of these slang terms to infer use of a drug that Rs had already been asked about (e.g., LSD, PCP) could be problematic, however, because the meaning of a particular slang term could vary by region, could apply to more than one drug, or could change over time. | Use of street or slang terms to infer lifetime use of a specific drug was generally restricted to situations where that term was supplied to respondents as a synonym for that drug. For example, "angel dust" was listed in question LS01b as an alternative term for PCP. Thus, situations where Rs specified angel dust were used to infer lifetime use of PCP, if question LS01b had not already been answered as "yes." However, other potential slang terms for PCP (e.g., "boat") that were not listed in question LS01b were not used in editing. Additional situations where slang terms were classified with particular drugs included "shrooms" (coded as Psilocybin/mushrooms), "X" and "X-TC" (coded as Ecstasy), "Roofies" (coded as Rohypnol), and "rock" (coded as crack cocaine).<br><br>In addition, question ST01 in the Stimulants module listed "speed" as an alternative term for methamphetamine. However, the decision was made that indications of speed in and of themselves should not be used to infer lifetime methamphetamine use because speed may often be used to refer to other stimulants or to stimulants as a group, not just methamphetamine. |
| Rs can specify use of some other drug within a particular category of drugs but the drug being mentioned applies to another drug category covered in the survey (e.g., specifying valium, a tranquilizer, when asked about nonmedical use of pain relievers). | No editing across modules was done when these types of responses occurred. However, the "OTHER, Specify" data within a module indicate to analysts when Rs have specified use of a drug that fits another category.<br><br>This approach assumed that some Rs may specify the use of other drugs according to their *functional* properties. For example, Rs may specify certain tranquilizers in the Sedatives section because tranquilizers can cause drowsiness. Similarly, the definition of hallucinogens (i.e., drugs that "often cause people to see or experience things that are not real") could apply to other drugs besides hallucinogens that alter one's perception of reality. |
| For the psychotherapeutics (i.e., pain relievers, tranquilizers, stimulants, and sedatives), Rs may indicate lifetime use of one or more prescription-type psychotherapeutics (e.g., Darvon, Darvocet, or Tylenol with codeine) but specify only over-the-counter (OTC) medications as the other drug that they used (e.g., specifying only Tylenol as the other prescription pain reliever that they used nonmedically). | Because Rs were instructed in the Psychotherapeutics section not to include use of OTC medications, edits in this situation inferred that Rs logically had not used some other drug in that category. However, Rs who reported lifetime nonmedical use of a prescription-type medication in response to a previous question were still at least lifetime abusers of that category of drugs. This edit also was done in situations where Rs answered at least one prescription-type gate question as DK or REF. This edit was *not* done if Rs reported use of "some other" psychotherapeutic drug and (a) specified use of a prescription-type medication and an OTC drug, or (b) had some response of DK or REF along with the indication of OTC use. In this latter situation, the assumption was that a response of DK or REF meant that the R was still potentially a nonmedical user of a prescription-type drug. |

*(continued)*

**Exhibit 1 (Continued)**

| Issue | Edits Implemented |
|---|---|
| For the psychotherapeutics, the *only* indication of lifetime use may be an OTC drug that was specified. (In this situation, unlike the one described above, Rs have denied ever using other prescription-type medications that were covered in the particular category.) | Because Rs were instructed in the Psychotherapeutics section not to include use of OTC medications, edits in this situation inferred that the R had never used prescription or street psychotherapeutics.<br><br>This edit was not done in situations where one or more other gate questions for a type of psychotherapeutic medication was answered as DK or REF because that was not conclusive evidence that the R had *never* used. |

Edits of gate variables that involved assignment of codes of 4 and 81 were limited to the psychotherapeutic drugs. Both of these edits applied to situations where respondents specified use of over-the-counter (OTC) medications despite being instructed not to report about use of OTCs.

Codes of 81 were assigned when the following situation occurred:

- the respondent answered "no" to all of the questions about lifetime use of specific medications in that category except for use of any other medication in that category (e.g., any other pain reliever besides the ones shown on Pill Card A); and
- the *only* thing the respondent specified was an OTC medication, subject to the qualifications discussed below.

Analogous to the situation described above where codes of 91 were assigned to all of the gate variables when respondents had never used anything in that category, codes of 81 were assigned to all of the gate questions in a Psychotherapeutics section if the only indication of use came from OTCs.[8]

The rationale for this edit was that respondents were instructed not to report about use of OTCs, as noted above. Respondents who used only OTC drugs would be correctly following instructions if they answered in a manner that caused them to be skipped out of remaining questions for that category of psychotherapeutic drugs—even if they used OTCs other than for their intended purpose. Consequently, the exclusive OTC users who were not skipped out of subsequent questions about use of that particular psychotherapeutic medication would comprise some unknown (and possibly unrepresentative) subset of exclusive OTC users.

This edit applied if the only responses in the "OTHER, Specify" data for a type of psychotherapeutic medication were OTCs, with the remaining "OTHER, Specify" variables having values of blank or bad data (i.e., where the bad data code denoted a nonsensical answer that the

[8]In the Stimulants section, this edit also involved assigning codes of 81 to the lifetime methamphetamine variable METHDES, even though lifetime nonuse was not really logically inferred. That is, respondents would already have answered this question as "no," indicating that they never used methamphetamine, Desoxyn, or Methedrine. However, the code of 81 was assigned to METHDES for consistency with the assignment of codes of 81 to the other stimulant gate variables.

respondent keyed). This edit was *not* implemented if respondents had "OTHER, Specify" responses of "don't know" or "refused" in addition to specification of OTCs; responses of "don't know" or "refused" that accompanied indications of OTC use in the "OTHER, Specify" data were interpreted to mean that the respondent was still a potential abuser of some prescription-type medication, especially in situations where respondents may not have known what they ingested.

Similarly, if other qualifying prescription-type medications were specified in addition to OTCs, that respondent's status as a user was retained (e.g., if a respondent reported nonmedical use of a prescription-type pain reliever in addition to use of aspirin in the Pain Relievers section of the interview). Further, the OTC responses were retained in the respondent's "OTHER, Specify" variables. In addition, if a respondent reported use of a drug that may be available over the counter in certain strengths but is available in other strengths only by prescription, then the respondent's status as a nonmedical user of that category of prescription-type psychotherapeutics did not change. For example, specification of ibuprofen or Motrin without a dosage could refer to use in prescription form, and we assumed this to be the case in the editing. However, specification of Advil (i.e., an OTC dosage of ibuprofen) would be an unambiguous indication of use of an OTC drug. Similarly, the decision was made to treat certain drugs as OTCs if they at one time had been available only by prescription but were now available over the counter without a prescription counterpart (e.g., Benadryl).

As noted in Section 3.5, this step of the editing procedures also involved wiping out lifetime drug data if potential patterned responses had previously been identified. For the single gate sections pertaining to snuff through heroin,[9] these edits involved wiping out data for the lifetime drug question and the follow-up probe, if respondents had previously refused the lead question (e.g., questions AL01 and ALREF for alcohol). For the multiple gate sections pertaining to hallucinogens through sedatives, these edits involved wiping out data in all lifetime variables in that section (e.g., data in questions LS01A through LS01H for hallucinogens and any associated "OTHER, Specify" data for hallucinogens).

## 4.2 Edits of Recency-of-Use Variables

Edits of the variables that establish when respondents last used a drug of interest are probably the most critical. These recency-of-use variables are the precursors for the final measures that establish the prevalence of use in the past 30 days, past 12 months, and lifetime.

Under PAPI, as a general rule, if a respondent indicated in one question on a drug's answer sheet that he or she had never used a substance and indicated in another question that he or she had used a

[9]Because of the requirement of the usable case criteria that respondents had to have defined data for lifetime use or nonuse of cigarettes (Section 3.4), any decision to wipe out data for cigarettes would have resulted in cases being reclassified as not usable. This did not occur in 2001.

substance, logical editing coded the person as a user of that substance. If a respondent reported two (or more) different answers on the same answer sheet with respect to how recently he or she had used a substance, the editing procedures typically assigned the category indicating the more recent use. Relatively little further statistical imputation was done to the PAPI recency variables following the editing step.

These edit procedures might compensate for the tendency of some respondents to underreport drug use or recent use. However, these procedures also could create false positives, such as respondents who truly had not used a drug in the past 30 days or past 12 months but were inferred to be users in one of these periods because of how they marked an answer sheet. In particular, these PAPI edit procedures could appreciably affect estimates of behaviors that were relatively rare (e.g., past month crack cocaine use), where the numbers of cases not requiring editing were small relative to the number of cases assigned to a category through editing.

The skip logic in the CAI instrument limited the kinds of information that were available for use in editing the CAI recency-of-use variables. In particular, respondents who answered a gate question as "no" (i.e., never used that drug) were not given the opportunity to answer additional questions as though they were users of that drug. Similarly, respondents who reported that they last used a drug "more than 12 months ago" were not given the opportunity to answer further questions in that module about use in the past 12 months or past 30 days, as though they were more recent users than what they had originally indicated.

The CAI instrument included follow-up probes for respondents who were lifetime users of a given drug but did not know or refused to report when they last used it. Respondents who initially did not know when they last used a drug were asked to give their "best guess" of when they last used it. Respondents who initially refused to report when they last used a drug were asked to reconsider answering the question. If respondents changed their initial answer of "don't know" or "refused" in response to these probes, the CAI logic routed them according to their revised answer. For example, if respondents initially refused to report when they last used marijuana but then reported last using it in the past 30 days, they were routed to questions about their frequency of marijuana use in the past 12 months and past 30 days, as if they had reported use in the past 30 days in the first place.

If respondents gave a definite period in these follow-up probes for when they last used the drug of interest, that information served as the starting point for subsequent editing of the drug's recency variable. In the absence of any inconsistencies between the recency-of-use answers in these follow-up probes and other data within a given drug's module (see below), these answers from the follow-up probe were accepted as final.

Respondents were assigned a code of 9 if they reported lifetime use of a drug but continued on follow-up to answer "don't know" or "refused" regarding when they last used it. This code of 9 had the following meaning: Used at some point in the lifetime LOGICALLY ASSIGNED. These respondents were eligible to be statistically imputed to be users in any period, including in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago.

The procedures used to edit the CAI recency-of-use variables were referred to as the "flag and impute" procedures. Under these procedures, the limited situations where potential inconsistencies existed between a respondent's answer to a drug's recency question and other data in that module were identified and flagged. Exhibit 2 lists the usual types of inconsistencies that could occur between a drug's recency variable and other variables in that drug's module, and how these inconsistencies were handled through the flag and impute procedures. In the situations described in Exhibit 2, these inconsistencies were handled by statistically imputing final values for the affected recency variable and the other variable(s) where the data were inconsistent with the respondent's original answer to the recency question.

Prior to implementing these flag and impute rules, initial edits checked for situations where a respondent's original answer to an age-at-first-use question might have been inconsistent with his or her recency of use, but a revised age at first use was not. For example, if a respondent reported first using a drug at his or her current age, the respondent was asked to verify this age at first use. If the respondent reported that this age at first use was *not* correct but then on follow-up did not know at what age he or she first used, or refused to answer, the edits updated the age at first use to reflect this "don't know" or refusal response. The rationale for this edit was that the respondent indicated that the initial answer was not correct. If the respondent reported last using a drug more than 12 months ago, the respondent's initial answer to the age at first use would have been inconsistent with the answer to the recency question. However, the respondent's follow-up answer of "don't know" or "refused" to an age at first use would not necessarily be inconsistent with a reported recency of more than 12 months ago. By implementing this type of edit prior to checking the age at first use for consistency with the recency, we aimed to prevent spurious inconsistencies from being flagged and unnecessary edits being done to the recency variables.

Initial edits in the tobacco module in 2001 also took into account errors in the CAI routing logic when respondents triggered consistency checks involving their age-at-first-use data. Although these errors were fixed through CAI instrument "patches" that were sent to the field during 2001 data collection, some data existed that reflected the earlier logic errors. The routing errors and how they were handled through initial editing are discussed below. Again, these edits were handled toward the beginning of the tobacco editing routine to prevent data from being flagged and imputed unnecessarily.

- If respondents answered "no" to questions CGCC07 or CGCC11 ("The computer recorded that you were DAILYCIG years old when you first started smoking cigarettes every day. Is this correct?"), they were supposed to be routed to subsequent consistency checks (CGCC10 or CGCC14, respectively) to correct their age when they started smoking cigarettes every day. However, some respondents were incorrectly skipped past these subsequent items. In these situations, the age at initiation of daily cigarette smoking was set to bad data.

- Similarly, if respondents said "no" to question CGSN01 ("The computer recorded that you were SNUFFAGE when you first used snuff. Is this correct?"), they were supposed to get routed to CGSN04 ("Please answer this question again. How old were you the first time you used snuff?"). Instead, they were getting skipped past CGSN04. The age-at-first-use for snuff was set to bad data for these cases.

**Exhibit 2. How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables**

| Type of Inconsistency | Edits Implemented |
|---|---|
| Recency originally indicates use in the past 30 days, but use on 0 days in the past 30 days is confirmed (suggesting less recent use). | Logically inferred that the respondent (R) was at least a past year user and potentially a past month user. The 30-day frequency was set to missing (i.e., bad data). |
| Recency originally indicates use more than 30 days ago but within the past 12 months, but the 12-month frequency indicates use on more than 335 days in that period (suggesting past month use). | Inferred that the R was at least a past year user and potentially a past month user. The 12-month frequency was set to missing. |
| Recency does not indicate use in the past 30 days, but the R reports first using the drug in the same month as the interview took place (suggesting past month use). | If the recency originally indicated use in the past 12 months, it was inferred that the R was at least a past year user and potentially a past month user. If the recency originally indicated lifetime use (or was missing), it was inferred that the R was at least a lifetime user (and potentially a past year or past month user). The values in the month- and year-of-first use that triggered the inconsistency with the recency-of-use answer were overwritten with bad data codes. |
| Recency does not indicate use in the past 30 days, but the R has other data suggesting initiation of use in the past 30 days (e.g., if first use was indicated at the R's current age and the R's last birthday was fewer than 30 days ago, or based on comparison of the 12-month frequency and the maximum number of days that the R could have used the drug). | Inferred that the R was at least a past year user and potentially a past month user. |

(*continued*)

**Exhibit 2 (Continued)**

| Type of Inconsistency | Edits Implemented |
|---|---|
| Recency does not indicate use in the past 12 months, but the age at first use equals the R's current age (suggesting past year use). | Inferred that the R was at least a lifetime user (and potentially a past year or past month user).<br><br>For tobacco products, if the lead 30-day question (e.g., CG05 for cigarettes) indicated that the R did not use in the past 30 days, then the edits assigned a code of 19 to indicate that the R was a user more than 30 days ago (but was potentially a past year user).<br><br>The age at first use that triggered the inconsistency with the recency-of-use answer was set to bad data. If the month of first use and year of first use were answered (i.e., not blank), the values in these variables were overwritten with bad data codes. (Month and year data were considered to be linked with the age-at-first-use data. Therefore, if the age at first use was questionable, then the month and year were considered to be questionable as well.) |
| Recency does not indicate use in the past 12 months, but the R reported first using in a month and year that falls within 12 months of the interview date. | Inferred that the R was at least a lifetime user (and potentially a past year or past month user).<br><br>For tobacco products, if the lead 30-day question (e.g., CG05 for cigarettes) indicated that the R did not use in the past 30 days, then the edits assigned a code of 19, to indicate that the R is a user more than 30 days ago (but was potentially a past year user).<br><br>The values in the month and year of first use that triggered the inconsistency with the recency-of-use answer were overwritten with bad data codes. |
| For tobacco products, recency indicates use more than 3 years ago, but age at first use indicates that the first use was within 2 years of the R's current age (suggesting use within the past 3 years). (Note that edits checking for indications of use at the R's current age were given priority over the condition described here because the former response pattern suggests use in the past year.) | Inferred that the R last used at some point more than 12 months ago (but potentially in the past 3 years); this was signified by a code of 14. The age at first use that triggered the inconsistency with the recency-of-use answer was set to bad data. If the month of first use and year of first use were answered (i.e., not blank), the values in these variables were overwritten with bad data codes. |
| For chewing tobacco and snuff, the brand of chewing tobacco that Rs reported using most often in the past 30 days was really a snuff brand, or vice versa. | Created a recoded any smokeless tobacco recency (SLTREC) that incorporated data from the chewing tobacco and snuff recency variables CHEWREC and SNFREC, respectively. Thus, for example, Rs who reported using chewing tobacco in the past 30 days but specified a snuff brand as the brand they used most often in that period would still be considered a past month user of some type of smokeless tobacco product. |

Note: Indications of most recent use include answers from follow-up probes for the recency questions.

- When respondents had inconsistent data for the ages when they first tried a cigarette (edited variable CIGTRY) or started smoking cigarettes every day (edited variable CIGAGE), the logic for resolving the inconsistency was reversed. In one situation, respondents who reported that the age when they first tried a cigarette was correct were asked to revise their age when they first tried a cigarette, instead of being asked to revise the age when they started smoking cigarettes every day. When this occurred, CIGAGE was set to bad data. Similarly, respondents who reported that the age when they first started smoking cigarettes every day were asked to revise that age for initiation of daily smoking, instead of being asked to revise the age when they first tried a cigarette. When this occurred, no editing was done because subsequent CAI edits set CIGTRY to be consistent with younger ages reported in CIGAGE.

Exhibit 2 also lists a special situation for chewing tobacco and snuff. When the CAI instrument was first fielded in 1999, we observed considerable cross-reporting of chewing tobacco and snuff brands among past month users, suggesting that respondents were not always clear about the differences between these two types of smokeless tobacco. For example, respondents could report using chewing tobacco in the past 30 days but specify a snuff brand as the brand of "chewing tobacco" they used most often in that period. However, this cross-reporting was identifiable only for respondents who reported past month use of either smokeless tobacco product but was assumed to be operating for respondents who reported less recent use. For this reason, a recoded smokeless tobacco recency variable SLTREC was created from the respective chewing tobacco and snuff recency variables (CHEWREC and SNFREC, respectively). Thus, if a respondent reported use of chewing tobacco in the past 30 days but specified use of a snuff brand in the past 30 days, the respondent was still a past month smokeless tobacco user.

In creating the recoded SLTREC, indications of more recent use of chewing tobacco or snuff were given precedence over indications of less recent use. In situations where one recency variable indicated use in a definite period (e.g., more than 30 days ago but within the past 12 months) and the second recency variable indicated use in an indefinite period (e.g., use at some point in the lifetime, which could have included use in the past 30 days, past 12 months, or past 3 years), the final assignment to SLTREC indicated a less definite recency value. The rationale for this procedure was that the respondent was potentially a user in a more recent period. For example, if a respondent indicated use of chewing tobacco more than 30 days ago but within the past 12 months and the flag and impute rules had assigned a code to the snuff recency to indicate that the respondent last used snuff at some point in his or her lifetime, the recoded SLTREC indicated use at some point in the past 12 months. That is, the report of chewing tobacco use in the past 12 months (but not past 30 days) could be used to narrow down the use of any smokeless tobacco to some point in the past 12 months, but the respondent could still have used in the past 30 days. Similarly, if one of the recency variables had a missing value but the other did not, the SLTREC variable was assigned a code to indicate that there was some uncertainty about when the respondent last used smokeless tobacco. Suppose, for example, that a respondent reported last using chewing tobacco more than 12 months ago but within the past 3 years but refused to report whether he or she had ever used snuff. In this situation, the SLTREC variable was given a code to indicate that the respondent used smokeless tobacco at some point in the lifetime because the respondent may have used snuff within the past 12 months or past 30 days.

Exceptions to the general flag and impute principles involved situations where inconsistencies existed between related recency variables (any cocaine and crack cocaine, any hallucinogen use and LSD, PCP, or Ecstasy use, any stimulant use and methamphetamine use). These are presented in Exhibit 3, along with a description of how the data were edited when specific types of inconsistencies occurred between related recency variables. In these special situations, indications of use of the specific drug (e.g., crack cocaine) that were more recent than that indicated for the general drug category (e.g., cocaine in any form) were used to logically infer more recent use of the general drug category. For example, not all

respondents might make the connection that crack cocaine fits within the broader category of cocaine in general. Therefore, if a respondent reported last using any cocaine more than 30 days ago but within the past 12 months and also reported last using crack cocaine in the past 30 days, the edit procedures inferred that the respondent logically had used cocaine in any form in the past 30 days. Overall, however, the relative prominence of editing and statistical imputation was reversed in the CAI editing and imputation procedures compared with that in PAPI.

In addition to the situations described in Exhibit 3, special edits were applied in situations in which respondents were (a) users only of LSD, only of PCP, or only of Ecstasy in the Hallucinogens module (e.g., LSD=1 and all other hallucinogen gate variables coded as 2), or (b) users only of methamphetamine in the Stimulants module (e.g., METHDES=1 and all other stimulant gate variables coded as 2). In the Hallucinogens module, respondents who were "pure" users only of LSD, PCP, or Ecstasy were asked the general hallucinogen recency but were skipped out of the corresponding LSD, PCP, or Ecstasy recency. In these situations, the value from the edited hallucinogen recency HALLREC was assigned to the recency variable (e.g., LSDREC for LSD) that had been skipped.

**Exhibit 3. How the Flag and Impute Edit Procedures Handled Inconsistencies Between Related Recency Variables**

| Recency Reported by Respondent | | Edited Recency | |
|---|---|---|---|
| **Specific Recency (i.e., crack, LSD, PCP, methamphetamine)** | **General Recency (i.e., any cocaine, any hallucinogen, any stimulant)** | **Specific Recency (i.e., crack, LSD, PCP, methamphetamine)** | **General Recency (i.e., any cocaine, any hallucinogen, any stimulant)** |
| (1) Indicates use in past month. | Indicates use that is less recent than the past month. | Retains the recency reported by the respondent (R). | Logically infers the R to be a past month user. Assigns a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED). |
| (2) Indicates use more than 30 days ago but within the past 12 months. | Indicates use at some point in the past 12 months (i.e., some other data suggesting past month use). | Retains the recency reported by the respondent. | Retains the nonspecific value indicating that the R has used at some point in the past 12 months. |
| | Indicates use more than 12 months ago. | Retains the recency reported by the respondent. | Logically infers the R to be *at least* a past year user (and potentially a past month user). Assigns a code of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED). |
| | Indicates use at some point in the lifetime (i.e., other data suggesting past month or past year use). | Retains the recency reported by the respondent. | Logically infers the R to be *at least* a past year user. Assigns a code of 8, as indicated above. |

(*continued*)

**Exhibit 3 (Continued)**

| Recency Reported by Respondent | | Edited Recency | |
|---|---|---|---|
| **Specific Recency (i.e., crack, LSD, PCP, methamphetamine)** | **General Recency (i.e., any cocaine, any hallucinogen, any stimulant)** | **Specific Recency (i.e., crack, LSD, PCP, methamphetamine)** | **General Recency (i.e., any cocaine, any hallucinogen, any stimulant)** |
| (3) Indicates use more than 12 months ago. | Indicates use at some point in the past 12 months. | Retains the recency reported by the respondent. | Retains a value indicating that the R is *at least* a past year user. |
| | Indicates use at some point in the lifetime. | Retains the recency reported by the respondent. | Retains a value indicating that the R is *at least* a lifetime user. |
| (4) Indicates use at some point in the past 12 months (i.e., other data suggesting past month use). | Indicates use more than 30 days ago but within the past 12 months. | Retains the value indicating use at some point in the past 12 months. | Assigns a value indicating use at some point in the past 12 months. |
| | Indicates use at some point in the past 12 months (i.e., other data suggesting past month use). | Retains the value indicating use at some point in the past 12 months. | Retains the value indicating use at some point in the past 12 months. |
| | Indicates use at some point in the lifetime (i.e., other data suggesting past month or past year use). | Retains the value indicating use at some point in the past 12 months. | Assigns a value indicating use at some point in the past 12 months. |
| (5) Indicates use at some point in the lifetime | Indicates use more than 30 days ago but within the past 12 months. | Retains the value indicating use at some point in the lifetime. | Assigns a value indicating use at some point in the past 12 months. |
| | Indicates use more than 12 months ago. | Retains the value indicating use at some point in the lifetime. | Assigns a value indicating use at some point in the lifetime. |
| | Indicates use at some point in the lifetime. | Retains the value indicating use at some point in the lifetime. | Retains the value indicating use at some point in the lifetime. |

Note: These edits take place after inconsistencies have been identified between a recency variable and nonrecency variable (e.g., between the recency and the age at first use). For hallucinogens/LSD/PCP and for stimulants/methamphetamine, these edits also take place after the R has revised one or more answers in response to a consistency check. Further, for hallucinogens/LSD/PCP and for stimulants/methamphetamine, any inconsistencies that remain between a given recency variable and other nonrecency variables following inconsistency resolution are transferred back into the recency variables prior to implementation of these edits. For example, if the original answer to the hallucinogen recency disagreed with the age at first use but the revised recency in response to the consistency check did not, then the recency would be updated to reflect the revised value. Prior to implementation of the edits shown in the exhibit, however, if the revised recency still disagreed with the age at first use, then the recency would be edited further to reflect the fact that the previous inconsistency still remained.

In the Stimulants module, respondents who used only methamphetamine were skipped out of the general stimulant recency question ST09. Therefore, when respondents were users only of methamphetamine, the value from the edited methamphetamine recency METHREC was assigned to the stimulant recency STIMREC. Similarly, if respondents reported lifetime use only of methamphetamine in question ST01 and "some other stimulant" in question ST05 but the respondent specified use only of

methamphetamine, the respondent could be inferred to be a "pure" methamphetamine user. In this situation, the value from METHREC also was assigned to STIMREC.

In addition, special patterns also could remain in the data for cocaine, hallucinogens, or stimulants after most inconsistencies had been addressed and some related data elements were missing. Specifically, respondents could indicate that they first used any cocaine, any hallucinogen, or any stimulant within 12 months of the interview date (e.g., first use at their current age) and indicate that they last used that drug more than 30 days ago but within the past 12 months (e.g., HALLREC = 2 for any hallucinogen use). Logically, then, if respondents who had used more specific drugs within a given category (e.g., LSD, PCP, or Ecstasy) had missing incidence data for one or more of their more specific drugs, it followed not only that they had to have *first* used these more specific drugs at some point in the past 12 months, but they also had to have *last* used them at some point within the past 12 months. In this situation, both the general recency (e.g., HALLREC) and more specific recency variables (e.g., LSDREC, PCPREC, or ECSREC) were set to values of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) so that they would be imputed consistently. If the general recency indicated past month use (e.g., HALLREC = 1) but a more specific recency (e.g., LSDREC) did not indicate use in the past 12 months, only the specific recency was set to a value of 8.

Similarly, respondents could indicate that they first used any cocaine, any hallucinogen, or any stimulant in the same month that they were interviewed and indicate that they last used the drug in the past 30 days but have missing incidence data for more specific drugs within a category. In this situation, respondents who had used a more specific drug within that category (e.g., LSD) also were inferred to have last used that drug in the past 30 days. The edits assigned a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED) to the more specific recency variables (e.g., LSDREC = 11).

## 4.3 Edits of Frequency-of-Use Variables

The CAI instrument included questions about the number of days that respondents used different drugs in the past 30 days or past 12 months (or the average number of days per week or days per month that they used in the past 12 months). These are referred to in this section as 30-day and 12-month frequency variables, respectively. In addition, the Alcohol section included a question about the number of days that respondents consumed five or more drinks per occasion in the past 30 days.

Data from these frequency questions can be used to distinguish between occasional and more frequent users of a drug. In particular, frequent users of alcohol and illicit drugs may represent a group who are potentially in need of substance abuse treatment or other services for their drug use. Similarly, regular users of tobacco products, such as people who smoked cigarettes every day in the past 30 days, probably represent a group that would have greater difficulty stopping their use of tobacco. In addition, the question about frequency of consumption of five or more drinks of alcohol per occasion in the past 30

days is used to construct measures of "binge" and heavy alcohol use in that period ("binge" alcohol use = consumption of five or more drinks in a single occasion on at least 1 day in the past 30 days; heavy alcohol use = consumption of five or more drinks in a single occasion on 5 or more days in that period).

For the 12-month frequency determinations, respondents could report their frequency of use in one of three ways: (1) use on an average number of days per week in the past 12 months, (2) use on an average number of days per month in the past 12 months, and (3) the total number of days they used in the past 12 months. In particular, respondents who used a drug regularly in the past 12 months might find it easier to report their frequency of use in one of the first two ways as opposed to figuring the total number of days they used in that entire period. Conversely, respondents who used on only a few days in the past 12 months might prefer the third reporting method. For respondents who chose this third reporting method, the 12-month frequency was the actual number of days that the respondent reported using the drug in the past 12 months (assuming no inconsistency with the 30-day frequency, as discussed below). For respondents who chose the first two reporting methods, the overall number of days that they used in the past 12 months was a calculated value. Specifically, answers in terms of the average number of days used per week in the past 12 months were multiplied by 52, and answers in terms of the average number of days used per month in the past 12 months were multiplied by 12 in order to yield a calculated 12-month frequency. Because these latter two response options were averaged responses over the past 12 months, no further adjustments were done to the calculated 12-month frequency value when respondents used the drug more than 30 days ago but within the past 12 months, and they did not initiate use at some point in the past 12 months.

If the lifetime gate question(s) and edited recency-of-use variable indicated that the respondent had never used the drug of interest,[10] then edits at this step assigned a code of 91 to the 30-day frequency variable (where applicable)[11] and a code of 991 to the final 12-month frequency variable (where applicable). For questions on drugs where respondents were asked to report their frequency of use in the past 12 months, codes of 91 (or 991) were assigned to the source variables pertaining to the preferred method of reporting the 12-month frequency (i.e., average number of days per week, average number of days per month, or total number of days used in the past 12 months), the average number of days per week, the average number of days per month, and total number of days used in the past 12 months.

Similarly, if the edited recency of use indicated that the respondent had used the drug but not in the period of interest, edits at this step assigned a code of 93 to the 30-day frequency variable, and codes of 93 (or 993) to the 12-month frequency variable and related source variables that were used to create

[10]For hallucinogens, inhalants, and the psychotherapeutics, this meant that the respondent had never used any of the drugs in that category.

[11]For alcohol, this edit also applied to other 30-day variables, including the variable on the number of days in the past 30 days that respondents had five or more drinks in a single occasion.

the 12-month frequency. However, if the respondent was potentially a user in the period of interest (i.e., there was some question about when the respondent last used the drug) and the CAI program had skipped the 30-day or 12-month frequency questions, then the skipped variables retained a blank code. For example, if a respondent reported last using marijuana more than 12 months ago, the CAI program would skip the questions pertaining to frequency of marijuana use in the past 12 months and past 30 days. If this respondent also reported first using marijuana at his or her current age, that would be inconsistent with the reported recency. As discussed in Section 4.2, the flag and impute edit rules would infer that this respondent was a user at some point in the lifetime, which could include use in the past 30 days or past 12 months. In this situation, the 12-month and 30-day marijuana frequency variables that had been skipped by the CAI program retained a blank value in case subsequent imputation might assign the respondent to a more recent category. In addition, as discussed in Section 2.2, if respondents refused the lifetime gate question(s) and were skipped out of the 12-month and 30-day frequency questions (where applicable),[12] the edits at this step assigned refusal codes to the skipped frequency questions (i.e., the refusal was propagated). However, if respondents were skipped out of the 12-month and 30-day frequency questions because they answered the lifetime gate question(s) as "don't know," the edits retained codes of "blank" in the frequency variables, for the reasons given in Section 2.2.

Exhibit 4 lists the major issues in editing the 12-month and 30-day frequency variables and how these issues were addressed. In particular, modules that contained both 12-month and 30-day frequency variables included consistency checks between these variables. A consistency check was triggered in situations where the number of days that respondents reported using the drug in the past 30 days exceeded the number of days that the respondent used in the past 12 months. If the respondent revised either the 12-month or 30-day frequency data (or both) to make them consistent (i.e., such that the 12-month frequency was greater than or equal to the 30-day frequency following any updates done by the respondent), data from the consistency checks were taken as final. If the 30-day frequency still was greater than the computed 12-month frequency despite a consistency check having been triggered, then the 12-month frequency was assigned a bad data code.

Beginning with the 2000 CAI instrument (and continuing in 2001), if respondents did not know how many days they used a drug in the past 30 days or refused to give an answer, they were asked to give their best estimate of the number of days that they used. Respondents could estimate their 30-day frequency by choosing the category most likely to contain the number of days they used the drug: 1 or 2 days, 3 to 5 days, 6 to 9 days, 10 to 19 days, 20 to 29 days, or all 30 days. These questions existed in 1999 for cigarettes and cigars but were expanded in 2000 to cover all drugs for which 30-day frequency information was requested (chewing tobacco, snuff, alcohol, marijuana, cocaine, crack cocaine, heroin, hallucinogens, and inhalants).

[12]For the tobacco variables through heroin, such a situation would occur if respondents initially refused the gate question and then refused again on follow-up.

**Exhibit 4. Issues in the Editing of 12-Month and 30-Day Frequency Variables**

| Issue | Edits Implemented |
|---|---|
| Number of days that the respondent (R) used in the past 12 months is greater than 365 days (e.g., if the R reported using 31 days per month on average, then 31 × 12 = 372 days). | Nonmissing values greater than 365 were trimmed back to a maximum of 365 days. A flag was set to indicate that this trimming was done. |
| Number of days that the R reported using in the past 30 days implies less use in the past 12 months than what the R reported in the 12-month frequency (e.g., if the R reported using on only 1 day in the past 30 days and did not use on 29 days, the maximum number of days that the R could have used in the past 12 months would be 336 days); applied to alcohol through inhalants (but not to LSD or PCP). | The 12-month frequency was trimmed to be consistent with the number of days that the R reported using in the past 30 days. The relevant flag variable indicated that this trimming was done. |
| Rs who answered the 30-day frequency questions as "don't kow" (DK) or "refused" (REF) had the opportunity to give their best estimate of the number of days they used in the past 30 days. These best estimate variables were categorical as opposed to being continuous (e.g., 1 = 1 or 2 days, 2 = 3 to 5 days, etc.). For alcohol through inhalants, the maximum possible value of the "estimated" 30-day frequency could imply less use in the past 12 months than what the R reported in the 12-month frequency. For example, if the R reported using a drug on "10 to 19 days" in the past 30 days, the R would not have used the drug on at least 11 days in that period. It would therefore be inconsistent for the R to have reported using the drug on more than 354 days in the past 12 months. | The 12-month frequency was trimmed to be consistent with the minimum number of days that the R estimated using in the past 30 days. For example, if the R estimated using a drug on "10 to 19" days in the past 30 days, it was assumed that the R could have used on 345 days in the past 12 months, which would be consistent with use on 10 days in the past 30 days (and nonuse on 20 days). This value would not be considerably different from that derived by assuming use on the maximum number of estimated days in the past 30 days (e.g., 354 possible days of use in the past 12 months, if the R reported using on 10 to 19 days in the past 30 days). |
| For Rs who estimated their 30-day frequency, the final 12-month frequency was greater than the estimated 30-day frequency. However, the number of days that the respondent reported using in the past 12 months was consistent with some, but not all, values in the range for the estimated 30-day frequency. | No editing was done to the 12-month frequency. For example, if an R reported using a drug on "10 to 19 days" in the past 30 days and reported using on 350 days in the past 12 months, that 12-month frequency is within the range of 345 to 354 days (i.e., based on the logical inference that the R did not use on 11 to 20 days in the past 30 days). Thus, a 12-month frequency of 350 days would be consistent with the estimated 30-day frequency, as long as the R used on 15 to 19 days in the past 30 days (and did not use on 11 to 15 days in that period). In this situation, minimum and maximum possible values were created for use by the imputation team in assigning a final 30-day frequency. This procedure narrowed the allowable range of the 30-day frequency for consistency with the 12-month frequency, with the final 30-day frequency being picked from within that narrowed range. In the above example, the allowable range for assigning the final 30-day frequency was 15 to 19 days, rather than 10 to 19 days. |

*(continued)*

**Exhibit 4 (Continued)**

| Issue | Edits Implemented |
|---|---|
| R initiated use at some point in the past 12 months. Relative to when the R was interviewed, the R therefore could not have used over the entire 12-month period. | The value of the 12-month frequency was reduced (i.e., prorated) according to the maximum possible allowable number of days that the R could have used in the past 12 months, relative to when the R was interviewed. If the R answered the 12-month frequency in terms of a total number of days used in that period and the answer was greater than the maximum number of days that the R could have used, the 12-month frequency was reduced to this upper limit.<br><br>The maximum possible allowable number of days was determined from the interview date and the month- and year-of-first use or the date of the R's last birthday, whichever yielded the smallest number of days that the R could have used. If the R reported first using in 2001 but did not report the month when he or she first used, it was assumed that the R potentially started use in January 2001. |
| R initiated use at some point in the past 30 days. Relative to when the R was interviewed, however, the 30-day frequency (or the minimum value of the "estimated" 30-day frequency) exceeded the maximum possible number of days that the R could have used the drug in that period. | If the R gave an exact number of days for the 30-day frequency, the value was reduced to agree with the maximum possible number of days that the R could have used the drug. If the R estimated the 30-day frequency as a range of a possible number of days, the range for the estimated 30-day frequency was reduced to agree with the maximum possible number of days that the R could have used. For example, if the R estimated using a drug on "20 to 29" days in the past 30 days but could have used the drug at most on 15 days in that period, the estimated 30-day frequency was revised to the range indicating use on "10 to 19" days. |
| The 30-day frequency is greater than 12-month frequency; this applied to alcohol through inhalants (but not to LSD or PCP). | The value for the 30-day frequency was retained and the 12-month frequency was set to a missing value (i.e., bad data). (Any of the corresponding 12-month frequency source variables that the respondent answered were set to bad data as well.) This edit conserved data reported by the R that suggests more frequent use in the past 30 days. Although we could have set both the 12-month and 30-day frequencies to missing (i.e., bad data) and imputed both, a drawback of this approach is that we might get a donor whose 30-day and 12-month frequency data were consistent but whose data were considerably different from what the recipient R originally reported, especially if the recipient R's original answer to the 30-day frequency was at or close to 30. |
| For cigarettes, chewing tobacco, snuff, and cigars, Rs who answered the 30-day frequency question as DK or REF were asked to give their best estimate of the number of days they used these tobacco products in the past 30 days. These best estimate variables were categorical as opposed to being continuous (e.g., 1 = 1 or 2 days, 2 = 3 to 5 days, etc.). Because there were no 12-month frequency questions for tobacco products, inconsistencies would not exist between these "estimated" 30-day frequency variables and other data for these tobacco products. | No editing was done to the corresponding 30-day frequency variables (i.e., leaving the original answer of DK or REF). A final value was assigned by the statistical imputation team. |

*(continued)*

**Exhibit 4 (Continued)**

| Issue | Edits Implemented |
|---|---|
| For alcohol through inhalants, if Rs gave their best estimate of the number of days they used in the past 30 days, it was possible for the bottom end of the range for an "estimated" 30-day frequency to exceed the value for the 12-month frequency (i.e., after any opportunities that the R had for inconsistency resolution). | The value indicating the range for the "estimated" 30-day frequency was retained, and the 12-month frequency was set to a missing value (i.e., bad data). For example, if the R reported using a drug on 5 days in the past 12 months but estimated use of that drug on "6 to 9" days in the past 30 days, there would be no way for these two answers to be consistent. Setting the 12-month frequency to bad data in this situation is consistent with the edit when Rs report an exact number of days for the final 30-day frequency that is greater than the final 12-month frequency. |
| The bottom end of the range for the final "estimated" number of days that the R used a drug in the past 30 days is *exactly* equal to the final 12-month frequency; this applied to alcohol through inhalants (but not to LSD or PCP). | The final, edited 30-day frequency was set to the value for the 12-month frequency. The code was retained for the "estimated" 30-day frequency to indicate to analysts where the value came from for the 30-day frequency. For example, if the R reported using a drug on exactly 6 days in the past 12 months and estimated using that drug on "6 to 9 days" in the past 30 days, the 30-day frequency was set to exactly 6 days. The two answers in this example are not necessarily inconsistent, as long as the R used on exactly 6 days in the past month. Thus, the response of "6 days" in the past 12 months could be thought of as "trapping" the possible number of days that the R could have used in the past 30 days. |
| The 30-day frequency has a final response of DK or REF (i.e., the R did not give a best estimate for the 30-day frequency), but the 12-month frequency indicates use on exactly 1 day. | The final, edited 30-day frequency was logically inferred to be 1 day. If the R last used in the past 30 days and used on only 1 day in the past 12 months, that 1 day of use had to have occurred in the 30 days prior to the interview. |
| The value for the final 12-month frequency is somewhere within the range for the final "estimated" number of days that the R used a drug in the past 30 days; this applied to alcohol through inhalants (but not to LSD or PCP). | No further editing was done because the data were not necessarily inconsistent. For example, if the R reported using a drug on 8 days in the past 12 months and estimated use of that drug on "6 to 9" days in the past 30 days, use of the drug on 6, 7, or 8 days in the past 30 days would still be consistent with use on 8 days in the past 12 months. In this situation, minimum and maximum possible values were created for use by the imputation team in assigning a final 30-day frequency. This procedure narrowed the allowable range of the 30-day frequency for consistency with the 12-month frequency, with the final 30-day frequency being picked from within that narrowed range. In the above example where the R reported using on 8 days in the past 12 months and 6 to 9 days in the past 30 days, the allowable range for the 30-day frequency would be 6 to 8 days, not 6 to 9, in order for the 30-day frequency to be consistent with the 12-month frequency. |
| In the Stimulants module, the R was a lifetime user only of methamphetamine and reported last using methamphetamine in the past 30 days or more than 30 days ago but within the past 12 months. | Values from the methamphetamine 12-month frequency variables were assigned to the corresponding stimulant 12-month frequency variables. The variable STBSTWAY (preferred way of answering the stimulant 12-month frequency questions, corresponding to question ST10) was assigned a code in the 20s to indicate that data had been assigned from the methamphetamine 12-month frequency variables. |

*(continued)*

**Exhibit 4 (Continued)**

| Issue | Edits Implemented |
|---|---|
| The 12-month frequency of methamphetamine use is greater than the corresponding frequency of any stimulant use. | The higher value from the methamphetamine frequency was assigned to the stimulant frequency. This also involved moving over the source variables from methamphetamine that were associated with the overall methamphetamine frequency. In addition, the methamphetamine variable associated with the preferred method of reporting the 12-month frequency (STBSTWAY) was bumped by a value of 20 to indicate that data had been moved over from the methamphetamine frequency variables. |
| The 30-day or 12-month frequency of crack use is greater than the corresponding frequency of any cocaine use. | The higher value from the crack frequency was assigned to the cocaine frequency. For the 12-month frequency, this also involved moving over the source variables from crack that were associated with the overall crack frequency. In addition, the cocaine variable associated with the preferred method of reporting the 12-month frequency (CCBSTWAY) was bumped by a value of 20 to indicate that data had been moved over from the crack cocaine variables. |
| The R was skipped out of the 12-month or 30-day frequency questions for cocaine because of his or her original answer to the cocaine recency, but nonmissing values existed in the corresponding frequency variables for crack. For example, if an R reported last using any cocaine more than 30 days ago but within the past 12 months but reported last using crack within the past 30 days, 30-day frequency data could exist for crack but not for any cocaine. | No editing was done to the 12-month or 30-day cocaine frequency that had been skipped. In the imputation stage, the cocaine 12-month or 30-day frequency was subsequently edited or imputed for consistency with the corresponding crack frequency data. |
| The R estimated the frequency of use of both cocaine and crack in the past 30 days, with the estimated frequency for crack being greater than the estimated frequency for any cocaine. | The "estimated" value for the crack 30-day frequency (CK30EST) was assigned as the estimate for the number of days that the R used any cocaine in that period. The code assigned to the "estimated" cocaine 30-day frequency variable CC30EST was bumped by a value of 10 to indicate that data had been moved over from CK30EST. |
| The "estimated" 30-day frequency of crack use is greater than the total number of days that the R reported using any cocaine in that same period. | The "estimated" value for the crack 30-day frequency (CK30EST) was assigned as the estimate for the number of days that the R used any cocaine in that period. The code assigned to the "estimated" cocaine 30-day frequency variable CC30EST was bumped by a value of 10 to indicate that data had been moved over from CK30EST. The edited cocaine 30-day frequency variable COCUS30A was assigned a bad data code to wipe out the R's original answer. For example, if the R reported using any cocaine on exactly 5 days in the past 30 days but reported using crack on "6 to 9 days" in that period (i.e., a code of 3 in question CK06DKRE), the R was logically inferred to have used any cocaine on "6 to 9 days" in the past 30 days. A code of 13 was assigned to CC30EST, corresponding to the code of 3 in CK30EST. In this example, it would be impossible for the R to have used crack on at least 6 days in the past 30 days but to have used any cocaine on only 5 days in that same period. |

*(continued)*

**Exhibit 4 (Continued)**

| Issue | Edits Implemented |
|---|---|
| The value for the cocaine 30-day frequency is somewhere within the range given by the "estimated" crack 30-day frequency. | Examination of preliminary data from 2000 indicated that all cocaine use in the past 30 days could not be automatically inferred to be crack use.<br><br>Therefore, if the value for the cocaine 30-day frequency was the minimum possible number of days that the R could have used crack in that period, the crack 30-day frequency CRKUS30A was set to be equal to the cocaine 30-day frequency COCUS30A. For example, if the R reported using cocaine on exactly 6 days in the past 30 days and estimated using crack on 6 to 9 days in that period, it could be logically inferred that the R used crack on exactly 6 days.<br><br>If the value for the cocaine 30-day frequency was greater than the minimum value for the estimated crack frequency, no further editing was done. However, information was provided to the imputation team regarding the narrower range in which to assign a final value for the crack 30-day frequency. For example, if the R reported using cocaine on 7 days and crack on 6 to 9 days, the allowable range for the final crack 30-day frequency would be 6 or 7 days. |
| Overall 12-month frequency does not agree with the reported preference for giving the answer (e.g., if the R originally indicated that he or she would give an answer in terms of a total number of days per year, answered the days per year question as DK/REF, and then gave an answer in average days per month). This is a function of the CAI logic, which keeps routing Rs through the series until they give a nonmissing answer or answer all 12-month frequency questions as DK/REF; this applied to the Alcohol through Psychotherapeutics sections. Similarly, the R could change his or her preferred way of reporting the 12-month frequency in response to a consistency check with the 30-day frequency (e.g., starting out as reporting in terms of the total number of days in the past 12 months but changing to average number of days per month in response to a consistency check); this applied to alcohol through inhalants (but not to LSD or PCP). | The value that best agreed with the final answer to the 12-month frequency was assigned to the best way variable. Thus, if an R started out as preferring to give an answer in total days per year but the final answer came from an average number of days per month, the preferred method of reporting would be consistent with the reporting in days per month. Any other 12-month frequency questions that the R answered that did not correspond with the final result were overwritten with LEGITIMATE SKIP Logically assigned codes. In the above example, the initial answer in terms of the total number of days used in the past 12 months would be overwritten, and the final answer of average number of days per month would be retained. |
| Frequency of consumption of five or more drinks in the past 30 days (DR5DAY, from AL08) is greater than the overall frequency of consumption of any alcohol in the past 30 days (ALCDAYS, from AL06), despite the presence of a consistency check. | If the R reported having five or more drinks per occasion on all 30 days, then the edits inferred that the R drank on all 30 days. If the R reported having five or more drinks per occasion on fewer than 30 days, then the value from the five-drink frequency DR5DAY was retained and the overall 30-day frequency ALCDAYS was set to missing (i.e., bad data).<br><br>This approach conserved data reported by the R that could indicate heavy alcohol use (i.e., consumption of five or more drinks on 5 or more days in the past month). |

*(continued)*

**Exhibit 4 (Continued)**

| Issue | Edits Implemented |
|---|---|
| Frequency of consumption of five or more drinks in the past 30 days (DR5DAY) is greater than the maximum possible value for the estimated frequency of consumption of alcohol in the past 30 days (AL30EST, from AL06DKRE). | If the R reported having five or more drinks per occasion on all 30 days, then the edits inferred that the R drank on all 30 days (i.e., the variable ALCDAYS was assigned a value of 30). This edit also was done if the R reported having five or more drinks per occasion on all 30 days and the R continued to answer the follow-up question AL06DKRE as DK or REF.<br><br>If the R reported having five or more drinks per occasion on fewer than 30 days, then the value from DR5DAY was retained, and the estimated 30-day frequency variable AL30EST was assigned a bad data value. For example, if an R reported having five or more drinks per occasion on 10 days in the past 30 days but estimated drinking alcohol on 6 to 9 days in that period, it would be impossible for these two answers to be consistent. |
| R drank on 1 day in the past month and the usual number of drinks was five or more on this 1 day, but frequency of consumption of five or more drinks per occasion is 0 (or is answered as DK/REF). | A value of 1 day was assigned to DR5DAY. If the R reported having five or more drinks on the 1 day when he or she drank, then the R logically had to have had five or more drinks on exactly 1 day in the past month. |
| R drank on 1 day in the past month and the usual number of drinks was less than five, but the frequency of consumption of five or more drinks per occasion is answered as DK/REF. | A code of 80 was assigned to DR5DAY, where 80 = NO OCCASION OF 5+ DRINKS IN PAST 30 DAYS Logically assigned. If the R reported having fewer than five drinks on the 1 day when he or she drank, then the R logically had no occasions of consuming five or more drinks in the past month. |
| Frequency of consumption of five or more drinks in the past 30 days (DR5DAY) falls within the range of the estimated frequency of consumption of alcohol in the past 30 days (AL30EST). | No further editing was done because the data were not necessarily inconsistent (see NOTE at end of exhibit). For example, if the R reported drinking any alcohol on 6 to 9 days in the past 30 days and reported having five or more drinks per occasion on 7 days in that period, these data would be consistent as long as the R drank any alcohol on 7, 8, or 9 days in the past 30 days. In this situation, minimum and maximum possible values were created for use by the imputation team in assigning a final 30-day frequency for alcohol. This procedure narrowed the allowable range of the alcohol 30-day frequency variable ALCDAYS for consistency with the five-drink frequency variable DR5DAY. The final 30-day frequency for alcohol was picked from within that narrowed range. In the above example where the R reported having five or more drinks per occasion on 7 days in the past 30 days and 6 to 9 days in that period, the allowable range for the 30-day frequency would be 7 to 9 days, not 6 to 9, in order for ALCDAYS to be consistent with DR5DAY. |

*(continued)*

**Exhibit 4 (Continued)**

| Issue | Edits Implemented |
|---|---|
| The R drank on more than 1 day in the past month and the usual number of drinks per day was five or more, but the reported frequency of five or more drinks per occasion is 0. | *Both* the usual number of drinks (i.e., NODR30A, from AL07) and DR5DAY (from AL08) were set to bad data. This approach was consistent with cognitive testing results that suggested that Rs sometimes incorrectly answer AL07 by indicating the total number of drinks they had *over the entire 30-day period* instead of the usual number of drinks they had per day. Consequently, these edits did not automatically assume that these Rs were binge alcohol users.<br><br>If DR5DAY had a value of DK/REF and the usual number of drinks reported was five or more, then DR5DAY retained the corresponding DK/REF value. |
| The number of days that the R reported having five or more drinks equals the total number of days that the R reported drinking any alcohol (DR5DAY = ALCDAYS), but the reported usual number of drinks is less than five (or is answered as DK/REF). | A code of 975 was assigned to NODR30A (i.e., the edited version of AL07) to indicate usual consumption of at least five drinks. |

NOTE: For five cases in 2001, a consistency check was triggered between the answers in questions AL06DKRE (estimated 30-day frequency) and AL08 (frequency of five or more drinks) because the CAI logic compared the minimum value of estimated 30-day frequency with the answer in AL08. These respondents changed their answers as a result of these consistency checks being triggered. For these five cases, data from the consistency checks were disregarded and the original answers were allowed to stand because the original answers were consistent.

For alcohol through inhalants, a consistency check was triggered if the number of days that respondents reported using a drug in the past 12 months was lower than the minimum value for the number of days that respondents estimated using that drug in the past 30 days. For example, it would be inconsistent for a respondent to report using marijuana on "6 to 9" days in the past 30 days and also to report using it on fewer than 6 days in the past 12 months.

Exhibit 4 discusses issues related to the introduction of these questions that asked respondents to estimate their 30-day frequency for alcohol through inhalants, if they previously did not give an exact answer. For example, if the minimum value for the estimated 30-day frequency for a drug (e.g., 6 days, if respondents estimated using a drug 6 to 9 days) continued to be greater than the value for the 12-month frequency (e.g., 5 days), the edited 12-month frequency was assigned a bad data value.

In addition, if the value for the 12-month frequency fell within the range of the estimate for the 30-day frequency (e.g., if a respondent reported using on 8 days in the past 12 months and on 6 to 9 days in the past 30 days), maximum and minimum values were created for the estimated 30-day frequency. In the above example where a respondent reported use on 8 days in the past 12 months but estimated using the drug on 6 to 9 days in the past 30 days, use on 6 to 8 days (as opposed to 6 to 9 days) in the past 30 days would be consistent with the respondent's answer to the 12-month frequency. This information on

the maximum and minimum possible number of days that a respondent could have used a drug in the past 30 days was subsequently used by the imputation team to assign a final value to the 30-day frequency.

Exhibit 4 also discusses issues related to the question on the usual number of drinks consumed in a given day in the past 30 days (question AL07 and the final variable NODR30A). Although this is not a frequency variable per se, information from this variable was used to edit the 30-day frequency data for alcohol. For example, if a respondent reported having five or more drinks per occasion on *exactly* the same number of days that he or she reported drinking any alcohol in the past 30 days, then it would logically follow that the respondent's usual number of drinks per day had to have been five or more. Similarly, if a respondent drank on only 1 day in the past 30 days and reported having fewer than five drinks on that 1 day in question AL07, but the respondent answered question AL08 as "don't know" or "refused," it logically would follow that the respondent could not have had five drinks on any occasion in the past 30 days.

Although a recoded recency variable SLTREC was created for any smokeless tobacco use, we did not create a measure of the number of days that respondents used any smokeless tobacco in the past 30 days. If respondents reported use of both chewing tobacco and snuff in the past 30 days, it would have been possible for use of both smokeless tobacco types to have overlapped to varying degrees in the past 30 days. However, we did not know this degree of overlap for such respondents. To create a recoded 30-day frequency of any smokeless tobacco use, we therefore would have had to make assumptions (e.g., picking the maximum of the two) that could not have been confirmed from the data.

## 4.4 Edits of Additional Core Drug Use Variables

This section describes the following additional issues that were relevant to the editing or processing of the remaining core drug data:

- processing of variables that were relevant to all lifetime users of a drug,
- processing of the tobacco brand data, and
- processing of miscellaneous 30-day cigarette variables.

As described previously, if respondents reported never having used a drug, the CAI skip logic precluded the possibility of respondents providing data that would suggest that they were users. Consequently, final processing of all remaining core variables involved assigning codes of 91 (or 991, etc.) when respondents had never used a drug. Similarly, for tobacco brand and miscellaneous cigarette use questions that pertained to use in the past 30 days, processing also involved assignment of codes of 93 (or 993, etc.) when respondents had used a type of tobacco but definitely not in the past 30 days.

### 4.4.1 Processing of Lifetime Variables

In all core modules except for pipe tobacco, respondents were asked how old they were when they first used the drug of interest. If respondents reported first using the drug within 1 year of their current age, they were asked to report the specific month and year when they first used, with the allowable years ranging from 1999 to 2001. If respondents reported first using the drug at their current age and their birth month was earlier than the interview month (i.e., they reached their current age in the same year that they were interviewed), the CAI program assumed that the first use of the drug occurred in the current year (i.e., 2001). These respondents were asked only for the month that they first used in the current year. The remaining respondents who first used a drug within 1 year of their current age could be routed to one of two possible questions on the specific year they first used. They were then routed to a question to report on the specific month that they first used the drug in the year they had previously reported.

Because the routing logic to the different versions of the month-of-first-use and year-of-first-use questions was mutually exclusive, we created a single, composite set of month-of-first-use and year-of-first-use variables from the individual raw variables. In addition, if respondents indicated a specific year that they first used a drug, the final year-of-first-use variables for 2001 were recoded to replace raw codes with values for the specific years (i.e., 1999 through 2001). If respondents confirmed that they first used a drug at their current age and were interviewed subsequent to their birthday, we assigned a code of "2001" to the year of first use; this was done even if respondents did not know what month they first used in the current year, or if they refused to report what month they first used in the current year. If the month- and year-of-first-use questions had been skipped because respondents first used the drug more than 1 year younger than their current ages, we assigned legitimate skip codes to the final month- and year-of-first-use variables.

Exhibit 5 presents information on additional issues involved in editing variables relevant to lifetime users of a drug and how these issues were addressed. Several of these additional issues involve the age-at-first-use, month-of-first-use, and year-of-first-use variables, such as situations where the age at first use was inconsistent with the respondent's current age or where the reported month and year of first use of a drug was inconsistent with the age at first use.

In addition, we created recoded variables for respondents' ages when they first used any smokeless tobacco product (i.e., chewing tobacco or snuff), and the month and year when they first used, if applicable. If respondents had a missing value for one of the types of smokeless tobacco (i.e., "don't know," "refused," "bad data"), we retained the missing value in the recoded smokeless tobacco variables described above. For example, if a respondent had used both chewing tobacco and snuff, reported an age when he or she first used chewing tobacco but refused to report the age when he or she first used snuff, the respondent may have used snuff at a younger age than was reported for chewing tobacco. The recoded month and year that respondents first used smokeless tobacco were subsequently edited to be

**Exhibit 5. Miscellaneous Core Edit Issues Involving Lifetime Use Variables**

| Issue | Edits Implemented |
|---|---|
| Age at first use is greater than the respondent's (R's) reported age; this applied to all drugs except pipes and also applied to LSD, PCP, and Ecstasy within the Hallucinogens module. (Beginning in 2000, Rs who were lifetime users of LSD or PCP were asked to report the ages when they first used these specific hallucinogens, in addition to their ages when they first used any hallucinogen. In 2001, a similar question was added for Ecstasy.) | The final age was accepted as the standard, and the inconsistent age at first use was set to bad data. Any month and year data associated with the age at first use also were set to bad data. |
| Rs in the Hallucinogens module were lifetime users only of LSD, only of PCP, or only of Ecstasy. The CAI logic skipped these respondents out of questions related to the age, month, and year that they first used the specific hallucinogen. | Data on initiation of hallucinogen use were moved over to the corresponding specific hallucinogen variables that had been skipped. For example, if the R had used only LSD, the hallucinogen age (from HALLAGE) was assigned to the LSD age-at-first-use (LSDAGE). Similarly, if the R had provided data on the month and year that he or she had first used any hallucinogen, these data were assigned to the LSD month and year variables LSDMFU and LSDYFU. |
| Age at first use of a more specific variable (e.g., crack cocaine) is younger than age at first use of a related variable (e.g., any cocaine); this applied to first use of cigarettes and age at initiation of daily smoking, and to ages at first use for cocaine/crack, hallucinogens/LSD/PCP/Ecstasy, and stimulants/methamphetamine. | The younger age at first use was assigned from the more specific variable. For example, if the age at first use of crack cocaine was younger than the age at first use of any cocaine, then the edits assigned the value from the crack age at first use to the cocaine age at first use.<br><br>An exception is noted below for cigarettes. |
| Age at first use of any cigarettes (i.e., CIGTRY) is answered as "don't know" (DK) or "refused" (REF), but the R reports starting to smoke daily at age 1 (i.e., CIGAGE = 1). | The value of DK/REF was retained in CIGTRY (i.e., the age at first use), and the daily cigarette age variable CIGAGE was set to bad data. The rationale for making this exception was that a respondent would logically have to have tried a cigarette at age 1 or younger in order to have started smoking daily at age 1. |
| The Rs in the Cocaine/Crack, Hallucinogens, or Stimulants sections indicated first use of any drug in that category (e.g., any hallucinogen) at their current age but had missing data for the age when they first used specific drugs in that category (e.g., LSD) | The Rs were logically inferred to have first used the specific drug at their current age. As discussed in Section 4.2, these Rs also were inferred to have last used these drugs at least at some point in the past 12 months. |
| The Rs in the Cocaine/Crack, Hallucinogens, or Stimulants sections indicated first using any drug in that category (e.g., any hallucinogen) in the current year (i.e., 2001) but had missing data for the year when they first used specific drugs in that category (e.g., LSD) | The Rs were logically inferred to have first used the specific drug in the current year; these Rs also were inferred to have last used these drugs at some point in the past 12 months (Section 4.2). If the Rs also reported first use of the general drug (e.g., any hallucinogen) in the same month that they were interviewed and they had missing month-of-first-use data for specific drugs within that category, the Rs also were logically inferred to have first used the specific drug in the same month that they were interviewed. Otherwise, if the Rs had originally reported a month when they first used a specific drug but that month had been wiped out due to the "flag and impute" edits, the month-of-first use (e.g., LSDMFU) also was reset to the originally reported value if it matched the month-of-first-use for any drug in that category (e.g., HALMFU). |

*(continued)*

**Exhibit 5 (Continued)**

| Issue | Edits Implemented |
|---|---|
| The Rs in the Cocaine/Crack, Hallucinogens, or Stimulants sections indicated first using any drug in that category (e.g., any hallucinogen) in the year immediately prior to the current one (i.e., 2000) but that point of initation was within 12 months of the interview date. These Rs also had missing data for the year and month when they first used specific drugs in that category (e.g., LSD) | If the Rs had originally reported year and month data for when they first used a specific drug but these data had been wiped out due to the "flag and impute" edits, this information was restored *only* if the original answers matched the month- and year-of-first use for any drug in that category. For example, if LSDMFU and LSDYFU had been set to bad data but the original answers for the LSD month and year matched the month and year in HALMFU and HALYFU, LSDMFU was equated to HALMFU, and LSDYFU was equated to HALYFU. |
| Month- and/or year-of-first-use are inconsistent with the age at first use, but both are consistent with recency (i.e., when compared with the interview date, the month and year of first use suggest a different age of initiation than what the R reported in the age at first use); this applies to all drugs except pipes and also applied to LSD, PCP, and Ecstasy in the Hallucinogens module and to methamphetamine in the Stimulants module. | The age at first use was accepted as the standard in 2001. The value for the age-at-first-use was retained, and the month and year were set to missing (i.e., bad data). (Consistency checks have been added beginning in 2002 that allow respondents to revise the age at first use or the month and year information to make them consistent.)<br><br>This approach conserved age data reported by the R that suggests the R initiated use within 1 year of his or her current age. This approach also implicitly assumes that the age at first use is more valid than the month and year. Although we could have set all incidence data to missing and impute everything (which would not make a judgment about which answer is the more valid one), this would have caused us to lose information suggesting that the R had recently initiated use. |
| The month- and/or year-of-first-use questions are answered because the initial answer to an age at first use is within 1 year of the R's current age, but the final, edited age at first use is more than 1 year younger than the R's current age (e.g., if the raw age at first use for cocaine was within 1 year of the R's current age but the age at first use of crack was more than 1 year younger than the R's current age). | The original answers to the month and year of first use were overwritten with logically assigned legitimate skip codes. If the R had originally answered the relevant age at first use as being more than 1 year younger than his or her current age, the CAI program would have skipped the R past the questions about the month and year of first use. |
| The age at first use has a value of DK or REF, including situations where this assignment has been made from the consistency check data. | If the month- and year-of-first-use questions were skipped because the R answered the age at first use as DK or REF to begin with, the DK or REF value from the age at first use was propagated onto the skipped month- and year-of-first-use variables. This edit was designed to indicate the reason that the month- and year-of-first-use variables had been skipped. In addition, because the R may have first used the drug within 1 year of his or her current age, the month- and year-of-first-use questions may have been relevant to the R.<br><br>If the R had answered the month- and year-of-first-use questions but the final age at first use had a value of DK or REF (i.e., due to a consistency check response), the month- and year-of-first-use data were overwritten with the corresponding DK or REF value from the edited age at first use. Retaining the month- and year-of-first-use data in this situation would imply that the R used within 1 year of his or her current age. |

*(continued)*

**Exhibit 5 (Continued)**

| Issue | Edits Implemented |
|---|---|
| The month of first use has been skipped because the R answered the year of first use as DK or REF. | The DK or REF value from the year of first use was propagated onto the skipped month of first use. That is, if the R did not know in what year he or she first used the drug, it was assumed that the R would not know the month either. Similarly, a refusal to answer the year of first use was interpreted to be a blanket refusal to answer the month as well as the year. |
| For cocaine/crack, hallucinogens/LSD or PCP, or stimulants/ methamphetamine, month- and year-of-first-use data exist for related drugs, but prior edits described above were implemented to assign the younger age at first use from the more specific drug variable (e.g., crack). | The associated month and year of first use from the more specific drug (e.g., crack) were assigned to the more general drug's month and year data (e.g., any cocaine). However, this edit was not done in the Hallucinogens module when the sole hallucinogen use involved use of LSD or PCP. Consequently, some inconsistencies could remain in the edited hallucinogens data. |
| For cocaine/crack, hallucinogens/LSD or PCP, or stimulants/ methamphetamine, the ages at first use are the same for the related drugs, but the month or year of first use is earlier for the more specific drug (e.g., crack). | The earlier month and year data from the more specific drug (e.g., crack) were assigned to the more general drug's month and year data (e.g., cocaine). As above, this edit was not done in the Hallucinogens module when the sole hallucinogen use involved use of LSD or PCP. Consequently, some inconsistencies could remain in the edited hallucinogens data. |
| The CAI program had skipped lifetime cigarette questions pertaining to periods where the R had smoked daily for at least 30 days (CG15) or whether the R had smoked 100 or more cigarettes (CG16a). Answers to prior questions already indicated that the R had engaged in these behaviors. For example, if the R had reported smoking every day in the past 30 days, the R had to have had a point in his or her life when he or she smoked cigarettes every day for at least a month. Similarly, it could be inferred from answers to questions about the number of days that Rs smoked in the past 30 days and the number of cigarettes they usually smoked per day that they had smoked at least 100 cigarettes in their lifetime. | It was logically inferred in CIGDLYMO (the edited version of CG15) or in CIG100LF (the edited version of CG16a) that the R had engaged in the behavior of interest (i.e., ever smoked cigarettes daily for at least 30 days or smoked 100 or more cigarettes in the R's lifetime). A code of 5 was assigned. This code had the following meaning:<br><br>5 = Yes LOGICALLY ASSIGNED (from skip pattern)<br><br>This code was used instead of a legitimate skip code to signify that the CAI logic had skipped the R out of the question because the R had already provided data to indicate that the question should have been answered affirmatively. |
| The R was routed to the 100-cigarette question (CG16a) because the R estimated the number of days that he or she smoked cigarettes in the past 30 days. The R did not answer CG16a as "yes," but the minimum number of days that the R could have smoked in the past 30 days and the minimum usual number of cigarettes smoked per day suggest that the R smoked 100 or more cigarettes in his or her lifetime. | It was logically inferred in CIG100LF (i.e., the edited version of CG16a) that the R had smoked 100 or more cigarettes in his or her lifetime. For example, if an R reported smoking "10 to 19 days" and reported smoking "16 to 25 cigarettes per day," then the R would have smoked at least 100 cigarettes in the past 30 days, even if he or she smoked on only 10 days and smoked only 16 cigarettes per day. |

consistent with the age at first use that was chosen. If respondents initiated use of both types of smokeless tobacco at the same age and were asked the month and year that they first used (i.e., the first use was within 1 year of their current age), the recoding procedures picked the earliest year. If they reported first using both types of smokeless tobacco in the same year, the recoding procedures picked the earliest month.

### 4.4.2 Processing of Tobacco Brand Variables

The CAI instrument included questions to identify the specific brands of tobacco that were most commonly used by current (i.e., past month) users of cigarettes, chewing tobacco, snuff, and cigars. For types of tobacco other than cigarettes, respondents who were past month users could choose from a list of brands shown on the computer screen or they could indicate use of "a brand not on this list."

Respondents who gave the latter answer were asked to type in the name of the specific brand that they used. For cigarettes, the listing of brands was split between two different computer screens. Those respondents who identified a brand on the first list were not necessarily shown the second list (see below for exceptions). However, if respondents could not find their most commonly used brand of cigarettes on the first screen, they could indicate use of a brand not on the list. They were then routed to a list of brands on the second computer screen. If they could not identify their brand on the second list, they could again indicate use of a brand not on the list and were then asked to type in the name of the brand of cigarettes they used. For all tobacco types, if respondents reported use of a brand not on the list and specified an answer, they were asked no further questions about their preferred brand for that tobacco type.

In developing a coding scheme for the "OTHER, Specify" tobacco brand data, we assigned the same basic set of codes to a particular brand of tobacco regardless of whether that response came from the Cigarettes, Chewing Tobacco, Snuff, or Cigars section within the Tobacco module. Stated another way, the coding took into account situations where respondents might have specified a brand that actually corresponded to a different type of tobacco (see below). The final coding classification scheme was as follows:

- codes of 101-199 and 1001-1999 were reserved for cigarette brands,
- codes of 201-299 and 2001-2999 were reserved for chewing tobacco brands,
- codes of 301-399 and 3001-3999 were reserved for snuff brands,
- codes of 401-499 and 4001-4999 were reserved for cigar brands,
- codes of 501-599 and 5001-5999 were reserved for pipe tobacco brands, and
- codes in the 600-699 and 700-799 series were reserved for miscellaneous tobacco and non-tobacco responses.

In particular, the coding procedures allowed for coding to resume in the thousands series if available codes were exhausted in the hundreds series. For example, questions CG11 and CG11a for cigarettes had already listed 60 brands of cigarettes, not counting a brand not on the list.

If a respondent reported use of a brand not on the list and answered "don't know" or "refused," or gave some response that was coded as bad data when asked to specify the name of this other brand, we assigned a final code to indicate that the respondent was a user of this type of tobacco but did not specify the brand name (e.g., a code of 4999 for "Cigar/Cigarillo, brand unspecified"). In addition, these residual "brand unspecified" codes were used for other ambiguous responses, such as those indicating that the respondents' usual brand varied or that they smoked or used whatever brand was least expensive. These residual brand-unspecified codes also were applied to a subgroup of cases at the end of the coding process for whom a specific, final brand code could not be arrived upon. For example, if a respondent entered a response to the cigarette brand question(s) that could not be coded (but was not necessarily "bad data"), we coded the respondent as a user of cigarettes, brand unspecified.

Under this coding approach, the first codes in the series were reserved for brands that were covered in the instrument. Specifically, for cigarettes, codes of 101 through 126 were used for Basic through Winston, corresponding to the brands listed in question CG11. Cigarette brand codes of 127 through 160 were used for Alpine through True, corresponding to the brands listed in question CG11a.[13] New specify codes for cigarettes that did not correspond to an existing brand listed in questions CG11 or CG11a started at 161. Similarly, for chewing tobacco, codes of 201 through 211 corresponded to Beech-Nut through Work Horse. For snuff, codes of 301 through 309 corresponded to Copenhagen through Timber Wolf. For cigars, codes of 401 through 428 corresponded to Antonio y Cleopatra through Winchester.

A particular advantage of this approach was that it would allow analysts to identify situations where respondents specified tobacco brands that pertained to another type of tobacco (e.g., specifying a snuff brand under chewing tobacco or little cigar brands under cigarettes). This coding scheme was particularly relevant for the chewing tobacco and snuff brand data, where snuff brands could be reported as the brand of chewing tobacco used most often in the past 30 days, or vice versa. Thus, if a respondent specified a brand of snuff in the Chewing Tobacco section, the final chewing tobacco brand variable was assigned a code in the 300 series (i.e., for snuff brands).

For coding "OTHER, Specify" responses, we developed a common computer-assisted coding procedure to assign appropriate codes for cigarettes, chewing tobacco, snuff, cigars, pipe tobacco, and other types of responses. In situations where respondents specified an answer that corresponded to a

[13] The numbering of codes corresponding to responses in question CG11a started with 127 instead of 128 because the code of 27 in question CG11, meaning "a brand not on this list," was simply a toggle to question CG11a. By resuming the coding of brands at 127 for brands listed in question CG11a, there was no break in the codes.

brand on a list, no distinction was made in the coding between situations where respondents chose that particular brand from the list and those where respondents reported use of a brand not on the list and then specified use of that brand.

In addition, the coding procedures did not make distinctions in terms of details, such as cigarette length (e.g., 100s), regular or menthol forms, or light, ultralight, or full-flavor varieties. For example, a respondent who usually smoked the menthol form of a cigarette brand listed in questions CG11 or CG11a may have considered the nonmenthol form to be the regular brand and the menthol form to be some other brand. However, the main intent of these questions was to determine *overall* use of *any* type of cigarette within a particular brand label. Thus, respondents who typically smoked the regular, nonmenthol version, a light version, or a menthol version would all be considered smokers of that basic brand. Further, the wording in question CG11 said, "the brand is the name on the pack." Consequently, we could not determine if respondents who keyed an existing brand from questions CG11 or CG11a might have keyed something else if they were given more detailed options for light, ultralight, full-flavor, menthol, or 100-mm versions.

Questions were added in 2000 to determine whether respondents smoked full-flavor, light, or ultralight cigarettes (CGTAR1 or CGTAR2), and whether they usually smoked regular or menthol cigarettes in the past 30 days (CGMENTH1 or CGMENTH2; see Section 4.4.3). We used data from the first two quarters of 2000 to evaluate whether the addition of these new questions required a change in our coding procedures for cigarette brands. Of concern was whether respondents might specify something in their cigarette brand response that was inconsistent with their answers to these new questions (e.g., specifying a light or ultralight brand but then reporting in CGTAR2 that they smoked full-flavor cigarettes most often in the past 30 days). When respondents made a distinction with respect to light, ultralight, full-flavor, or menthol cigarettes in the brand that they specified, they almost always answered CGTAR2 and CGMENTH2 in a manner consistent with what they specified. Therefore, the decision was made not to change the coding procedures for cigarette brands in 2000 to add more detail about the type of cigarette smoked within a given brand. This continued to be the practice for 2001.

Unlike the situation described above where respondents specified use of a brand not on the list, those who entered a brand from an available listing were asked to confirm their answer. If they confirmed their answer, they were asked no further questions about the brand they used for that particular type of tobacco. However, if respondents indicated that their previous answer was not correct, they were routed back through the series of 30-day brand questions for that type of tobacco. Thus, respondents had the opportunity to make corrections in situations where they may have miskeyed a number, such as if they keyed the number immediately above or below the number of the brand they meant to choose. For each type of tobacco that respondents reported using in the past 30 days, they were allowed to make corrections up to a total of three times. Respondents exited the loop once they confirmed an answer or

specified use of a brand not on the list. Respondents also exited the loop if they answered "don't know" or "refused" when asked to confirm their answer.

Because of this routing logic, we assigned the brand that respondents confirmed that they used most often in the past 30 days. If respondents were rerouted through the series of questions and confirmed their answer on their second or third pass through the questions, the final tobacco brand coding procedures retained the final answer that respondents confirmed and disregarded whatever previous answers the respondent had given but did not confirm. Respondents who answered "don't know" or "refused" when asked to confirm what brand they used were assigned that corresponding code to the final brand variable for that type of tobacco. We did not assign a brand-unspecified code in this situation because respondents who answered "don't know" or "refused" when asked to confirm what brand they used most often *had* reported use of a brand from the available listing, so they could not truly be thought of as not specifying what brand they used; rather, this group did not confirm their answer. In contrast, respondents who reported use of a brand not on the list but then did not know what specific brand they used or refused to specify a brand had provided some information to support the conclusion that they were users of some other, but unspecified, brand of that type of tobacco.

If respondents did not confirm what brand of a given tobacco type they used most often in the past 30 days despite three passes through the series of questions, we did not assign a final tobacco brand code within that series. In particular, we did not assign a brand-unspecified code. As in the situation when respondents answered "don't know" or "refused" when they were asked to confirm their answer, these respondents had given answers to indicate what brand they used most often in the past 30 days, but they never confirmed what they used. Current (i.e., past month) users of cigarettes, chewing tobacco, snuff, or cigars who did not confirm a brand despite three passes through the data were assigned a final code of 9000 to that tobacco type's brand. The code of 9000 was intended to indicate that these respondents did not confirm their brand. However, this edit was implemented only rarely in the 2001 CAI data (i.e., for two respondents in the cigarette brand data). In comparison, there were 18,842 respondents who were classified as past month cigarette users at the machine-editing stage. This pattern did not occur at all in the chewing tobacco, snuff, and cigar brand data in 2000.

As indicated previously, we also were aware that confusion sometimes existed in terms of what constituted chewing tobacco and snuff. Therefore, we also created a recoded "any smokeless tobacco brand" variable to indicate the brand of any smokeless tobacco that respondents reported using most often in the past 30 days. If respondents reported use of both smokeless tobacco types in the past 30 days, they were asked to indicate which they had used most often (i.e., the reported chewing tobacco brand or the reported snuff brand). We therefore assigned the final smokeless tobacco brand code according to the answer to this question. For example, if a respondent reported that the brand in the chewing tobacco section was the brand that he or she used most often, but the respondent specified use of a brand that was really a snuff brand, we still assigned that code to the smokeless tobacco brand. In this example, then, the

smokeless tobacco brand would indicate that the respondent used a particular snuff brand most often in the past 30 days, even though this response came from the Chewing Tobacco section.

### 4.4.3 Processing of Miscellaneous Cigarette Variables

Respondents who reported that they smoked cigarettes in the 30 days prior to the interview were asked a series of additional questions about their cigarette use in that period (i.e., other than the brand of cigarettes that they smoked most often). These questions covered the following topics:

- the average number of cigarettes they smoked per day,
- the type of cigarette they smoked (light, ultralight, or full flavor),
- whether they smoked regular or menthol cigarettes most often, and
- whether they smoked a "roll-your-own" cigarette.

Respondents who smoked cigarettes in the past 30 days were asked one of two possible questions regarding the number of cigarettes they smoked per day. Respondents who smoked on only 1 day in the past 30 days were asked to report the number of cigarettes they smoked on that 1 day. Respondents who smoked on more than 1 day (or who estimated the number of days they smoked in the past 30 days) were asked to report the average number of cigarettes they smoked per day. A single, composite variable (CIG30AV) was created from these two questions, using data from whatever question the respondents were asked. No further editing was done to the data from these two questions.

Similarly, past month cigarette users were routed to one of two possible questions about the type of cigarette they smoked in the past 30 days (CGTAR1 or CGTAR2). They also were routed to one of two possible questions regarding whether they usually smoked regular or menthol cigarettes (CGMENTH1 or CGMENTH2). Routing to these questions was mutually exclusive (i.e., respondents were routed to one or the other question in a set but not both). Therefore, composite variables were created for the cigarette type (CIG30TYP) and whether respondents smoked regular or menthol cigarettes most often (CIG30MEN).

An error was identified in the CAI program (corrected in June 2001) in which respondents who chose cigarette brand options 27 through 34 in question CG11a were skipped out of the CGTAR1 and CGMENTH1 questions. Respondents affected by this error prior to the fielding of the CAI correction in June 2001 retained values of 98 (blank) in CIG30TYP and CIG30MEN.

No attempt was made to edit CIG30TYP or CIG30MEN for consistency with the cigarette brand from the variable CIG30BRN. In developing these new items, the instrument development team consulted with tobacco research experts regarding which brands offered or did not offer menthol, light,

or ultralight varieties. No conclusive information was obtained. As was discussed in Section 4.4.2, the decision also was made in 2000 to continue the practice of not coding the cigarette brand "specify" data for particulars such as regular or menthol forms, or light, ultralight, or full-flavor varieties. Therefore, in situations in 2001 where respondents specified that level of detail regarding the brand of cigarette that they smoked most often in the past 30 days, that information was not used to edit CIG30TYP or CIG30MEN.

The cigarette section also included a question (CG14) about whether respondents smoked part or all of a "roll-your-own" cigarette in the past 30 days. The edited variable CIG30ROL corresponded to this question. The cigarette brand question CG11a[14] included response categories for two roll-your-own brands of cigarette tobacco. Respondents who chose either of these roll-your-own brands were skipped out of question CG14; by choosing a roll-your-own brand from the list of cigarette brands, these respondents had already indicated that they had smoked a roll-your-own cigarette in the past 30 days. Therefore, if question CG14 had been skipped and the cigarette brand was one of the roll-your-own brands from CG11a, we assigned a code of 5 (Yes LOGICALLY ASSIGNED [from skip pattern]) to the edited variable CIG30ROL.

However, respondents could specify a cigarette brand not on the list in questions CG11 and CG11a and then specify a roll-your-own brand in CG12. In this situation, respondents were routed to question CG14. If respondents specified a roll-your-own brand and CG14 was already answered as "yes," no further editing needed to be done. If CG14 was not answered as "yes" when respondents had specified a roll-your-own cigarette brand, we replaced the respondents' original answers with a code of 3 (Yes LOGICALLY ASSIGNED). This code of 3 signified to analysts that we had overwritten the respondents' original answers in the raw variable CG14 to make them consistent with the cigarette brand information recorded in the variable CIG30BRN. This edit occurred very rarely in 2001 (for only 1 respondent out of 18,829 who had nonmissing values in the edited variable CIG30ROL).

[14]Respondents could be routed to questions RCG11, RRCG11, RCG11a, RRCG11a, RCG12, or RRCG12 if they cycled through the cigarette brand questions more than once. For brevity, however, we limit our reference here to the first set of cigarette brand questions: CG11, CG11a, or CG12.