

2011 NATIONAL SURVEY ON DRUG USE AND HEALTH

PROCEDURES FOR EDITING INTERVIEWER- ADMINISTERED DATA IN THE 2011 NSDUH COMPUTER-ASSISTED INTERVIEW

Prepared for the 2011 Methodological Resource Book

Contract No. HHSS283200800004C
RTI Project No. 0211838.207.003
Deliverable No. 39

Authors:

Larry A. Kroutil
Chuchun Chien
Wafa Handley
Michael R. Bradshaw

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

January 2013

2011 NATIONAL SURVEY ON DRUG USE AND HEALTH

PROCEDURES FOR EDITING INTERVIEWER- ADMINISTERED DATA IN THE 2011 NSDUH COMPUTER-ASSISTED INTERVIEW

Prepared for the 2011 Methodological Resource Book

Contract No. HHSS283200800004C
RTI Project No. 0211838.207.003
Deliverable No. 39

Authors:

Larry A. Kroutil
Chuchun Chien
Wafa Handley
Michael R. Bradshaw

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

January 2013

Acknowledgments

This report was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ, formerly the Office of Applied Studies), by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina, under Contract No. HHSS283200800004C. At RTI, Larry A. Kroutil, Chuchun Chien, Wafa Handley, and Michael R. Bradshaw co-authored this report. Thomas G. Virag served as the RTI Project Director. Richard S. Straw copyedited the document, and Debbie Bond word processed it.

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

Table of Contents

Section	Page
List of Exhibits.....	iv
1. Introduction.....	1
2. General Edit Issues for the Interviewer-Administered Data.....	5
2.1 Implementation of General Legitimate Skip Fills	5
2.2 Handling of Missing Data.....	6
2.3 Handling of Responses to "OTHER, Specify" Variables	7
2.4 Handling of Interview Errors	7
3. Edit Issues for Specific Interviewer-Administered Sections	9
3.1 Core Demographics Variables	9
3.2 Noncore Demographics Variables	10
3.2.1 Moves in the Past Year and State Residency	10
3.2.2 Immigrant Status.....	13
3.2.3 Noncore Education.....	13
3.2.4 Employment and Workplace.....	23
3.2.5 Health Insurance	29
3.2.6 Edited Income	32
3.2.7 Field Interviewer Debriefing Questions.....	36

List of Exhibits

Exhibit	Page
1. Location of Documentation on Editing in the 2011 Methodological Resource Book.....	4
2. Edit Issues Pertaining to Moves in the Past Year and State Residency	12
3. Edit Issues Pertaining to the Noncore Education Section.....	16
4. Edit Issues Pertaining to the Employment and Workplace Section.....	27
5. Edit Issues Pertaining to the Health Insurance Section	33

1. Introduction

This report is the third in a series that documents procedures developed for editing the computer-assisted interviewing (CAI) data from the 2011 National Survey on Drug Use and Health (NSDUH). The first report, *General Principles and Procedures for Editing Drug Use Data in the 2011 NSDUH Computer-Assisted Interview*,¹ serves as the starting point for background on basic CAI editing issues and procedures. As such, it provides background on issues surrounding the transition from data collection based on paper-and-pencil interviewing (PAPI) to a CAI format. The first document in the series also discusses the following topics:

- general principles associated with editing the CAI data, including the assignment and meaning of standard NSDUH codes and principles for assigning relevant "not applicable" types of codes;
- initial processing steps, including (a) general procedures for coding of "OTHER, Specify" data, (b) creation of edit-ready raw variables, (c) initial processing of age-related variables, (d) identification of usable cases, (e) investigation of potentially problematic response patterns, and (f) edits of date-dependent variables when the interview date was judged to be questionable; and
- edits involving the key self-administered drug use variables in the cigarettes through sedatives sections, including edits of (a) the lead lifetime use variables (i.e., gate questions), where respondents indicated whether they had ever used the drug of interest; (b) the recency-of-use variables, where respondents who indicated lifetime use of the drug indicated when they last used that drug; (c) 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 (d) remaining variables in a module.

The second document in the series discusses procedures for editing supplementary modules that were self-administered by the respondents.² 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 a computer screen and enter their responses directly into a 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 practice prevented interviewers (or others in the household)

¹ Kroutil, L. A., Handley, W., & Bradshaw, M. R. (2013). General principles and procedures for editing drug use data in the 2011 NSDUH computer-assisted interview. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 10, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.207.003). Research Triangle Park, NC: RTI International.

² Kroutil, L. A., Handley, W., Bradshaw, M. R., Chien, C., & Felts, B. J. (2013). Procedures for editing supplementary self-administered data in the 2011 NSDUH computer-assisted interview. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 10, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.207.003). Research Triangle Park, NC: RTI International.

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 in which interviewers read the questions and respondents gave their answers aloud to the interviewers, who then entered the responses into the computer. The logic for determining which questions should be asked was controlled by the computer program based on the responses entered by the interviewers. Consequently, interviewers could concentrate on asking questions and recording respondent answers, without having to concern themselves with comprehending and following skip pattern instructions.

This third document describes procedures for editing these interviewer-administered sections of the survey. 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 one year to the next to permit measurement of trends in drug use, including trends among key demographic subgroups. In contrast, the content of noncore sections could undergo more change 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 one year to the next.

Section 2 of this report discusses general issues associated with editing the interviewer-administered data. Section 3 discusses specific issues associated with the editing of individual interviewer-administered modules, where applicable.

As was the case with the NSDUH instrument as a whole, the interviewer-administered sections were divided into core and noncore demographics sections. The core demographics section consisted of key data on respondent age, gender, Hispanic origin, race, marital status, number of times married, military service history, highest educational grade attained, and perceived health. The noncore demographics section contained the following sections:

- moves in the past year and State residency,³
- immigrant status,
- noncore education (i.e., education-related questions other than the highest grade attained),
- employment and workplace,
- household roster information,
- proxy information (for determining who from the household should answer health insurance and income questions),

³ The item that the field interviewer (FI) completed for the State location of the sampled dwelling unit (SDU) was actually placed toward the beginning of the interview (question FIPE4). Because FIPE4 was used in editing the data on State residency, editing of State location data is discussed in conjunction with the editing of those variables. In addition, data from FIPE4 were relevant to the editing of variables in the health insurance and income sections.

- health insurance,
- income (including questions about telephone numbers serving the household), and
- FI debriefing questions (completed by the FI after the conclusion of the interview).

This document discusses procedures for logically editing data from these core and noncore interviewer-administered sections, except for variables pertaining to age, gender, Hispanic origin, race, the household roster information, and proxy information. For these latter variables, both editing (where applicable) and/or preparation of final, statistically imputed variables were handled as part of the statistical imputation procedures.

Logical editing uses data from elsewhere within the same respondent's record to reduce the occurrence of missing or ambiguous data or to resolve inconsistencies between related variables.⁴ Documentation for most procedures for logically editing data in the 2011 NSDUH is included in this editing and coding section (Section 10) of the 2011 methodological resource book (MRB). The imputation report (Section 11 of the 2011 MRB) contains descriptions of additional editing procedures for variables that subsequently undergo statistical imputation.⁵ For some variables, these procedures may involve assumptions (e.g., picking the midpoint from a range) or other additional edits to prepare the variables for imputation. [Exhibit 1](#) provides a crosswalk to identify the location of relevant documentation on editing procedures in the 2011 MRB.

⁴ There are a few situations where data from outside the respondent's record is used in logical editing. Some editing procedures involve data from the screener, where an eligible member of the dwelling unit reports basic information about all members of the dwelling unit. In situations where two members of the same dwelling unit are selected for the survey and complete the interview, data from the second respondent's record may be used in logical editing of some variables in the first respondent's record, or vice versa. This procedure allows use of information from both respondents to determine relationships among household members.

⁵ Frechtel, P. A., Archambault, H. J., Carpenter, L. A., Edwards, S. L., Laufenberg, J. S., Martin, P. C., Moore, A. S., & Scott, V. M. (2013). Imputation report. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for Substance Abuse and Mental Health Services Administration under Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.207.006.007). Research Triangle Park, NC: RTI International.

Exhibit 1. Location of Documentation on Editing in the 2011 Methodological Resource Book

Variable Group	Location within MRB Section 10 (Editing and Coding)¹	Location within MRB Section 11 (Imputation) Where Logical Edits/ Additional Edits Are Discussed
Core Demographics	Report 3: Interviewer-Administered Data, Section 3.1	Sections 3.2 and 4.2
Drugs	Report 1: General Principles/Core Drug Data Report 2: Supplementary Self-Administered Data, Section 3.1.1	Section 5.2
Nicotine Dependence	Report 2: Supplementary Self-Administered Data, Section 3.4	(No additional edits discussed)
Immigrant Status	Report 3: Interviewer-Administered Data, Section 3.2.2	Sections 4.2.2 and 4.2.3
Employment	Report 3: Interviewer-Administered Data, Section 3.2.4	Section 4.2.1
Roster	None	Sections 7.2 to 7.4
Health Insurance	Report 3: Interviewer-Administered Data, Section 3.2.5	Section 9.2
Income	Report 3: Interviewer-Administered Data, Section 3.2.6	(No additional edits discussed)
Roster Pair	None	Sections 10.2 to 10.4

MRB = methodological resource book.

¹Section 10 contains three reports. The first report describes general editing principles and procedures for editing the core substance use data. The second report describes procedures for editing supplementary self-administered interview data. The third report describes procedures for editing interviewer-administered data.

2. General Edit Issues for the Interviewer-Administered Data

The following general issues were relevant to the editing of the interviewer-administered data:

- implementation of general legitimate skip fills,
- handling of missing data, and
- handling of responses to "OTHER, Specify" variables.

2.1 Implementation of General Legitimate Skip Fills

An important aspect of editing the interviewer-administered data involved identifying variables that had been legitimately skipped by the computer program based on respondent characteristics (e.g., age, gender) or other answers that respondents gave to prior questions. For example, respondents under the age of 15 were not asked questions about their current marital status or the number of times that they had been married. In addition, if respondents aged 15 or older reported in question QD07 that they had never been married, there was no need for them to be asked the question about the number of times they had been married.

The following general code was 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 or based on some other criteria (e.g., age of the respondent):

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

In the above example, if a respondent was younger than 15 and the marital status questions had been skipped, codes of 99 were assigned in the editing process to the variables pertaining to marital status and the number of times married. Similarly, if a respondent had never been married and the item had been skipped pertaining to the number of times the respondent had been married, a code of 99 was assigned to the edited variable NOMARR (i.e., number of times married).

The following analogous code also was assigned through the editing procedures:

89 (or 989, or 9989, etc.) = LEGITIMATE SKIP Logically assigned.

The value of 89 signified that existing values were overwritten during data editing. For example, if a respondent was somehow routed into the marital status questions but that respondent subsequently was classified as being younger than 15, any answers that the respondent gave to these items were overwritten with codes of 89. These codes signified that the youth logically was not eligible to be asked these questions.

As in the general procedures described in the first part of the documentation of data editing (see footnote 1 in Section 1), edits in these types of situations required the ability to determine *unambiguously* that a question did not apply. For example, if respondents did not know their current marital status or refused to report it, the computer-assisted interviewing (CAI) skip logic treated these responses as though the respondents had never been married. From the standpoint of respondent burden, there often may be little value in asking further questions about a particular topic if respondents could not indicate unambiguously whether the topic was relevant at all. In addition, asking respondents in this situation about the number of times they had been married would imply that they had been married at least once.

On the other hand, responses of "don't know" or "refused" 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. Consequently, such responses could be thought of as *potentially* affirmative responses, as opposed to inferring that they are negative responses. For this reason, when respondents answered a lead question as "don't know" or "refused," missing values were retained for the questions that the CAI program skipped (see Section 2.2).

2.2 Handling of Missing Data

The occurrence of missing data was not completely eliminated in CAI because respondents had the option of answering "don't know" or "refused" to questions when asked for a response. In addition, questions often were skipped if respondents answered a lead question as "don't know" or "refused," as noted above.

In situations where respondents answered "don't know" or "refused" to a lead question, the following standard codes for missing data generally were applied:

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]).

When a lead question retained a code of 97 after other editing had been done, refusal codes were assigned to the skipped questions within that branch (i.e., the refusal was propagated). That is, it was logically inferred that a refusal to the lead question was a blanket refusal to answer any questions on that topic. When a lead question retained a code of 94 after other editing had been done, values of blank were retained in the questions that had been skipped.

The following additional missing data code could be assigned to interviewer-administered variables:

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

"Bad data" codes usually were assigned when responses were inconsistent with other data.

2.3 Handling of Responses to "OTHER, Specify" Variables

There were two types of "OTHER, Specify" questions in the interviewer-administered sections:

- those where respondents did not get the opportunity to choose the "other" response (and specify something) if they already chose another category from the list, and
- those where the "OTHER, Specify" item was a follow-up to a lead question that typically was answered as "yes" or "no"; depending on the nature of the lead question, either an affirmative or a negative response to the lead question could govern whether respondents were asked to specify something.

Question QD24SP (specify other reason for leaving school without getting a high school diploma) is an example of the first type of "OTHER, Specify" question. Respondents were first asked question QD24, "Please look at this card and tell me which one of these reasons best describes why you left school before receiving a high school diploma." If respondents chose a response from the list of options in QD24 except for "other reason," they were not routed to QD24SP. For this type of "OTHER, Specify" question, data from the lead question (e.g., QD24) and the specify question (e.g., QD24SP) were combined into a single, final variable (LFSCHWHY). "OTHER, Specify" responses that corresponded to existing response categories were coded starting with number 21, with the coding proceeding in the order of the existing response categories. For example, if a respondent did not choose category 7 from QD24 ("I had to get a job [or work more hours]") but specified a response that corresponded to that category, a code of 27 was assigned to the coded response. The final, edited variable LFSCHWHY would have a code of 27 to signify that (a) the respondent left school because he or she needed to get a job (or work more hours), and (b) the respondent specified this as some other reason for leaving school, as opposed to choosing category 7 directly. When respondents chose the other category in the lead question but specified something that got coded as a missing value (i.e., don't know, refused, bad data, blank), the final variable retained a code corresponding to other (as opposed to assigning a missing value).

Question QD15 in the noncore demographics section (other country of birth) is an example of the second type of "OTHER, Specify" question. Only those respondents who reported in question QD14 that they were not born in the United States (QD14 = 2) were routed to QD15 and asked to report the other country where they were born. Conversely, respondents who reported that they were born in the United States (QD14 = 1) were skipped out of QD15, and the edited variable BORNINOT (corresponding to QD15) was assigned a legitimate skip code.

2.4 Handling of Interview Errors

For a small number of cases, staff who monitored the data collected by the field interviewers (FIs) alerted the editing team to data collection protocol violations that FIs made in the interviewer-administered data. When such cases were included in the dataset as final respondents (see footnote 1), the interviewer-administered variables that were affected by the protocol violation were assigned codes for bad data.

3. Edit Issues for Specific Interviewer-Administered Sections

As discussed previously, the interviewer-administered sections were divided into core and noncore demographics sections. Processing of core demographics variables is discussed first, followed by discussion of specific issues pertaining to variables in the noncore demographics sections.

3.1 Core Demographics Variables

Core demographics variables that were handled by the editing task included marital status, number of times married (if respondents had ever been married), U.S. military service history, current military status (if respondents had ever been in the U.S. military), highest educational grade attained, and perceived health. Consistent with the editing procedures in prior years, minimal processing of these variables was done beyond that of assigning legitimate skip codes, as described in Section 2.1.

Processing of the variables pertaining to military service is discussed here in detail, however, because respondents who were currently on active duty in the U.S. military were not eligible to be interviewed in the National Survey on Drug Use and Health (NSDUH). Legitimate skip codes were assigned to the variables pertaining to lifetime U.S. military service and current military status if respondents were younger than 17. In addition, legitimate skip codes were assigned to the current military status variable if respondents were aged 17 or older and reported that they had never been in the U.S. armed forces.

Respondents who reported that they had been in the U.S. armed forces then were asked whether they were (a) still on active duty, (b) in a military reserves component, or (c) separated or retired from active duty or the reserves. Unlike the situation in most places in the interview, responses of "don't know" or "refused" to the question about lifetime military service were treated as potentially having served in the military. Thus, these respondents also were asked about their current military status.

If respondents reported that they were currently on active military duty, the interviewers were asked to confirm this answer with the respondents. The interview was terminated if respondents confirmed that they were on active duty in the U.S. military. Consequently, there were no final respondents in the final NSDUH data who reported that they currently were on active military duty. However, some final respondents were civilians who were currently in the military reserves or were separated or retired from the military. In addition, the industry and occupation variables in the noncore employment section may include military-related codes for some respondents (see Section 3.2.4).

Another noteworthy aspect of the processing of the core demographics variables was that the core education variable EDUC (highest grade completed) was not edited with respect to education variables in the noncore demographics section (e.g., current grade), nor was it edited with respect to the respondent's age. However, a second variable, EDTEDUC, was created as

part of the noncore demographics processing (see below). Consequently, the core education variable would not be affected by changes that might occur in the content of noncore education variables in subsequent years. Nevertheless, the EDTEDUC variable might in some situations be a more accurate reflection of the highest grade that respondents had completed.

3.2 Noncore Demographics Variables

As noted previously, the following noncore demographics sections were handled as part of the data editing process:

- moves in the past year and State residency,
- immigrant status,
- noncore education (i.e., education-related questions other than the highest grade attained),
- employment and workplace,
- household roster information,
- health insurance,
- income (including questions about telephone numbers serving the household), and
- field interviewer (FI) debriefing questions (completed by the FI after the conclusion of the interview).

Questions in the income section pertaining to the number of telephone numbers serving the household (TELNO) also were handled through the data editing code. However, processing of TELNO was limited to assigning a final, mnemonic variable name.

3.2.1 Moves in the Past Year and State Residency

This section covered issues related to changes of residence. Specifically, question QD13 (edited variable MOVESPYR) asked respondents to report the number of times that they had moved in the past 12 months. If a respondent moved at least once in the past 12 months, the respondent was asked to report the State where he or she was living a year prior to his or her interview date (question QD13a; edited variable LIVSTYRA). If a respondent reported in the social environment module (for adults aged 18 older) or the youth experiences module (for youths aged 12 to 17) that he or she moved at least once in the past 5 years, the respondent was asked to report the State where he or she was living 2 years prior to the interview date (question QD13b; edited variable LIVST2YA). Respondents who moved from another State to their current one (from the checkpoint⁶ variable FIPE4; edited variable STATELOC) within the past 2 years were asked to report the month and year when they moved to their current State (question QD13c; edited variables MOVSTTMO and MOVSTTYR).

⁶ "Checkpoint" refers to an item completed by the interviewer about the location of the sampled dwelling unit (SDU) or characteristics of the sample within the SDU; these checkpoints are not seen by the respondent and are used to determine what the respondent is asked in subsequent sections of the interview.

Interviewers were instructed to use the FI checkpoint FIPE4 at the beginning of the interview to report the State where the SDU was located. The State that interviewers entered in FIPE4 (edited variable STATELOC) sometimes mismatched the State residence information that was used to sample a given case. These mismatches were investigated by field staff during data collection. Some of these mismatches existed for a valid reason, such as if a respondent had been selected in an SDU in one State but had moved to another State. In these situations, if FIPE4 reflected the State where the respondent was currently living, STATELOC retained the value from FIPE4. Otherwise, if the State information in FIPE4 was entered incorrectly, STATELOC was set to bad data. In turn, the setting of STATELOC to bad data affected the editing of other State residency variables.

As noted above, information from question QD13c was captured in two edited variables pertaining to the month (MOVSTTMO) and year (MOVSTTYR) when respondents moved to their current State. Assignment of values to these two variables took into account the logic that the computer-assisted interviewing (CAI) program used to assign specific months and years to the response categories in QD13c, which depended on the interview date and the answers in questions QD13a and QD13b. Specifically, if QD13a had been answered (i.e., for the State where the respondent was living a year ago) and the State where the respondent lived a year ago did not equal the respondent's current State residence (from FIPE4), then the respondent saw the response categories that were filled in QD13c as follows: the first response option (QD13c = 1) was filled with the month and year that occurred 12 months ago (i.e., same month as the interview month but in the year 2009), the next response option was filled with the month that occurred 11 months ago, and so on, up through the last response option (QD13c = 13), which was the current month in which the respondent was interviewed.

Suppose, for example (not necessarily actual data), that a respondent was interviewed in May 2011, QD13a was answered, and QD13a did not equal the State in FIPE4. The first response option in QD13c that the respondent saw would have been "May 2010," corresponding to the same month as the interview month but in the prior calendar year. If a value of 4 was keyed in QD13c, that response corresponded to "August 2010" based on this interview month. In this situation, then, the edited variable MOVSTTYR was coded as 2010 and MOVSTTMO was coded as 8.

Otherwise, the month that the CAI logic filled in the response options for QD13c still began with the interview month. However, the year that was filled in the response options began with 2009. Response option 13 in QD13c was filled with the interview month in 2010. Suppose, for example, that QD13a was blank because the respondent did not move in the past 12 months (QD13 = 0), but the State that the respondent lived in 2 years ago (from QD13b) did not equal the State where the SDU was located. If the interview was conducted in April 2011, then the first response option in QD13c would be filled with April 2009. If a value of 2 was keyed in QD13c, that response would correspond to the respondent moving to the current State in May 2009. In this example, MOVSTTYR was coded as 2009 and MOVSTTMO was coded as 5.

A key aspect of processing the variables LIVSTYRA, LIVST2YA, MOVSTTMO, and MOVSTTYR involved assignment of legitimate skip codes based on the CAI logic. This assignment of legitimate skip codes is described below.

- If respondents did not move in the past 12 months (MOVESPYR = 0), then LIVSTYRA was assigned a legitimate skip code.
- If respondents did not move in the past 5 years (i.e., SNMOV5YR = 0 from the social environment module for respondents aged 18 or older or YEMOV5YR = 0 from the youth experiences module for respondents aged 12 to 17), then LIVST2YA was assigned a legitimate skip code.
- If respondents moved zero times or moved within the same State, then MOVSTTYR and MOVSTTMO were assigned legitimate skip codes. This condition would not hold if respondents reported in either QD13a or QD13b that they moved to their current State from outside of the United States.

Exhibit 2 discusses additional edit issues that were relevant to LIVSTYRA, LIVST2YA, MOVSTTMO, and MOVSTTYR.

Exhibit 2. Edit Issues Pertaining to Moves in the Past Year and State Residency

Issue	Edits Implemented
Question QD13 was answered as "don't know" (DK) or "refused" (REF). However, SNMOV5YR or YEMOV5YR indicated that the respondent (R) did not have any moves in the past 5 years.	It could be logically inferred that the R did not have any moves in the past 12 months. A value of 0 was logically assigned to MOVESPYR (corresponding to QD13). An "editing indicator" variable EIMOV5YR was created to indicate when this logical inference was made. The default value for EIMOV5YR was 1, when no editing was done to MOVESPYR. EIMOV5YR was assigned a value of 2 when this edit was implemented.
The number of moves in the past 12 months (from MOVESPYR) was greater than the number of moves in the past 5 years (from SNMOV5YR or YEMOV5YR).	No editing was done to MOVESPYR, SNMOV5YR, or YEMOV5YR. In particular, interviewers would not have access to information from the self-administered variables SNMOV5YR and YEMOV5YR to be aware of this inconsistency. However, a "flag" variable (MOVYRFLG) was created that indicated when this inconsistency occurred. The default value was 98 (i.e., blank) when MOVESPYR was consistent with either SNMOV5YR (for Rs aged 18 or older) or YEMOV5YR (for Rs aged 12 to 17). MOVYRFLG had a value of 1 when MOVESPYR was inconsistent with the SNMOV5YR or YEMOV5YR.
Question QD13b (edited variable LIVST2YA) was skipped because Rs reported no moves in the past 5 years in SNMOV5YR or YEMOV5YR. However, MOVESPYR indicated one or more moves in the past year.	As in the above issue, Rs logically should have reported that they had no moves in the past year if they had reported no moves in the past 5 years. Therefore, LIVST2YA retained a code of 98 (i.e., blank) rather than being assigned a legitimate skip code, even though SNMOV5YR or YEMOV5YR indicated no moves in the past 5 years.
STATELOC had been set to bad data because the information that an interviewer keyed for the State where the sampled dwelling unit (SDU) was located was inconsistent with the State that was on record for that case.	MOVSTTYR and MOVSTTMO (from QD13c) also were set to bad data. When STATELOC was set to bad data, this called into question whether Rs should have been asked to report the month and year when they moved to their current State.

3.2.2 Immigrant Status

Edits described in this section for the immigrant status variables have applied since 2004, when the content of the immigrant status variables changed. Question QD14 asked whether respondents were born in the United States. If they were not born in the United States, question QD15 asked respondents for their country of birth. The former question QD16 (length of time that respondents had lived in the United States, corresponding to edited variable LIVEDUSA) was replaced in 2004 with three variables pertaining to (a) whether respondents had lived in the United States for at least 1 year (question QD16a or edited variable LIVUS1YR); (b) the number of years that respondents had lived in the United States, if they reported in QD16a that they had lived in the United States for at least 1 year (question QD16b or edited variable LIVUSYRS); and (c) the number of months that respondents lived in the United States, if they answered QD16a as "no," indicating that they had not lived in the United States for at least 1 year (question QD16c or edited variable LIVUSMOS).

An important aspect of processing the immigrant status variables involved assigning legitimate skip codes where relevant. For example, if respondents reported that they were born in the United States (i.e., the edited variable BORNINUS was answered as "yes"), the edited variables BORNINOT, LIVUS1YR, LIVUSYRS, and LIVUSMOS were assigned legitimate skip codes. Similarly, if LIVUS1YR = 1 (i.e., "yes"), LIVUSMOS was assigned a legitimate skip code. If LIVUS1YR = 2 (i.e., "no"), LIVUSYRS was assigned a legitimate skip code. When LIVUS1YR was coded as 94 ("don't know") or 97 ("refused"), the appropriate code for "don't know" or "refused" was assigned to the variables LIVUSYRS and LIVUSMOS that had been skipped.

If respondents reported that they were born outside the United States, however, it was possible for them to specify an answer in question QD15 that logically would mean that they were born in the United States. If this inconsistency occurred in the data (i.e., it had not been resolved by the interviewer), then the edited variable BORNINUS was logically inferred to be answered as "yes."⁷ The edit procedures also logically inferred that the edited variables BORNINOT, LIVUS1YR, LIVUSYRS, and LIVUSMOS should have been skipped.

It also was possible for respondents to report in question QD16b that they had lived in the United States for a number of years greater than their current age. When this situation occurred, the edited variable LIVUSYRS was assigned a bad data code to indicate that the answer was inconsistent with the respondent's age. No editing was done to LIVUSYRS when LIVUSYRS was equal to the respondent's age because that answer in LIVUSYRS could be interpreted to mean that the respondent came to the United States as a baby.

3.2.3 Noncore Education

Question QD17 asked whether respondents were currently enrolled in school. Since 2001, respondents who did not report in question QD17 that they were currently enrolled in school have been asked follow-up questions (if they were aged 12 to 25 and their highest reported grade from question QD11 was grade 1 to 15) to determine if they were on a holiday or

⁷ If respondents reported being born in Alaska or Hawaii and were born before 1959 (i.e., before Alaska and Hawaii became States), these respondents still were considered to have been born in the United States.

vacation break from school (question QD17a), and if so, whether they intended to return to school once their break was over (question QD17b). Because of the addition of these new follow-up questions, the name of the school enrollment variable has been called SCHENRL since 2001. Prior to 2001, this variable was called ENROLED.

If respondents originally reported in QD17 that they were not enrolled (QD17 = 2) but reported in QD17b that they intended to return to school once their vacation or break was over (QD17b = 1), SCHENRL was set to a value of 1 ("yes") to indicate that the respondents should be considered enrolled. Otherwise, the response from QD17 was carried over to SCHENRL. That included situations in which respondents reported in QD17a that they were not on vacation break from school, or who reported in QD17b that they were on break but did not intend to return to school once their break was over.

Respondents who reported that they were enrolled were asked to report their current grade in school (or the grade they would be in once they returned from school break), whether they were a full- or part-time student, and the number of days that they missed school in the past 30 days because they were sick or because they skipped school (questions QD18 through QD21). For question QD18, respondents who reported in QD17 that they were currently enrolled in school (QD17 = 1) were asked to report the grade of school they were currently attending. For respondents who were on vacation break from school but intended to return to school once their break was over (QD17b = 1), question QD18 asked for the grade that they would be in once they returned from their vacation break.

Prior to 2001, QD18 asked respondents only for their current grade. Because question QD18 was worded differently for different groups of respondents, the name of the corresponding variable has been called EDUCATND since 2001. Prior to 2001, this variable was called EDUCNOW.

Similarly, since 2002, the wording of the question about full-time or part-time student status (question QD19) has varied for respondents who were currently attending school and those who were on break from school but intended to return to school. Respondents who were currently attending school were asked, "Are you a full-time student or a part-time student?" Those who were on break from school but intended to return to school were asked, "Will you be a full-time or a part-time student?" Therefore, the name of the corresponding variable has been called SDNTFTPT since 2002. Prior to 2002, this variable was called STUDNT.

Respondents who were aged 25 or younger, had completed the 12th grade or lower (from question QD11), and were not enrolled in school (see above) were asked whether they had received a high school diploma (question QD22). Respondents in this age group who reported that they left school without receiving a high school diploma were asked whether they had received a GED certificate of high school completion, why they left school before receiving a high school diploma, and their age when they left school (questions QD23 through QD25).

Thus, if respondents were currently enrolled in school, the edited variables corresponding to questions QD22 through QD25 (HSDIPLMA, HSGED, LFSCHWHY, and LFTSCHAG) were assigned legitimate skip codes. Similarly, respondents aged 26 or older were considered to have legitimately skipped out of questions QD22 through QD25 because of the age requirement for

administration of these questions, regardless of whether they might not have finished high school. In addition, if respondents were not currently enrolled in school, the edited variables corresponding to questions QD18 through QD21 (EDUCATND, STUDNT, SCHDSICK, and SCHDSKIP) were assigned legitimate skip codes, provided there were no other data to suggest that they were enrolled (see below).

[Exhibit 3](#) discusses additional edit issues that were relevant to the noncore education variables. In particular, the current school grade question QD18 could be inconsistent with the highest grade that the respondent reported completing in question QD11. In most situations, one might expect the current grade in QD18 to be one grade level higher than the response in QD11. In addition, no editing was done when the current grade reported in QD18 was the same as the highest grade reported in QD11 because respondents could have been repeating a grade.

Since 2002, a "hard error" has been included in the education section for situations in which the highest grade from QD11 was higher than the current (or anticipated) grade from QD18. (In 2001, a hard error was triggered if QD11 and QD18 differed by 2 or more years in either direction.) FIs were prompted to verify the answers with the respondents and correct any information in QD11 or QD18. If the answers were correct as recorded, the FIs could "suppress" the hard error and continue with the interview. When FIs suppressed a hard error message, however, they were requested to enter a comment documenting why the information that had been entered in QD11 and QD18 was correct. These comments were reviewed on a case-by-case basis to determine if (a) the answers should be accepted and no editing should be done to EDTEDUC (corresponding to QD11) or EDUCATND (corresponding to QD18); (b) the value for EDTEDUC or EDUCATND should be edited for consistency with the comments entered by the FI; (c) EDTEDUC or EDUCATND should be set to bad data based on the FI comments; or (d) normal education edits should be invoked (see below and [Exhibit 3](#)). Any edits based on the FI comments were done on a case-level basis using the respondent ID, not on a more global basis. These case-level edits superseded any of the usual edits discussed in [Exhibit 3](#) that otherwise would have been done. However, this hard error was suppressed for fewer than 30 cases, and specific case-level edits were done for fewer than 15 of these cases.

The general education edits discussed below that had been in place since 1999 were invoked if the hard error between QD11 and Q18 had been triggered, the answers from QD11 and QD18 had not been corrected, or the FI's comments indicated that a correction needed to be made, but what needed to be corrected was not clear from the FI's comments. Similarly, answers to QD11 and QD18 were accepted when FIs provided a plausible reason for the discrepancy between the two answers, such as if respondents were in college and transferred to another school but some prior credits did not transfer.

In addition, if the FI's comments indicated that the respondent was now in some sort of technical or vocational school, the school enrollment variable SCHENRL was set to a value of 4 (No LOGICALLY ASSIGNED). This edit was done because interviewers were instructed not to include vocational or technical schools as types of schools in which respondents could be enrolled. When SCHENRL was set to a value of 4, any data in EDUCATND, STUDNT, SCHDSICK, and SCHDSKIP were overwritten with values of 89 (LEGITIMATE SKIP Logically assigned). Where possible, when respondents were inferred not to be enrolled in school because their current enrollment was in a technical or vocational school, FI comments

Exhibit 3. Edit Issues Pertaining to the Noncore Education Section

Issue	Edits Implemented
<p>The current grade (QD18) was potentially inconsistent with the highest grade that the respondent (R) reported completing (QD11), and (a) the hard error between QD11 and QD18 was not triggered (e.g., the current grade from QD18 was two or more grades higher than the highest grade from QD11); or (b) the hard error was triggered and suppressed, but the field interviewer (FI) did not provide sufficient information to determine what corrections needed to be made.</p>	<p>An algorithm was developed that compared the self-reported current and highest grades with the R's current age (see text). A noncore edited variable for the highest grade completed (EDTEDUC) also was created. Edits were generally implemented as follows:</p> <ul style="list-style-type: none"> • When both the current grade and the highest completed grade were potentially consistent with the R's age, the edits picked the response from QD18 or QD11 that would yield the most consistent data. The second variable in the pair then was edited for consistency with the response that was picked as being most consistent. • When the current grade was more consistent with the R's current age than was the reported highest grade from the core demographics, then EDTEDUC was logically assigned a code to indicate that the R had completed the lower grade that was adjacent to his or her current grade. • When the highest completed grade was more consistent with the R's current age than was the reported current grade, then the edited current grade (EDUCATND) was logically assigned a code to indicate that the R was in the next highest grade relative to the one he or she had completed, or else EDUCATND was coded as bad data. • When neither the current grade (QD18) nor the highest completed grade (QD11) was consistent with the R's age, either EDTEDUC or EDUCATND (or both) were coded as bad data. If the current grade was exactly two grades higher than the last grade but the highest grade was lower than the expected highest grade, then EDTEDUC was coded as bad data. If the current grade was more than two grades higher than the last grade but the current grade was lower than the expected current grade, then EDUCATND was coded as bad data. If the current grade was lower than the highest grade, the one that was closest to the expected grade was chosen, and the other was set to bad data. If both EDTEDUC and EDUCATND were both close to their expected grades, both were set to bad data.
<p>The R reported being currently enrolled in school, and a hard error was triggered between QD11 and QD18. The FI's comments for suppressing the hard error indicated that the R was currently enrolled in technical or vocational school.</p>	<p>The R was logically inferred not to be currently enrolled in school. A special code of 4 was assigned to the edited school enrollment variable SCHENRL. For the following variables, it was logically inferred that they should have skipped: EDUCATND (current grade), SDNTFTPT (full- or part-time status), SCHDSICK (number of days in the past 30 days that the R missed school because the R was sick), and SCHDSKIP (number of days the R skipped school in the past 30 days). Consequently, any data in these items were wiped out in the edited variables.</p>
<p>The R did not know or refused to report whether he or she was enrolled in school, reported being on a holiday or break from school (QD17a = 1), but reported that he or she did not intend to return to school once the break was over (QD17b = 2).</p>	<p>The R was defined as not being enrolled in school (SCHENRL=2). The variables pertaining to the current grade through the number of days that the R skipped school in the past 30 days (EDUCATND, SDNTFTPT, SCHDSICK, and SCHDSKIP) were assigned legitimate skip codes.</p>

(continued)

Exhibit 3. Edit Issues Pertaining to the Noncore Education Section (continued)

Issue	Edits Implemented
<p>The R reported not being currently enrolled in school. In the question about reasons for leaving school without getting a high school diploma (LFSCHWHY), however, the R specified that he or she was still in school (LFSCHWHY = 778).</p>	<p>The R was logically inferred to be currently enrolled in school. A special code of 3 was assigned to the edited school enrollment variable SCHENRL. For the following variables, it was logically inferred that they should have skipped: HSDIPLMA (receipt of a high school diploma), HSGED (receipt of a GED certificate), and LFTSCHAG (age when the R left school). Consequently, any data in these items were wiped out in the edited variables. Data were not wiped out for LFSCHWHY (reason for leaving school) because that was the variable responsible for inferring that the R was currently enrolled.</p>
<p>The R reported not being currently enrolled in school. In the question about reasons for leaving school without getting a high school diploma (LFSCHWHY), however, the R specified that he or she was being homeschooled (LFSCHWHY = 601).</p>	<p>The R was logically inferred to be currently enrolled in school. A special code of 5 was assigned to the edited school enrollment variable SCHENRL. As above, any data in HSDIPLMA, HSGED, and LFTSCHAG were wiped out. Data were not wiped out for LFSCHWHY because that was the variable responsible for inferring that the R was currently enrolled. If the R was aged 12 to 18, a code of 14 was assigned to the variable SCHTYPE (Homeschooled, LOGICALLY ASSIGNED) to indicate that the R was logically inferred to be homeschooled. No editing was done to SCHTYPE if the R was older than 18. In addition, this edit was not implemented if interviewers specified that the respondent had <i>completed</i> homeschooling. In these latter situations, LFSCHWHY was assigned a code other than 601 (see below).</p>
<p>The R reported not being currently enrolled in school, reported receiving a high school diploma, but reported completing the 10th or 11th grade.</p>	<p>No editing was done, and the variable pertaining to receipt of a high school diploma (HSDIPLMA) retained a value of 1 (i.e., "yes"). The rationale was that the R may have gone through school on an accelerated pace or may have otherwise qualified for a high school diploma with fewer than 12 years of education (e.g., if the R went to school in another country).</p>
<p>The R reported not being enrolled in school but having received a high school diploma. However, the R had completed only the 9th grade or lower.</p>	<p>The R was logically inferred in HSDIPLMA not to have received a high school diploma. HSDIPLMA was assigned a code of 4 (No LOGICALLY ASSIGNED).</p>
<p>The R reported not being enrolled and not having received a high school diploma. In the question about reasons for leaving school without getting a high school diploma, however, the R specified that he or she had gotten a diploma. That included situations where the R may have received a diploma in another country.</p>	<p>The R was logically inferred to have received a high school diploma, provided that the R had completed the 10th grade or higher. The edited variable HSDIPLMA was assigned a code of 3 (Yes LOGICALLY ASSIGNED). Although Rs would have skipped the question about getting a GED if they had answered "yes" to the question about getting a high school diploma (QD22), no editing was done to the GED variable HSGED (QD23) when Rs were logically inferred to have gotten a high school diploma. In addition, no editing was done to LFSCHWHY and LFTSCHAG in this situation.</p>

(continued)

Exhibit 3. Edit Issues Pertaining to the Noncore Education Section (continued)

Issue	Edits Implemented
<p>The R reported not being enrolled in school, not having received a high school diploma, and not having received a GED certificate. In the question about reasons for leaving school without getting a high school diploma, however, the R specified that he or she had received a GED (LFSCHWHY = 606).</p>	<p>The R was logically inferred to have gotten a GED certificate. For this edit to be implemented, however, the R had to have indicated explicitly that he or she had actually received a GED, not that he or she was working on a GED.</p>
<p>The R reported not being enrolled in school, not having received a high school diploma, and having received a GED certificate. In the question about reasons for leaving school without getting a high school diploma, however, the R specified that he or she had completed homeschooling but had not received a GED (LFSCHWHY = 621).</p>	<p>The R's response of "yes" in HSGED was considered to be questionable based on the response in LFSCHWHY. HSGED was set to a value of 11 (i.e., "bumped" by 10). For this edit to be implemented, however, the R had to have indicated explicitly that he or she had actually received no GED after having completed homeschooling.</p>
<p>The R reported leaving school at an age greater than his or her current age.</p>	<p>The edited variable corresponding to the age at leaving school was assigned a bad data code.</p>
<p>The R reported leaving school at age 3 or younger, or the R reported leaving school at an age that was considered too young for the highest grade that he or she reported completing (e.g., completed the 11th grade but reported leaving school at age 13 or younger).</p>	<p>The edited variable corresponding to the age at leaving school was assigned a bad data code.</p>
<p>The R was aged 23 or younger and reported not being enrolled in school. However, the interview was conducted in June, July, or August (i.e., when school was not in session). The R also originally reported getting a high school diploma but was inferred not to have received one (i.e., the R completed the 9th grade or lower). In addition, the highest completed grade was within 1 year (in either direction) of the expected completed grade for that R's age.</p>	<p>A code of 52 was assigned to the school enrollment variable SCHENRL only if there was some ambiguity (i.e., a response of "don't know" or "refused") in the follow-up questions QD17b or QD17c, regarding the R being on break from school or intending to return to school when the R's break was over, respectively. This code of 52 was intended to indicate to analysts that there was some uncertainty about the R's current enrollment status due to (a) the interview being conducted in summer months when most schools are not in session, and (b) ambiguity in the R's answers to QD17b or QD17c.</p> <p>Otherwise, no editing was done to SCHENRL (i.e., the R continued to be classified as not enrolled in school) when the R indicated in QD17b that he or she was not on break from or the R was on break but indicated in QD17c that he or she did not intend to return to school once this break was over.</p>

(continued)

Exhibit 3. Edit Issues Pertaining to the Noncore Education Section (continued)

Issue	Edits Implemented
The number of days in the past 30 that the R missed school because the R was sick and the number of days that the R skipped school added up to more than 30.	No editing was done in this situation.
The R reported being currently enrolled in school but reported skipping school all 30 days in the past 30 days.	A code of 11 was assigned to the school enrollment variable SCHENRL. This code was intended to indicate to analysts that there was some uncertainty about the R's current enrollment status.
The R reported being currently enrolled in school but reported in question QD20 that he or she missed school because of sickness for more than 30 days. This pattern was observed in earlier years because a code of 90 was used to mean "school not in session," and the computer-assisted interviewing (CAI) program code did not allow for discontinuities in the allowable range.	Values of 31 days were set to 30 days. Values greater than 31 days but fewer than 90 (i.e., school not in session) were replaced with bad data codes. This logic was in place, but values greater than 31 and fewer than 90 did not occur in the data.

also were used to edit the variables pertaining to receipt of a high school diploma (HSDIPLMA) or receipt of a GED certificate of high school completion (HSGED). For example, if the FI comments indicated that respondents had received a high school diploma, HSDIPLMA could be assigned a code of 3 (Yes LOGICALLY ASSIGNED), and the remaining variables HSGED, LFSCHWHY, and LFTSCHAG could be assigned legitimate skip codes. In the absence of information in the FI comments that would permit editing of additional variables, HSDIPLMA, HSGED, LFSCHWHY, and LFTSCHAG were left as blank because these respondents who were logically inferred not to be enrolled were skipped out of questions that were relevant to respondents who were not enrolled.

The following potential patterns of inconsistent or questionable data could occur between QD18 and QD11 despite the presence of the "hard error" check between the two questions:

- the hard error was triggered, but the case was allowed to proceed through the general education edits for the reasons described above;
- the hard error was not triggered, the current grade in QD18 was exactly two grades higher than the highest grade completed (from QD11), but the respondent was at a current grade level that would be expected for someone at his or her age (e.g., if a 12 year old reported last completing the 4th grade and reported currently being in the 6th grade); or
- the hard error was not triggered, and the current grade in QD18 was more than two grade levels higher than the highest grade from QD11.

An algorithm was developed to handle these types of situations when they occurred. This algorithm is discussed in detail below. In particular, having accurate data on respondents' current

grade levels would be important for comparing NSDUH data with drug use data from in-school surveys, such as Monitoring the Future, that are administered to students in specific grades.

For respondents aged 12 to 23, a series of arrays was set up that mapped out the highest grade and current grade that would be *expected* relative to a respondent's current age, assuming an orderly progression from one grade level to the next highest level. The following matrix maps the current age with expected grades:

Current Age	12	13	14	15	16	17	18	19	20	21	22	23
Expected Completed Grade	6	7	8	9	10	11	12	13	14	15	16	17
Expected Current Grade	7	8	9	10	11	12	13	14	15	16	17	17

For example, one might expect most people in the United States to have completed the 6th grade by the time they are 12. It would therefore not be unreasonable for someone to be aged 12 and to be currently in the 7th grade, depending on when the respondent was interviewed. An upper age limit was set at 23 because a grade level of 17 (college or university, 5th year or higher) was the upper limit of the education levels.

In addition, the algorithm allowed for some deviation relative to the expected ages, as described below. Thus, if a respondent was aged 12, had completed the 5th grade, and was currently in the 6th grade, that would be an acceptable pattern because the respondent might have had his or her 12th birthday at some point during the 6th grade.

Separate edits were done depending on whether a respondent was aged 12 to 18 or was older than 18. The rationale for doing edits separately for these two different age groups was that the typical progression from one grade level to the next would be less likely to hold for adults and at higher educational levels. Suppose, for example, that a respondent completed 3 years of college but changed majors and not all of the prior credits applied to the new major. It would be possible for the respondent to report having completed 3 years of college and to be currently enrolled at a level lower than the third year of college, depending on how the respondent interpreted these questions. Similarly, a respondent who got a bachelor's degree in one field and went back to school for a second bachelor's degree might report having completed 4 years of college but also might report currently being enrolled at some level below the 4th year of college.

Edits When Respondents Were Aged 12 to 18. For respondents aged 12 to 18, the highest grade completed and the current grade were considered to be consistent with the respondent's age if what was reported was within 1 year of the grades given in the matrix described above. Thus, if a respondent was aged 12, the algorithm would consider completion of any grades 5 through 7 to be sufficiently consistent with the respondent's age. Similarly, for respondents aged 12, the algorithm would consider current enrollment in any grades 6 through 8 to be sufficiently consistent with the respondent's age.

Therefore, the following four data combinations were possible:

- both the completed grade and the current grade were consistent with the respondent's age;
- the highest completed grade was consistent with the respondent's age, but the current grade was not;
- the current grade was consistent with the respondent's age, but the highest completed grade was not; or
- neither the highest completed grade nor the current grade was consistent with the respondent's age.

Separate edits were done according to the four combinations of data patterns described immediately above. The following edits were done if both the completed grade and current grade appeared to be consistent with the respondent's age:

- If the current grade was more than two grade levels higher than the highest completed grade, the current grade was edited to be consistent with the highest grade because the latter was a core variable. For example, if a respondent was aged 17, reported completing the 10th grade, and reported a current grade of 13 (i.e., first year in college), the edits logically inferred that the respondent currently was in grade 11. The edited variable for current grade (EDUCATND) was assigned a code of 31 (i.e., 11th grade LOGICALLY ASSIGNED).
- If the current grade was lower than the highest completed grade, the edit code gave precedence to the reported grade that would yield the most consistent result relative to the respondent's age. In particular, if accepting the report of the highest grade and inferring that the respondent was currently in the next highest grade would yield a current grade that was inconsistent with the respondent's age, then the noncore-created variable EDTEDUC (i.e., edited highest grade completed) was assigned a value consistent with the current grade. Suppose, for example, that a 12-year-old respondent reported currently being in the 6th grade but completed the 7th grade. Accepting the answer that the respondent was currently in the 6th grade and completed grade 5 would be more consistent with the respondent's current age than would be the converse (i.e., accepting that this respondent had completed the 7th grade and inferring that he or she was currently in the 8th grade). In this example, EDTEDUC would be assigned a code of 25 (i.e., 5th grade LOGICALLY ASSIGNED).

If the highest completed grade was consistent with the respondent's age but the current grade was not, the highest completed grade was accepted by default. This was done if the current grade was lower than the highest completed grade or the current grade was more than two grade levels higher than the highest completed grade. The edited current grade EDUCATND was therefore assigned a value to indicate a current grade level that was 1 year higher than the highest completed grade. For example, if the respondent reported completing grade 10, EDUCATND would be assigned a code of 31 (i.e., 11th grade LOGICALLY ASSIGNED).

If the current grade was consistent with the respondent's age but the highest completed grade was not, the edit procedures accepted the current grade by default. Thus, if a 12-year-old

respondent reported last completing the 4th grade and reported currently being in the 6th grade, this edit would identify the current grade of 6 as being consistent with an age of 12; completing the 4th grade would not be identified as consistent with an age of 12. In this example, EDTEDUC would be assigned a code of 25 (i.e., 5th grade LOGICALLY ASSIGNED).

If neither the current grade nor the reported highest grade was consistent with the respondent's age, the following was done:

- If the current grade was lower than the highest grade reported, the algorithm picked the answer that was closest to the expected grade, based on the matrix shown above. The variable with the more inconsistent data was assigned a bad data code. This edit allowed for situations where respondents may have fallen behind where they would be expected to be grade-wise (e.g., if they had been held back a year).
- If the current grade was exactly two grade levels higher than the reported highest completed grade and the highest completed grade was higher than what would be expected for the respondent's age, no further editing was done. Otherwise, the created noncore variable EDTEDUC was assigned a bad data code. This edit was designed to allow for situations where a respondent might be on an accelerated track.
- If the current grade was more than two grade levels higher than the reported highest grade and it was lower than the expected current grade, then the value was retained for the current grade. The variable EDTEDUC was assigned a bad data code. In other situations, both EDTEDUC and EDUCATND (i.e., the edited current grade) were assigned codes of bad data. The rationale for the first edit was that, if EDUCATND was lower than the expected current grade, EDUCATND would be more consistent with the *expected* current grade and the respondent's age than what the reported highest grade would be.

Edits When Respondents Were Older Than 18. Minimal editing of EDTEDUC and EDUCATND was done for respondents aged 19 or older. Other than the edits described below, no other editing of the educational level data was done for respondents aged 19 or older.

If the current grade was lower than the highest completed grade and the current grade was at the 12th grade or lower, then EDUCATND (i.e., the current grade) was assigned a bad data code. Otherwise, no further editing was done when the current grade was lower than the highest grade. For example, if a respondent reported completing the 12th grade but reported currently being in grade 1, the latter response would probably indicate a typographical error. The first edit described in this paragraph would assign a bad data code to EDUCATND.

If the current grade was more than two grade levels higher than the highest completed grade and the current grade was above the 12th grade, the edits compared what the highest grade completed would be relative to the current grade, if the highest completed grade were actually increased by 10 years. If bumping the highest completed grade by 10 years yielded a completed grade that was still less than or equal to the reported current grade, then the variable EDTEDUC was assigned a code of bad data. In this situation, the interpretation was that a typographical error was made for the highest grade. Otherwise, no further editing was done. The first edit described in this paragraph was based on observed patterns that suggested that keying errors may

have been made in QD11 (highest grade completed). For example, there were respondents who reported completing grade 1 and currently being in their 13th or higher years of school. Again, this pattern suggested that the second digit did not get keyed in QD11. This edit gave respondents credit for being enrolled in a grade above the high school level.

3.2.4 Employment and Workplace

Respondents aged 15 or older were asked questions about their current employment, employment history, and characteristics of their workplace (if applicable). Question QD26 asked whether respondents worked in the week prior to the interview. If respondents reported that they did not work in the past week, they were asked in question QD27 whether they had a job or business. Respondents then were routed through different branches of work-related questions depending on how they answered these two key questions. For example, respondents who worked in the past week were asked questions to determine full-time or part-time work status (e.g., whether they usually worked 35 or more hours per week), whether they ever had a period of unemployment in the past 12 months, the number of days they missed work in the past 30 days because they were sick or because they did not want to be at their workplace, and characteristics of their workplace, particularly with respect to alcohol and other drug policies at their workplace. Similarly, respondents who did not work in the past week and did not have a job were routed into questions relevant for people who currently were not working, such as why they did not have a job, whether they made specific efforts to find work in the past 30 days, and the month and year when they last worked for pay.

The employment and workplace questions and logic underwent important changes in 2001. These changes are discussed below. Associated variable naming conventions and edits have been in place since 2001, unless indicated otherwise.

- In the questions pertaining to reasons that respondents did not work in the past week despite having a job (QD30) or reasons that respondents did not have a job in the past week (QD31), respondents who reported "some other reason" for not working in the past week or not having a job were not asked to specify what these other reasons were. Prior to 2001, respondents were asked to specify these other reasons, and these "OTHER, Specify" answers were taken into account to determine respondents' employment status (edited variable JOBSTAT in 1999 and 2000). Because these "OTHER, Specify" data were no longer available beginning in 2001, the names of these variables were changed to WRKNOWRK (corresponding to QD30) and WRKNOJOB (corresponding to QD31). These variables previously had been named WRKNORS1 and WRKNORS2, respectively. In addition, the name of the recoded employment status variable was changed from JOBSTAT (in 1999 and 2000) to JBSTATR in 2001.
- In question QD31 (edited variable WRKNOJOB), a new category was created in 2001 for respondents who did not have a job in the past week because they did not want one. Therefore, JBSTATR included a category for persons who endorsed this response in QD31. In addition, response category 3 in QD31 since 2001 has read as "KEEPING HOUSE OR TAKING CARE OF CHILDREN FULL-TIME" instead of the wording "KEEPING HOUSE FULL-TIME" that was used prior to 2001. This

change might have affected how respondents answered QD31. However, no changes were made to employment status categories in JBSTATR due to this wording change.

- Since 2001, in the questions about the number of employers that respondents had in the past 12 months (question QD35, if respondents reported being self-employed; question QD36 otherwise), respondents have not been not allowed to report that they had "0" employers in the past 12 months. The name of the edited variable corresponding to these questions (WRKJOBS) did not change. Due to this change, however, no respondents needed to be inferred to have had at least one job in the past 12 months. Since 2001, therefore, the code of 975 (At least one LOGICALLY ASSIGNED) has no longer applied to WRKJOBS.
- Since 2001, questions on the year and month that respondents last worked for pay (QD39a and QD39b, respectively) have had a numeric format. Prior to 2001, this information was captured in an alpha format (question QD39), with interviewers being instructed to enter the month and year data in the format of "MM/YYYY." The old alpha format required considerable data cleaning because interviewers did not always enter the information in the requested format. In addition, the routing logic for asking respondents for the year and month when they last worked for pay changed in 2001. Prior to 2001, respondents who reported that they did not work in the past week (QD26 not answered as "yes") were asked to provide this information. In 2001, the logic changed to ask this information of respondents who did not report that they had a job in the past week (QD27 not answered as "yes"). This logic change affected the assignment of legitimate skip codes to the year and month variables. For these reasons, the variables pertaining to the year and month that respondents last worked for pay were changed to WRKLSTYR (formerly WRKLASYR) and WRKLSTMN (formerly WRKLASMO), respectively.
- In 2005, the logic for asking about the respondent's month of last employment (question QD39b) underwent a further change. Since 2005, this question has been asked only for those respondents who reported in question QD39a (edited variable WRKLSTYR) that they last worked in the current year or the prior year (i.e., 2010 or 2011); in 2001 through 2004, all respondents who gave a valid year in question QD39a were asked to report the month they last worked in QD39b. Consequently, more respondents in 2005 legitimately skipped out of question QD39b compared with respondents in prior years. For this reason, the edited variable corresponding to question QD39b has been called WRKLSTMN since 2005. In 2004, this variable was called WRKLSTMO. If the year in WRKLSTYR was more than 1 year prior to the current survey year, WRKLSTMN was assigned a legitimate skip code. Because of the changes to the skip logic in 2005, the distribution of WRKLSTMN also was not comparable with the distribution of WRKLSTMO prior to 2005.

In addition, coding procedures changed in 2003 for the industry in which respondents worked (currently or in the past year) and for respondents' occupations (for their current or previous job). Since 2003, the Census Bureau has coded these responses. In addition, the industry and occupation (I&O) codes and classification procedures since 2003 have been based on those from the 2000 census. Therefore, the names of the I&O variables were changed, beginning with the 2003 NSDUH. The former names of the I&O variables were as follows:

WRKINDUS (industry in which the respondent is currently working), WRKOCCUP (respondent's current occupation), WRKINDYR (industry in which the respondent formerly worked in the past year), and WRKOCCYR (respondent's former occupation in the past year). Since 2003, these variables have been called WRKIDSTY, WRKOCUPT, WRKIDSYR, and WRKOCUYR, respectively. Recoded versions of these variables (WRKIDST2, WRKOCUP2, WRKIDSY2, and WRKOCUY2) were created for the NSDUH public use file. Despite these changes, the procedures for editing these I&O variables have not changed.

An important aspect of editing the work-related variables involved identification of situations where questions had been legitimately skipped. A second key aspect of processing the work-related variables was to use the data to establish respondents' current work status. As noted above, a single, recoded work status variable named JBSTATR was created that served as the starting point for creation of a final, statistically imputed employment status variable (EMPSTAT4). JBSTATR was created from the following final variables: WRKEDWK (whether the respondent worked in the past week), WRKHAVJB (whether the respondent had a job if he or she did not work in the past week), WRKHRSUS (whether the respondent usually worked 35 or more hours per week), WRKNOWRK (reason for not working in the past week despite having a job), WRKNOJOB (reason for not having a job in the past week), WRKEFFRT (made specific efforts to find work in the past 30 days), and WRKEDYR (whether the respondent had a job in the past 12 months).

Based on the data in these variables, respondents aged 15 or older were assigned to one of the following categories in JBSTATR:

- worked at a full-time job in the past week;
- worked at a part-time job in the past week;
- had a job but out because of some temporary absence from work, such as vacation or being sick;
- had a job but out because of a layoff, and the respondents were looking for work;
- had a job but out because of a layoff, and the respondents were not looking for work;
- had a job but out because the respondents were waiting to report to a new job;
- had a job but out because the respondents were self-employed and did not have any business in the past week;
- had a job but out because the respondents were in school or training in the past week;
- did not have a job, unemployed or on layoff, and looking for work;
- did not have a job, unemployed or on layoff, and not looking for work;
- did not have a job because the respondents were keeping house or taking care of children full time;
- did not have a job because the respondents were in school or training (e.g., as full-time students, as opposed to a temporary absence from work due to school or training);
- did not have a job because the respondents were retired;

- did not have a job because the respondents were disabled; or
- did not have a job because the respondents did not want one (see above).

If respondents reported that they did not work in the past week for some other reason despite having a job, JBSTATR was assigned the following nonspecific codes, depending on whether information was available regarding the usual number of hours worked: 190 (has full-time job, reason for not working unknown), 191 (has part-time job, reason for not working unknown), or 199 (has job, no further information). Similarly, if respondents reported that they did not have a job for some other reason, they were assigned a nonspecific code of 290 (unemployed, no further information).

In addition, respondents who reported in question QD31 that they did not have a job but were looking for work were not classified as being unemployed unless they reported in WRKEFFRT that they had made specific efforts in the past 30 days to find work (such as making contacts with someone about a job, sending out resumes or job applications, or placing or answering ads). If respondents reported that they did not have a job but were looking for work but WRKEFFRT was not answered as "yes," they were classified as not in the labor force (code 299) in JBSTATR.

If respondents did not know or refused to report whether they worked in the past week, WRKEDYR was checked for indications of whether respondents worked in the past year. Respondents who indicated in WRKEDYR that they did not work in the past 12 months were classified as not having a job (JBSTATR = 290). Otherwise, if respondents did not provide information on whether they worked in the past week (i.e., QD26 answered as "don't know" or "refused"), JBSTATR was assigned the corresponding code of "don't know" or "refused."

[Exhibit 4](#) discusses additional issues that were relevant to the processing of the work-related variables. As noted above, for example, the question pertaining to the month that respondents last worked for pay was changed in 2005. In addition, if respondents reported in question QD39a that they never worked for pay, interviewers were instructed to enter a response of 9991. When the month question QD39b had been skipped because a response of 9991 had been entered in QD39a, the edited month variable WRKLSTMN was recoded as 91. Documentation for 9991 (or 91) was as follows:

9991 = NEVER WORKED AT A JOB OR BUSINESS.

A refinement to the editing procedures for the employment and workplace section also has been implemented since 2003 for respondents who reported that they did not work in the past week (WRKEDWK = 2). The variable pertaining to the number of hours that respondents worked in the past week (WRKHRSWK) was assigned a legitimate skip code regardless of how respondents answered the question about having a job (QD27). Prior to 2003, a legitimate skip code was assigned to WRKHRSWK only if respondents reported that they did not work in the past week (WRKEDWK = 2) and QD27 was answered as "no"; the prior logic did not assign legitimate skip codes to WRKHRSWK if respondents answered QD27 as "don't know" or "refused." Logically, however, if respondents reported that they did not work in the past week, they would not have worked any hours at a job during that period, regardless of how they answered question QD27.

Exhibit 4. Edit Issues Pertaining to the Employment and Workplace Section

Issue	Edits Implemented
<p>The respondent (R) reported working in the past week in question QD26. However, the R subsequently reported being without a job at some point in the past 12 months and reported being without a job during all 52 weeks in the past 12 months. Because all 52 weeks of the 12-month period prior to the interview would include the week prior to the interview, it would be inconsistent for an R to report working in the past week but not working for all 52 weeks in the past year.</p>	<p>The edited variable pertaining to the number of weeks without a job in the past 12 months (WRKUNWKS) was assigned a bad data code.</p>
<p>The R reported working in the past week. However, the R subsequently reported missing work for all 30 of the past 30 days because he or she was sick or did not want to be at work (or both). Because the past week was included in the 30 days prior to the interview, it would be inconsistent for an R to report working in the past week but missing work for every day in the past month.</p>	<p>The following edits were implemented in this situation:</p> <ul style="list-style-type: none"> • If the R reported that he or she missed work for all 30 days in the past month because he or she was sick, the edited variable (WORKDAYS) was assigned a bad data code. • If the R reported missing work for all 30 days in the past month because he or she did not want to be there, the edited variable (WORKBLAH) was assigned a bad data code.
<p>The R did not know or refused to report in question QD26 whether he or she worked in the past year. However, the R also reported in question QD33 (edited variable WRKEDYR) that he or she did not have a job in the past 12 months.</p>	<p>The R was logically inferred not to have worked in the past week (WRKEDWK = 4) and not to have had a job in the past week (WRKHAVJB = 4), where 4 = No LOGICALLY ASSIGNED. Subsequent employment and workplace variables that could be assigned legitimate skip codes were edited as though QD26 and QD27 had been answered as "no."</p>
<p>The R answered question QD26 (worked in the past week) as "no" but answered question QD27 (having a job in the past week) as "don't know." Edit logic prior to 2003 left the variable pertaining to the number of hours worked in the past week WRKHRSWK as blank (i.e., a legitimate skip code was not assigned to WRKHRSWK). The prior logic for assigning legitimate skip codes to WRKHRSWK was part of the logic for assigning legitimate skips when both QD26 and QD27 were answered as "no." However, only the response to question QD26 truly applies to WRKHRSWK.</p>	<p>Since 2003, a legitimate skip code has been assigned to WRKHRSWK when QD26 is answered as "no" (QD26 = 2), independent of how QD27 was answered.</p>
<p>The reported year when the R last worked for pay was fewer than 5 years from the R's birth year (including situations where the year the R reported last working for pay was earlier than the year the R was born).</p>	<p>The edited variables WRKLSTMN and WRKLSTYR were assigned bad data codes.</p>

(continued)

Exhibit 4. Edit Issues Pertaining to the Employment and Workplace Section (continued)

Issue	Edits Implemented
<p>The R was not asked whether he or she was self-employed in the past 12 months because the R had already given an answer indicating that he or she had been self-employed. This could occur in one of the ways listed below.</p> <ul style="list-style-type: none"> • The R reported not working in the past week because he or she was self-employed and did not have any business (QD30 = 5). • The R reported in question INOC06 that the category that best described the business in which he or she worked was one in which the R was self-employed (INOC06 answered as 7 or 8). 	<p>The edited variable pertaining to self-employment in the past 12 months (WRKSLFEM) was assigned a code to indicate that "yes" could be logically inferred. This was done instead of assigning a legitimate skip code. This edit did not apply if INOC06 indicated that Rs worked without pay in a family business or farm.</p>
<p>The R did not report being self-employed at any time in the past 12 months but reported having a job. However, the industry and occupation (I&O) question pertaining to the R's last job (INOC08) indicated that the R was self-employed in an incorporated or unincorporated business (edited variable WRKBZCYR, corresponding to INOC08, had a value of 7 or 8).</p>	<p>The edited variable WRKSLFEM was logically inferred to have been answered as "yes," provided that the following conditions held:</p> <ul style="list-style-type: none"> • The R reported working in the past year (WRKEDYR = 1), such that reported self-employment in INOC08 would pertain to self-employment in the past year. • The year and month that the R reported last working for pay (WRKLSTYR and WRKLSTMN) also were consistent with the R reporting that he or she worked in the past year. <p>The following data in WRKLSTYR and WRKLSTMN were considered to be consistent (or at least not contradictory) with indications that the R worked in the past year (WRKEDYR = 1):</p> <ul style="list-style-type: none"> • The R reported last working for pay in the current interview year. • The R reported last working for pay in the previous year, and the month that the R reported last working for pay was within 12 months of the interview date, or was the same month as the interview date. • The R reported last working for pay in the previous year, but the month that the R reported last working for pay had a missing value. In this situation, WRKEDYR = 1, and an indication of self-employment in INOC08 was still allowed to infer in WRKSLFEM that the R had been self-employed in the past 12 months.

(continued)

Exhibit 4. Edit Issues Pertaining to the Employment and Workplace Section (continued)

Issue	Edits Implemented
	<p>WRKSLFEM was not logically inferred to be "yes" if the R reported working in the past year (WRKEDYR = 1), WRKBZCYR = 7 or 8, but any of the following occurred:</p> <ul style="list-style-type: none"> • The R reported last working for pay in the previous year, and the month that the R reported last working for pay was more than 12 months beyond the interview date. • The R had missing data for the year when he or she last worked for pay (e.g., if WRKLSTYR was refused). • A problem had been identified with the interview date that was stored by the computer-assisted interviewing (CAI) system while the interview was in progress.

3.2.5 Health Insurance

Respondents (or other household members serving as proxies) were asked whether they (or the respondents) were currently covered by different types of health insurance.⁸ If private health insurance coverage was reported, respondents were asked whether that included coverage for substance abuse treatment or mental health services. Data also were collected on periods when respondents never had health insurance coverage, former coverage that they may have had, and reasons for losing health insurance coverage or for never having had coverage.

The health insurance section underwent important changes in 2001. These are described below. Unless indicated otherwise, these changes have been in place since 2001. In addition, the content and associated editing of the health insurance variables did not change relative to the previous survey year.

- If respondents were aged 12 to 19, they were asked question QHI02A to determine whether they were covered by the Children's Health Insurance Program (CHIP). Government experts in the health insurance field advised the Substance Abuse and Mental Health Services Administration (SAMHSA) that it would be virtually impossible to produce separate estimates of coverage by the Medicaid program (question QHI02) and coverage by CHIP (QHI02A). For this reason, the variable CAIDCHIP was created from responses to QHI02 and QHI02A. Creation of CAIDCHIP and related issues are discussed below in further detail.
- As noted in Section 3.2.1, interviewers were instructed to use the field interviewer (FI) checkpoint⁹ FIPE4 (edited variable STATELOC) to report the State where the sampled dwelling unit (SDU) was located. Interviewers were requested to report this in the FI checkpoint FIPE4 at the beginning of the interview. This information from

⁸ For the sake of brevity, reference is made only to "respondents" in the remainder of this section. However, readers are advised that the health insurance information for a respondent may have been provided by another adult household member who was serving as a proxy for the respondent because the proxy was considered to be better able to answer the health insurance questions for the respondent.

⁹ See footnote 6 in Section 3.2.1.

FIPE4 was used to fill in information in questions QHI02 and QHI02A regarding State-specific Medicaid program or CHIP names to aid respondent identification of whether they were covered by Medicaid or CHIP.

- Since 2001, respondents who answered "no" to all questions about Medicare, Medicaid, CHIP (if applicable), military health coverage, and private health insurance have been asked follow-up question QHI11 to determine if they were covered by *any* type of health insurance. Responses to QHI11 have been used to determine subsequent routing in the health insurance section depending on whether respondents currently had or did not have health insurance. The variable HLTINNOS was created from QHI11.
- A recoded "any health insurance" variable, ANYHLTIN, was created from responses to MEDICARE (from QHI01), CAIDCHIP (from QHI02 and QHI02A), CHAMPUS (from QHI03), PRVHLTIN (from QHI06), and HLTINNOS (from QHI11). If any affirmative response was reported in any of the above variables, ANYHLTIN was coded as 1 ("yes"). Otherwise, if HLTINNOS had been answered as "no" (and by definition, preceding questions had been answered as "no"), ANYHLTIN was coded as 2 ("no"). If ANYHLTIN was not already coded as 1 or 2, it was coded as 97 ("refused") or 94 ("don't know"), as follows: (a) if a code of 97 occurred in any of the above health insurance items, ANYHLTIN was coded as 97; or (b) ANYHLTIN was coded as 94 if a code of 94 (but no code of 97) occurred in the above items. For remaining cases (e.g., if variables had been set to bad data, or a breakoff had occurred), ANYHLTIN retained a code of 98 (OTHER MISSING).
- Since 2001, question QHI16 from the 2000 survey (type of health insurance that respondents last had, if they were not currently covered by health insurance) has not been included in the interview. Therefore, issues that were relevant to the editing of health insurance variables based on the respondents' last coverage have not been relevant in editing health insurance variables since 2001.
- Question QHI17 (reason that respondents lost health insurance coverage, if they previously had it) was an "enter all that apply" question prior to 2001 in which respondents could report multiple reasons that they lost health insurance coverage. At present, however, this question asks respondents to report the *main* reason they stopped being covered by health insurance, and only one response could be chosen from the list. Therefore, a single variable, HLLOSRSN, now corresponds to QHI17.
- After 2000, "OTHER, Specify" variables pertaining to "other" reasons that respondents lost their health insurance or never had health insurance were no longer included in the interview. Consequently, additional data are not available to edit the variable HLLOSRSN or the variables pertaining to reasons for never having health insurance (HLNVCOST through HLNVNEED).

Because the names for Medicare and Medicaid sound similar, follow-up questions have been included in the health insurance section of the interview since 2003 for respondents who were (a) under the age of 65 and reported being covered by Medicare (which is for persons aged 65 or older or certain disabled persons), or (b) aged 65 or older and reported being covered by Medicaid (which is for low-income persons or disabled persons). In these situations, respondents

were asked to confirm their previous answer. If respondents indicated on follow-up that their previous answer to Medicare or Medicaid coverage was incorrect, respondents were determined not to be covered by that particular type of health insurance. However, if respondents did not know or refused to confirm whether their previous answer to the Medicare or Medicaid coverage question was correct, the editing procedures continued to classify them as having that type of coverage. Stated another way, the only response on follow-up that overruled a previous report of coverage by Medicare or Medicaid was for respondents to indicate that they did not have that type of coverage.

As noted above, the variable CAIDCHIP was created from responses to questions QHI02 (regarding Medicaid coverage) and QHI02A (regarding coverage by CHIP). This CAIDCHIP variable indicated whether respondents were covered by Medicaid *or* CHIP. This variable replaced the variable MEDICAID that existed prior to 2001. However, the statistical imputation team still used information from question QHI02 (coverage by Medicaid) to create the imputed health insurance variable IRINSUR for comparability with data prior to 2001.

If STATELOC had been set to bad data because of inconsistencies in the State information for the respondent (see Section 3.2.1), CAIDCHIP was usually assigned a bad data code as well. The rationale for this edit was that the CAI logic would supply an incorrect name for the State's Medicaid program or CHIP if the information in FIPE4 was incorrect. Consequently, the respondent would be answering QHI02 or QHI02A based on a version of the question that did not correctly correspond to where the respondent would be eligible for publicly funded health insurance coverage. For example, if a respondent was aged 12 to 19 and was living in California (FIPE4 = 5), the respondent should have been asked in QHI02A whether he or she was covered by "Healthy Families." However, if a value of 6 had been entered in FIPE4 (i.e., for Colorado), the respondent would be asked whether he or she was covered by "Child Health Plan Plus, or CHP+."

An exception to this assignment of bad data codes concerned the special situation in which respondents were routed to questions QHI15 (time since the respondent last had health insurance) and QHI17 (main reason for losing health insurance coverage). If responses to QHI15 or QHI17 indicated that the respondent did not currently have (or never had) health insurance coverage, CAIDCHIP retained a code of 2 (i.e., "no"), even if STATELOC had been set to bad data, for consistency with information from QHI15 or QHI17 that the respondent was not currently covered by any type of health insurance.

If STATELOC had a valid value, CAIDCHIP was assigned a code of 1 ("yes") if an affirmative response occurred in either QHI02 or QHI02A (if applicable). CAIDCHIP was coded as 2 ("no") if QHI02 was answered as "no" and (a) QHI02A also was answered as "no" (for respondents who were aged 12 to 19) or (b) QHI02A had been legitimately skipped (for respondents aged 20 or older). Otherwise, CAIDCHIP was coded as 97 ("refused") if a code of 97 occurred in either QHI02 or QHI02A, or 94 ("don't know") if a code of 94 (and no code of 97) occurred in these items. Remaining cases that did not meet any of these criteria were coded as 98 (i.e., blank).

An important additional aspect of editing the health insurance variables consisted of assigning legitimate skip codes based on the skip logic in this section. For example, if

respondents answered "no" (where applicable) to questions QHI01 through QHI06 and then reported in QHI11 that they were not currently covered by any kind of health insurance (QHI11 = 2), legitimate skip codes were assigned to HLCNOTYR (anytime in the past 12 months that respondents were without health insurance, corresponding to question QHI13) and HLCNOTMO (number of months that respondents were without health insurance in the past 12 months, corresponding to question QHI14). Similarly, if respondents reported some type of current health insurance coverage in QHI01 through QHI06, edited variables corresponding to questions QHI15 through QHI18 were assigned legitimate skip codes (i.e., HLCLAST through HLNVSOR).

As was the case in prior years, question QHI18 (reasons that the respondent never had health insurance) was an "enter all that apply" question. Therefore, the edited variables corresponding to question QHI18 (HLNVCOST through HLNVSOR) were assigned a code of 1 (Response entered) if the corresponding response category was chosen from QHI18. The variables were assigned a code of 6 (Response not entered) if the corresponding response category was not chosen but at least one response had been entered in QHI18.

[Exhibit 5](#) discusses additional issues that were relevant to the processing of the health insurance variables. For example, the data could indicate that respondents were covered currently by Medicare, Medicaid, CHIP (for respondents who were aged 12 to 19), some type of military health coverage (e.g., CHAMPUS or the VA), or private health insurance. If respondents were reported to have been *currently* covered by all of the types of insurance they were asked about, a flag was set and included on the data file. The original data were retained, but this flag was designed to alert analysts to the presence of this unlikely data pattern.

In addition, the only types of current health insurance coverage that were asked about in 1999 were Medicare, Medicaid, some type of military health coverage, or private health insurance. Therefore, a second flag was set for comparability with a similar flag set in the 1999 data. This second flag indicated when respondents reported that they were covered currently by all four of these types of health insurance that were asked about in 1999, even if they did not report being covered by CHIP (if aged 12 to 19) or they were aged 20 or older and were skipped out of question QHI02A.

3.2.6 Edited Income

Since 2004, editing of the income questions has been handled by the logical editing team. Prior to 2004, these variables were edited by the statistical imputation team in conjunction with the procedures for creating final, imputed income variables.¹⁰

Unlike other sections of the interview that asked about behaviors or situations in the 12-month period prior to the interview, the reference period for the income module was the previous calendar year (i.e., January 1, 2010, through December 31, 2010). As was the case with the health insurance section, an adult family member who was better able than the respondent to answer the income questions could serve as a proxy for the respondent. Respondents (or other

¹⁰ Grau, E. A., Barnett-Walker, K., Copello, E., Frechtel, P., Licata, A., Liu, B., Martin, P., & Odom, D. M. (2005). Imputation report. In *2003 National Survey on Drug Use and Health: Methodological resource book* (Section 5, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-98-9008, Deliverable No. 28, RTI/07190.574.200). Research Triangle Park, NC: RTI International.

Exhibit 5. Edit Issues Pertaining to the Health Insurance Section

Issue	Edits Implemented
The respondent (R) reported being currently covered by Medicare, Medicaid, Children's Health Insurance Program (CHIP) (if aged 12 to 19), military coverage, and private health insurance.	A flag (HLCALLFG) was provided to indicate that this pattern occurred, but no further editing was done to the data.
The R reported being currently covered by Medicare, Medicaid, military coverage, and private health insurance, the only types of current coverage that were asked about in 1999.	A flag (HLCALL99) was provided to indicate that this pattern occurred, but no further editing was done to the responses. This HLCALL99 variable was comparable to the HLCALLFG variable in 1999.
The R's only indication of current health insurance coverage came from reports of coverage by Medicaid or CHIP, but the State location variable STATELOC (corresponding to FIPE4) had been set to bad data.	Nonblank values in the variables pertaining to any period in the past 12 months when the R was without health insurance (HLCNOTYR, corresponding to question QHI13) and the number of months that the R was without health insurance in the past 12 months (HLCNOTMO, corresponding to question QHI14) were replaced with bad data codes. This edit was not done if the R indicated current coverage by Medicare, the military, or private health insurance.
The R had some indication of current coverage from at least one of the five sources of insurance listed above. However, the R also was reported to have had a period in the past 12 months when he or she was without health insurance. Further, it was reported that the R had been without health insurance for 12 of those months.	No editing was done when this pattern occurred. The rationale for not doing any editing was that the R may just recently have gotten insurance or have become qualified for insurance.
The R had no indication of current coverage from any of the five sources of insurance listed above. If the R (or proxy) answered "don't know" or "refused" when asked when the R last had coverage, the R was routed to questions about what coverage the R last had, and why the R lost health insurance coverage. That is, the skip logic assumed that the R had some prior history of coverage, but that may not necessarily have been the case.	If the R was reported to have had previously some form of health insurance or medical coverage, or if some reason was given why the R lost insurance coverage, legitimate skip codes were assigned to the variables pertaining to reasons that the R <i>never</i> had coverage. That is, the implicit assumption made in the computer-assisted interviewing (CAI) skip logic was verified by an answer indicating some prior history of health insurance coverage. However, if nothing was reported to indicate that the R previously had health insurance, the skipped variables pertaining to reasons for never having had insurance retained codes of blank.
The R was male but reported in QHI17 that he lost health insurance coverage because he "received Medicaid or medical insurance only while pregnant."	The edited variable HLLOSRSN (corresponding to QHI17) was set to bad data.

household members serving as proxies)¹¹ were asked whether they (or other family members living in the dwelling unit, if applicable) received income or benefits from specific sources, such as Social Security or from wages earned at a job or business, during the previous calendar year; these questions are subsequently referred to as "source-of-income" questions. In addition, this section of the interview included questions about the dollar amounts of total personal income and total family income (if applicable) during the previous calendar year.

¹¹ As was the case in the health insurance section, reference is made only to "respondents" in the remainder of this section. However, readers are advised that the income information for a respondent may have been provided by another adult household member who was serving as a proxy for the respondent because the proxy was considered to be better able to answer the income questions for the respondent.

Since 2008, respondents have received only a reduced set of questions on income. In 2006 and 2007, two different forms of the income questions were administered to a split sample of respondents: the "long-form" set of income questions that were administered in 2005 and earlier surveys and the reduced set of questions. The reduced set of items allows for the creation of variables related to the different sources of income for the entire family in the household. Most of these consisted of single questions about whether anyone in the family received income from a particular source; exceptions are noted below. Therefore, most of these variables do not allow for the creation of separate variables indicating whether the respondents themselves and whether the rest of their families received that particular source of income. Nevertheless, the same variable names were used in the income section in 2008 and subsequent years that were used for similar variables that existed in prior years (e.g., specific sources of household income), including those that existed in 2005 and earlier. More detailed information about the long-form income questions and their correspondence to the reduced set of income questions can be found in the documentation of procedures for editing the 2007 interviewer-administered data.¹²

The source-of-income variables were binary (i.e., yes/no). Information for most of these variables was captured through single questions about a given source of income. The exception to this single-question format was for measurement of the number of months on welfare (questions QI12AN and QI12BN and the edited variable WELMOS). Question QI12AN was asked if respondents did not report any family members' receipt of food stamps in question QI07N. Otherwise, question QI12BN was asked. In asking about the number of months on welfare, question QI12BN included the additional phrase, "not including food stamps."

Creation of WELMOS also was dependent on respondents' answers in the questions corresponding to the relevant edited variables for family members' receipt of cash assistance from a State or county welfare program (FAMPMT) and receipt of other (noncash) welfare or public assistance (FAMSVC). If FAMPMT or FAMSVC indicated that respondents received welfare payments, WELMOS was assigned the values from questions QI12AN and QI12BN. Otherwise, if family members in the household (including the respondent) definitely did not receive welfare payments or noncash benefits (e.g., job training, help with child care), WELMOS was assigned a legitimate skip code.

In addition to the source-of-income variables, binary variables were created from a pair of questions that asked whether the respondent's personal total income or the respondent's total family income was \$20,000 or more. For this pair of questions (QI20N and QI22, corresponding to edited variables PINC1 and FINC1), the second question in the pair applied to the entire family. For answering these questions about total personal or family income, respondents were shown a list of other sources of income that they had not been previously asked about (e.g., child support payments). Respondents were instructed in answering these questions about total income to consider both of these other sources of income and the sources of income they had been asked about previously.

¹² Kroutil, L. A., & Chien, C. (2009). Procedures for editing interviewer-administered data in the 2007 NSDUH computer-assisted interview. In *2007 National Survey on Drug Use and Health: Methodological resource book* (Section 10, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009.373). Research Triangle Park, NC: RTI International.

If a respondent had no other family members living in his or her dwelling unit, FINC1 was assigned a legitimate skip code. QI22 also was skipped and FINC1 was assigned a legitimate skip code if respondents reported in QI20N that they had personal incomes of \$20,000 or more. A third binary total family income variable, FAMINC1, was created and was equal to either PINC1 or FINC1, depending on whether other family members were present in the household.

Respondents also were asked to identify, both for themselves and for their families, finer categories of income within the two general categories of "\$20,000 or more" or "Less than \$20,000." If respondents answered binary total income questions as "Less than \$20,000," they were asked to report a finer category of income from \$0 up to \$20,000 by increments of \$1,000. Similarly, respondents who answered binary total income questions as "\$20,000 or more" were asked to report a finer category of income from \$20,000 up to \$50,000 in increments of \$5,000. If respondents' or their families' incomes were greater than \$50,000, they could select the following additional categories: "\$50,000 to \$74,999," "\$75,000 to \$99,999," or "\$100,000 or more." (Prior to 2004, the maximum income category was "\$75,000 or more.")

As with the binary total income questions, the finer-category income questions were asked in a pair. The first question in the pair was for the individual respondent (QI21A for personal incomes of less than \$20,000 and QI21B for personal incomes of \$20,000 or more). The second question was for the total combined family income (QI23A for family incomes of less than \$20,000 and QI23B for family incomes of \$20,000 or more). The variable PINC2 was created from responses to QI21A and QI21B, and the variable FINC2 was created from QI23A and QI23B.

Questions QI23A or QI23B were skipped if the respondent had no other family members in the household. Therefore, when the imputed household roster variable IRFAMSKP indicated that no other family members were living in the household, FINC2 was assigned a legitimate skip code of 99; that included situations in which data existed in QI23A or QI23B but IRFAMSKP subsequently indicated that no other family members lived in the household.

A third total family income variable with finer income categories, FAMINC2, also was created. If other family members were living in the household, FAMINC2 was set to be equal to the response from the mutually exclusive pair of total family income questions QI23A and QI23B. If no other family members lived in the household, FAMINC2 was set to be equal to the responses from QI21A or QI21B. Thus, unlike FINC2, FAMINC2 was not assigned legitimate skip codes if no other family members lived in the household. Since 2010, respondents who reported in question QI21B that their total personal income was \$100,000 or more were not asked question QI23B about their total family income. FAMINC2 was assigned a code of 29 (\$100,000 or more) in this situation.

Relatively little additional editing was done to the income variables, aside from assigning legitimate skip codes. The additional edits that were performed for income variables are described below.

- If STATELOC (corresponding to FIPE4; see Section 3.2.1) had been set to bad data, any nonblank values in the variable pertaining to cash assistance from a State or county welfare program (FAMPMT, corresponding to question QI08N) were set to

bad data. Because the CAI logic filled these questions with State-specific names for Temporary Assistance for Needy Families (TANF) programs, incorrect information in FIPE4 would result in these questions being filled with incorrect information for the TANF program in the States where these respondents actually lived.

- If FAMPMT had been set to bad data because STATELOC had been set to bad data and FAMSVC did not indicate welfare assistance, then nonblank values in WELMOS also were set to bad data.
- If the binary family income variable FINC1 indicated that the family income was less than \$20,000 and the corresponding personal income variable PINC1 had a missing value, the respondent's personal income (PINC1) was logically inferred to be less than \$20,000. Similarly, if the binary personal income variable PINC1 was \$20,000 or more and the family income variable FINC1 was skipped, the total family income variable FAMINC1 indicated that the family income was \$20,000 or more.
- If the finer category for total personal income (PINC2) was greater than the corresponding income category in FINC2, then FINC2 and FAMINC2 were set to bad data. For example, if PINC2 indicated that the respondent had an income of \$25,000 to \$29,999 and the respondent reported a total family income of only \$20,000 to \$24,999, then FINC2 and FAMINC2 were set to bad data.

3.2.7 Field Interviewer Debriefing Questions

The FI debriefing section was to be completed by the interviewer to obtain information about the potential quality of the interview. That included information about factors that might have affected the quality of the data, such as the degree of privacy in the interview setting. These questions were not to be read aloud to the respondent.

Only minimal processing was done to the data in this section. Specifically, raw variables were replaced with final, mnemonic variable names (e.g., PRIVACY for the variable pertaining to the interviewer's indication of how private the interview was). Where relevant, variables also were assigned legitimate skip codes based on the routing logic in this section.