TESTING ACASI PROCEDURES TO REDUCE INCONSISTENCIES IN RESPONDENT REPORTS IN THE NHSDA: RESULTS FROM THE 1997 EXPERIMENTAL FIELD TEST

Rachel A. Caspar, Judith T. Lessler, and Michael A. Penne, Research Triangle Institute
Rachel A. Caspar, RTI, P.O. Box 12194, Research Triangle Park, NC 27709-2194

Keywords: National Household Survey on Drug Abuse, computer-assisted interviewing, data editing, cognitive laboratory testing

1. Introduction

One of the most significant potential benefits of converting the National Household Survey on Drug Abuse (NHSDA) to a computer-assisted format is the opportunity to resolve inconsistent data at the time of the interview rather than editing the data to deal with inconsistencies after the fact. However, maintaining the privacy benefits of the ACASI component of the interview requires that the respondent be able to resolve inconsistencies for many items on his or her own. Thus, one of the goals of the NHSDA conversion work was to develop a method for resolving inconsistent data that the respondent could easily understand and complete without significant intervention by the interviewer.

Based on our own hypotheses about how inconsistencies should be identified and resolved, we developed a resolution methodology that combined two components. First, at the **verify** stage respondents are asked whether an answer they have entered is in fact correct. So, for example, when a 20-year old respondent reports that she was 51 the first time she drank alcohol (a clearly inconsistent answer), the computer was programmed to verify that this information was correct. If the respondent indicated that the information was incorrect, he or she was routed back to answer the question again (perhaps this time entering the age of her first drink as 15). A second component incorporates the **resolution** of seemingly inconsistent answers. For example, a respondent who indicated drinking alcohol on 15 days in the past 12 months, but then reported drinking alcohol on 25 days in the past 30 days would first be asked to verify the last entry keyed. If the respondent indicated that the entry was correct, then he or she was routed to a question that identified the inconsistency and provided the respondent with an opportunity to correct one or both of the entries.

In developing the actual text of these verification and resolution screens, we sought to incorporate several important features:

- the original responses that were identified as inconsistent are displayed onscreen for the respondent to enhance recall and comprehension
- the resolution screens are worded so as not to explicitly place the responsibility for the inconsistency on the respondent (e.g., "your answers"), but rather imply that the computer could be incorrect (e.g., "the computer recorded")
- respondents are asked to identify the **incorrect** response when two items are inconsistent to facilitate the flow of questioning
- when the respondent indicates that both answers are incorrect, the respondent is routed back to the two items in the same order they were presented the first time in order to maintain a consistent flow through the instrument
- respondents are explicitly notified when their answers are inconsistent rather than attempting to resolve the inconsistency indirectly without actually making the respondents aware of the problem
- inconsistencies are only identified between items that appear close to each other in the interview to minimize respondent confusion.

2. Cognitive Laboratory Testing

We first tested our resolution methodology in the cognitive laboratory. In order to maximize the efficiency of our laboratory testing, we felt we could not rely on respondents to give inconsistent answers in the laboratory testing. Even with an extremely large sample, the number of respondents who will provide inconsistent answers is quite small. Since only 40 respondents were recruited for this task, it seemed entirely possible that none of the respondents who came in to the laboratory would provide an inconsistent response. For this reason, we developed a laboratory task that incorporated the use of vignettes. The vignettes we used were essentially brief descriptions of a person and his or her drinking behavior. The laboratory subject was instructed to answer a series of questions about drinking alcohol as though he or she was the person in the vignette. The subject read the vignette

and then began answering the questions as they appeared on the computer screen. At a specific point in the questioning process, the subject was instructed to obtain additional information from the interviewer to be able to continue answering the questions. This additional information resulted in the subject providing inconsistent answers which then he or she was required to verify or resolve.

The vignette methodology is somewhat artificial. That is, subjects are not answering the questions based on their own experiences and thus we did not learn anything about why people may give inconsistent answers (e.g., poorly worded questions, difficulty in recalling the information, desire to conceal information, etc.). However, our primary objective in this testing was to determine whether respondents could easily navigate through the verification and resolution process without becoming either confused or annoyed.

Our initial round of testing indicated one very pronounced problem. Respondents found it very confusing to be asked which of their answers was **incorrect** when two answers were identified as inconsistent. Respondents reported that it was much more logical to be asked which answer was **correct**. Even when the researcher pointed out that the reason for the question was to determine which question to re-ask, the subjects were nearly unanimous in their preference for identifying the correct answer.

Our initial testing also pointed to some problems with the vignette task itself. A number of respondents had difficulties figuring out which information in the vignette was applicable to which survey question. In our effort to make the vignettes seem more "realistic," we had added information not specifically needed to complete the set of questions asked. In order to reduce the confusion caused by the vignettes, we scaled back the amount of information provided in each vignette to only that which was needed to complete the task.

Using the revised vignettes and the resolution process which asked the respondent to indicate which of his/her answers was correct, we began testing a second round of subjects. The second round of subjects seemed to find the resolution task much easier than subjects in the first round had found it. Subjects were able to easily select which of the answers was correct during the resolution process and understood that they were being routed back to the incorrect item to make the necessary correction. In general, respondents were not put off by the verification and resolution process. Some respondents went so far as to note that they would appreciate the computer pointing out inconsistencies in their responses. Although not unanimous, most respondents preferred the less direct wording of, "the computer recorded...." to the more direct wording that would say, "you reported that..." Respondents with little computer experience seemed to prefer this wording because they believed that the computer could make mistakes in how entries were stored. Respondents with greater computer literacy recognized that recording errors are made by the respondent and not the computer. However, the majority of these respondents still felt the less direct wording would be less confrontational and less likely to embarrass the respondent.

Verification screens were reported to be easier to complete than the resolution questions. Comments made by the subjects indicated that the verification screens were short and to the point. Resolution screens were reported to be "too wordy." Respondents indicated that there was so much text to read that it was easy to get confused. Therefore, for our second round of testing, we reduced the amount of text in the resolution questions. For example, in round two the resolution screen that identified inconsistencies between 12 month frequency and 30 day frequency was:

> The computer compared the answers for the last question and an earlier question. According to the answers it recorded, you drank one or more alcoholic beverages on more days in the past 30 days than in the past 12 months. This is not possible. Which of the following is correct?
>
> I drank alcohol on [XX] days in the past 12 months
> I drank alcohol on [YY] days in the past 30 days
> Neither answer is correct

Subjects noted that the scripting in the body of the question was repeated again in the categories. This redundancy was viewed as unnecessary and sometimes confusing.

Prior to conducting the third (and last) round of laboratory interviews, we made some additional changes to resolution screens to reduce the amount of text. Scripting was reduced again, and the revised version of the question shown above was:

> The answers for the last question and an earlier question disagree. Which answer is correct?
>
> I drank alcohol on [XX] days in the past 12 months
> I drank alcohol on [YY] days in the past 30 days
> Neither answer is correct

The revised wording seemed to work well in the laboratory. Subjects were still able to complete the resolution task with little trouble, and we did not have as many complaints about the wordiness of the screens.

Based on these three rounds of testing, we developed a method for resolving inconsistent responses in the NHSDA to be tested in the 1997 CAI Field Experiment.

3. 1997 CAI Field Experiment

As one component of the 1997 CAI Field Experiment, we examined the potential for improving data quality by having a random half of the respondents resolve inconsistent and questionable data during the interview. For the random half receiving the data quality checks, the ACASI program included edit checks that followed up on inconsistent answers provided by the respondent. The respondents were asked a set of questions designed to resolve the inconsistency. The design of these questions was described above. In addition, checks of questionable reports, such as, a suspiciously low age of first use for a substance (e.g., 2 years old) were checked to ensure, for example, that a response of "2" was not actually supposed to be "22." The inconsistency checks included the following:

- 30 day frequency of use greater than 12 month frequency of use for cigarettes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, and inhalants.
- Zero days used in past 30 days for persons reporting the past 30 days as their time of most recent use for cigarettes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, and inhalants
- Age at first use greater than current age for cigarettes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, and inhalants
- 12 month frequency of being very high or drunk greater than 12 month frequency of use for alcohol.
- Number of days consumed 5 or more drinks on the same occasion greater than 30 day frequency of use for alcohol
- Last use of LSD more recent than last use of any hallucinogen
- Last use of PCP more recent than last use of any hallucinogen

The checks of questionable reports included:

- Age at first use is suspiciously low for cigarettes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, and inhalants
- Age at first use equal to current age for cigarettes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, and inhalants

Clearly it is preferable to have respondents correct any inconsistencies in their data rather than having an analyst determine how to edit the data after the fact. In addition, while considerable effort must be expended to program these data quality checks, they have the potential to reduce the post-survey processing by reducing the number of edits. However, we were uncertain as to whether respondents would be able or willing to provide this type of information and speculated that it could increase either the number of breakoffs and or the overall length of the interview.

4. Results from the 1997 CAI Field Test

As noted earlier, inconsistency checks programmed for the inconsistency resolution treatment can be divided into two types. First, there are true inconsistencies in which a respondent's answers to each of two answers cannot both be true. For example, a respondent who has reported her current age as 22 cannot logically report her first use of marijuana at 35. Similarly, a respondent who reports drinking alcohol on 22 days during the past 30 days cannot logically report drinking alcohol on 3 days in the past 12 months. A second type of check that was programmed can be considered verify checks as opposed to true inconsistencies. In these cases, a respondent's answer is not technically inconsistent with previous information reported, but rather is simply "unusual" given what we have learned from the data collected in previous NHSDA studies. The two verify checks included in 1997 CAI Field Experiment prompt respondents to verify the accuracy their response when they indicate an age at first use as equal to their current age or an age at first use less than ten. In discussing the outcome of the inconsistency resolution treatment we first consider the two types of inconsistency checks separately since they may reflect different effects on data quality.

Exhibit 1 shows the number of respondents in each of the four inconsistency resolution treatments who triggered at least one inconsistency check during the course of the interview. From this Exhibit we see that

approximately 28 percent of the respondents assigned to an inconsistency resolution interview triggered at least one check item. Of those respondents just over half (55%) triggered a verify check as opposed to a true inconsistency item. More respondents triggered a true inconsistency check in Treatments 2 and 4. This may be the result of additional confusion created by the multiple use treatment. However, it is also possible that it only reflects the fact that there were more inconsistency checks programmed into these two treatments. Interestingly, the number of respondents who triggered verify items was higher for the two treatments that utilized the single gate question. Perhaps the three question approach does a better job of minimizing the reporting error that we believe goes on in the PAPI interview in which respondents mistakenly report their current age when the question is actually asking for their age at first use for a specific drug.

Across all treatments, youth respondents (ages 12 - 17) accounted for approximately 60 percent of the respondents who triggered a true inconsistency. Within each treatment, youth respondents account for more than half the respondents who triggered a true inconsistency. In Treatment 8 the difference was especially pronounced; youth accounted for 70 percent of the respondents who triggered an inconsistency in this treatment.

In addition to determining the number of respondents who triggered any of the inconsistency items, we also calculated the total number of inconsistency checks triggered overall. The average number of true inconsistency checks triggered was approximately 1.2 per respondent for a total of 150 true inconsistencies. The average number of verify checks triggered is approximately 1.2 per respondent as well.

Of particular interest is whether the data coming out of the inconsistency checks is "cleaner" than the data going in. For 1997 CAI Field Experiment we made the decision to route respondents through each inconsistency check only once. The program was not designed to route respondents back through an inconsistency check a second time if their answers were still inconsistent. We were concerned that multiple passes through a single inconsistency check might be overly burdensome to respondents. However, if the data coming out of the consistency checks is no more consistent than it was prior to the check, it would either be necessary to route respondents through the checks multiple times or simply consider the consistency checks an ineffective method for improving data quality and delete them from the CAI program entirely.

Exhibit 2 provides data on the status of data consistency following a consistency check. Each resolution process was examined and classified into one of four types. First, the data could be classified as consistent; that is, after being notified of the inconsistency the respondent revised one or both of his/her answers such that the answers did not conflict with each other. Second, the data could remain inconsistent, meaning that even after changing one or both of the answers the respondent's answers still conflict with one another. Some responses were classified as "indeterminate" because the nature of the resolution process did not require the respondent to fully resolve the inconsistency. For example, in Treatment 8, a respondent may report that he used marijuana during the past 30 days but when asked the 30 day frequency item he reports zero days. He is routed to a resolution screen that asks him to verify that he did not use marijuana in the past 30 days. If he indicates this is correct, that is the end of the resolution process even though he is never presented with the two conflicting items. While it seems safe to consider this respondent a non-user for the 30 day period, for the sake of clarity, we have classified these cases separately in Exhibit 2. Finally, there are a small number of situations where the respondent recorded a Don't Know or Refused response when presented with the resolution screen. In these situations, the program routed the respondent out of the resolution series and thus the case cannot be classified into any of the other three categories.

From Exhibit 2 we see that 30 percent of all inconsistencies were resolved as consistent; 17 percent were still inconsistent after the resolution process, and just over half (51%) were indeterminate. If we assume that the indeterminate cases can be considered more accurate data than the inconsistent data originally reported, then the quality of 81 percent of the inconsistent responses has been improved through the resolution process.

We hypothesized that respondents who were overly confused by the resolution process would enter a Don't Know or Refused response as a method for avoiding the resolution task. We found only a small number of the inconsistencies were "resolved" by the respondent in this manner (a total of four resolutions across all treatment versions). We take this as a positive sign that the resolution process was not overly confusing for respondents and believe that overall the improvement in data consistency is an improvement to the overall quality of the NHSDA data as well. However, before we finalized our decision to include inconsistency resolution

in future NHSDA data collection, we examined several operational factors.

First, we compared the time to complete the experimental portion of the ACASI interview by whether even one inconsistency was resolved during the respondent's interview. We found an increase of approximately three minutes when at least one inconsistency resolution is triggered compared to when none were triggered.

We next compared the rate of incomplete interviews (commonly known as "breakoffs) by whether the inconsistency resolution treatment was in use. Breakoff cases accounted for only a small number of the total number of interviews completed for 1997 CAI Field Experiment. However, if the inconsistency resolution treatment caused respondents not to complete the full interview, we would view this as serious drawback of the methodology. We analyzed the breakoff data in two ways. First, we compared the percent of finalized breakoff cases by resolution status. We also compared the percent of cases that were **ever** reported as a breakoff (even if the interviewer went on to finalize the case as a completed interview) by resolution status. If the resolution methodology resulted in an increased number of cases that required the interviewer to return to the household to finish an incomplete interview, the cost per completed interview would increase even if the overall response rate remained unchanged. This would be viewed as an undesirable effect of the resolution methodology. In each case, our analyses showed no impact of the inconsistency resolution treatment on the breakoff rate for the 1997 CAI Field Experiment.

Finally, for respondents who triggered at least one inconsistency item (a true inconsistency rather than a only a verify item), we asked the interviewer to provide any explanation for why the respondent might have provided inconsistent data. A review of these open-ended responses indicated the most common reason provided by the interviewers was that there were too many distractions during the interview. These distractions included the presence of young children, the presence of other adults, the telephone ringing, the doorbell ringing, and the television playing. Other explanations mentioned by the interviewers included: respondent boredom, literacy problems, respondent fatigue, and respondents who simply weren't paying close attention to what they were doing.

5. Conclusions

Based on the results reported here, we feel the inconsistency resolution methodology employed in the 1997 CAI Field Experiment was successful. The methodology improved the consistency of the data collected without adversely affecting respondent cooperation or burden. Using this methodology in future implementations of the NHSDA will allow SAMHSA to capitalize on the numerous benefits of the ACASI technology while minimizing one of the potential pitfalls; that respondent errors and inconsistencies are not identified and corrected at the time of interview. Currently work is underway to incorporate inconsistency resolution screens into the 1999 CAPI/ACASI NHSDA instrument.

We anticipate that future research in this area will be conducted to determine whether respondents can resolve inconsistencies between items that are not closely spaced in the interview. If this proves to be possible, future NHSDA instruments may incorporate inconsistency resolution screens of this type as well.

Exhibit 1 Number of and Percent of Respondents Who Triggered at Least One Inconsistency

	Consistency Checks Present				
	Single Gate Question		Multiple Gate Questions		Total
	Multiple Use Questions		Multiple Use Questions		
	Absent	Present	Absent	Present	
Treatment Version	3	4	7	8	
Number of Respondents	285	264	219	207	975
Total Number of Respondents Who Triggered an Inconsistency	75	82	52	68	277
Percent of Respondents	26.3	31.1	23.7	32.9	28.4
True Inconsistency	27	39	21	37	124
Percent of Respondents	9.5	14.8	9.6	17.9	12.7
Percent of Inconsistencies	36.0	47.6	40.3	54.4	44.8
Verify Only	48	43	31	31	153
Percent of Respondents	16.8	16.3	14.2	15.0	15.7
Percent of Inconsistencies	64.0	52.4	59.6	45.6	55.2

Source: National Household Survey on Drug Abuse: Development of Computer Assisted Interviewing Procedures; 1997 Experimental Field Test

Exhibit 2 Data Consistency Following the Resolution Process[1]

	Consistency Checks Present				
	Single Gate Question		Multiple Gate Questions		Total
	Multiple Use Questions		Multiple Use Questions		
	Absent	Present	Absent	Present	
Treatment Version	3	4	7	8	
Number of Respondents	285	264	219	207	975
Total Number of True Inconsistencies Triggered	31	50	24	45	150
Data Consistent	15	21	4	5	45
Percent of Inconsistencies	48.4	42.0	16.7	11.1	30.0
Data Inconsistent	5	6	7	7	25
Percent of Inconsistencies	16.1	12.0	29.1	15.6	16.7
Indeterminate	8	23	13	32	76
Percent of Inconsistencies	25.8	46.0	54.2	71.1	50.7
DK/REF	3	0	0	1	4
Percent of Inconsistencies	9.7	-	-	2.2	2.7

Source: National Household Survey on Drug Abuse: Development of Computer Assisted Interviewing Procedures; 1997 Experimental Field Test

[1]Some respondents triggered more than one inconsistency