

## **7. QUALITY CONTROL**

Through all phases of the 2001 National Survey of Veterans (NSV 2001), Westat paid strict attention to the accuracy of our work. Because we integrated quality control procedures into every aspect of the research process, we included descriptions of these procedures in our discussions of each project component in the previous chapters of this report. This chapter briefly presents the measures taken primarily to enhance quality. It also presents detailed information about two key quality control measures – the NSV 2001 pretest and managing the database.

### **Designing and Programming the Questionnaire**

The NSV 2001 questionnaire went through several design drafts and a number of iterations for the computer-assisted telephone interviewing system (CATI) program specifications. To assure that design decisions accurately reflected study goals, we kept a detailed log of questionnaire revisions. Westat and VA project staff thoroughly reviewed each draft to assess its effectiveness in addressing NSV 2001 data requirements.

The decision to use CATI was guided by the need for quality control. As noted in Chapter 5 (Data Collection), using a CATI system to conduct telephone interviews facilitates proper administration of the questionnaire. However, this is the case only if we develop and correctly implement the questionnaire specifications. We took several steps to make certain that the specifications were accurate. Questionnaire designers and CATI programmers worked together to develop the specifications for programming the CATI instrument. By using this team approach, we ensured that the CATI instrument accurately reflected the NSV 2001 questionnaire. Once the instrument was programmed, the team put it through numerous rounds of testing. Testing covered every aspect of the CATI program, including screen displays, transitions from one section to the next, skip patterns, data storage, sample management, result codes, randomization routines, and delivery of cases to interviewers. As part of the testing process, we documented all potential problems and had at least two project staff members verify that each was resolved. Finally, we maintained a history of all changes.

## **Training and Supervising Interviewers**

The NSV 2001 interviewers were central in our effort to collect accurate information, and to meet our data collection goal of 20,000 completed interviews. As a company, Westat spends a great deal of time honing our training techniques. Two important aspects of training are the high level of interviewer involvement and the consistent manner in which the training exercises are administered. Trainee involvement is critical because it allows trainers to observe and evaluate individual performance. Trainers can then decide early on who should be released from the program because of failure to meet Westat performance standards. Even more important, trainers can identify specific areas of difficulty, and tailor the program accordingly. Scripted exercises and role plays allow trainers to maintain a high level of consistency across multiple training sessions (there were eleven such sessions for the NSV 2001). Scripting the mock interviews in advance also ensured that interviewers had the opportunity to practice responding to various scenarios they were likely to encounter during administration of the NSV 2001, better equipping them to overcome initial respondent refusals.

During data collection, the primary method of ensuring that interviewers continued to accurately administer the NSV 2001 questionnaire was through our monitoring sessions. Project staff and telephone center supervisors monitored an average of 8 percent of all NSV 2001 interviews. We assessed interviewers' administration of the questionnaire, recording of responses, probing, handling of contacts, and professional demeanor. Monitors reported the results of each session on a monitoring form, then shared them with interviewers (who were unaware they were being monitored until after the session was over). Interviewers were apprised of their strengths as well as areas needing improvement. If needed, we provided additional training, or made adjustments to the interviewing process in general. In conjunction with the monitoring sessions, we used summary reports of the number of call attempts, successful contacts, refusals, and completions to assess interviewer performance. Based on these reports, we were able to identify interviewers with skills that could be matched to operational areas specific to the NSV 2001, such as proxy interviewing, tracing calls, refusal conversion, and language problem cases. Finding the best interviewers for these particular tasks ensured that they were carried out with the highest levels of accuracy and skill.

## **Sampling**

We prepared several sample design approaches and analyzed these before the final design was chosen. Our detailed analyses assessed coverage of the veteran population and

subgroups of interest, precision, cost, and operational demands. We developed detailed specifications for constructing the list and RDD frames to assure their accuracy and statistical validity. To minimize bias from time effects, we created sample release groups and made each one available for calling at separate intervals throughout the data collection period. As a means of tracking the List and RDD Samples separately, we developed a system of IDs that differentiated cases from each sample. Finally, we checked all steps in the sample file creation and other sample processing stages of data collection by designing and producing frequencies and cross-tabulations of appropriate variables.

### **NSV 2001 Pretest**

To determine if, under live interviewing conditions, our data collection procedures would work as we had anticipated, we conducted a pretest of the NSV 2001. The objectives of the pretest were to:

- Test item wording, survey flow, skip patterns, ease of administration, respondent comprehension of the survey questions, and other factors related to the instrument itself;
- Check that all functions of the CATI system, including data storage, call scheduling and case management, were working properly;
- Establish the average length of time it took to administer the instrument;
- Evaluate our training procedures; and
- Solicit feedback from interviewers about all aspects of the interviewing process.

An additional benefit of the NSV 2001 pretest was that it afforded the VA an excellent opportunity to observe the methodologies and procedures planned for the main data collection phase.

The NSV 2001 pretest was conducted at one Telephone Research Center between February 12, 2001 and March 4, 2001. During that period, the List Sample contact procedures were still in development, so the pretest was administered using RDD Sample cases only. The entire pretest effort was based on an initial RDD Sample of 60,000 cases, all of which were loaded into the CATI system. Of the 60,000 telephone numbers, 17,616 were eliminated from calling because they were either business or nonworking telephone numbers. Therefore, 42,384 telephone numbers remained available for dialing during the pretest.

Pretest interviewers called 21,609 telephone numbers. During the screening portion of the telephone interviews, interviewers completed 2,928 screening questionnaires and identified 901 potential veterans in 852 households as eligible to participate in the NSV 2001. This rate of 1.06 veterans per eligible household varied little throughout the entire data collection effort. At the conclusion of the pretest, interviewers had completed 519 extended interviews. Figure 7-1 is a flowchart that summarizes the magnitude of the pretest sample and workload, as well as the outcome of calls during the pretest. On March 16, 2001, Westat briefed the VA on these pretest call results.

The pretest revealed that the CATI instrument worked as we intended it to. Nor did we discover any problems with the CATI system's call scheduling, case management, or data storage functions. We did, however, modify our yield assumptions to reflect the actual completion rates. (See Chapter 8 for a more detailed discussion of completion rates.). We also learned from the pretest that the average length of the interview was slightly over the target of 30 minutes. Finally, we revised our training program to increase the focus on one area that presented difficulties for interviewers and respondents in the pretest – correctly identifying current household members.

### **Managing the Database**

Database managers for the NSV 2001 had two key responsibilities: ensuring that the data collected were consistent, valid, and within the specified ranges, and ensuring that result codes<sup>1</sup> accurately reflected the status of cases. In the course of this work, database managers also ensured that the CATI instrument was operating correctly, that the responses were being recorded correctly in the database, and that interviewers were administering the questions and using the response categories correctly. Database managers at Westat employed several methods to carry out these responsibilities. We programmed an automated data check, manually reviewed question by question response frequency

---

<sup>1</sup> A result code is the disposition that reflects the case's entire call history, as well as the status of that case at the time the last contact was made.

## NSV 2001 Pretest (RDD Only)

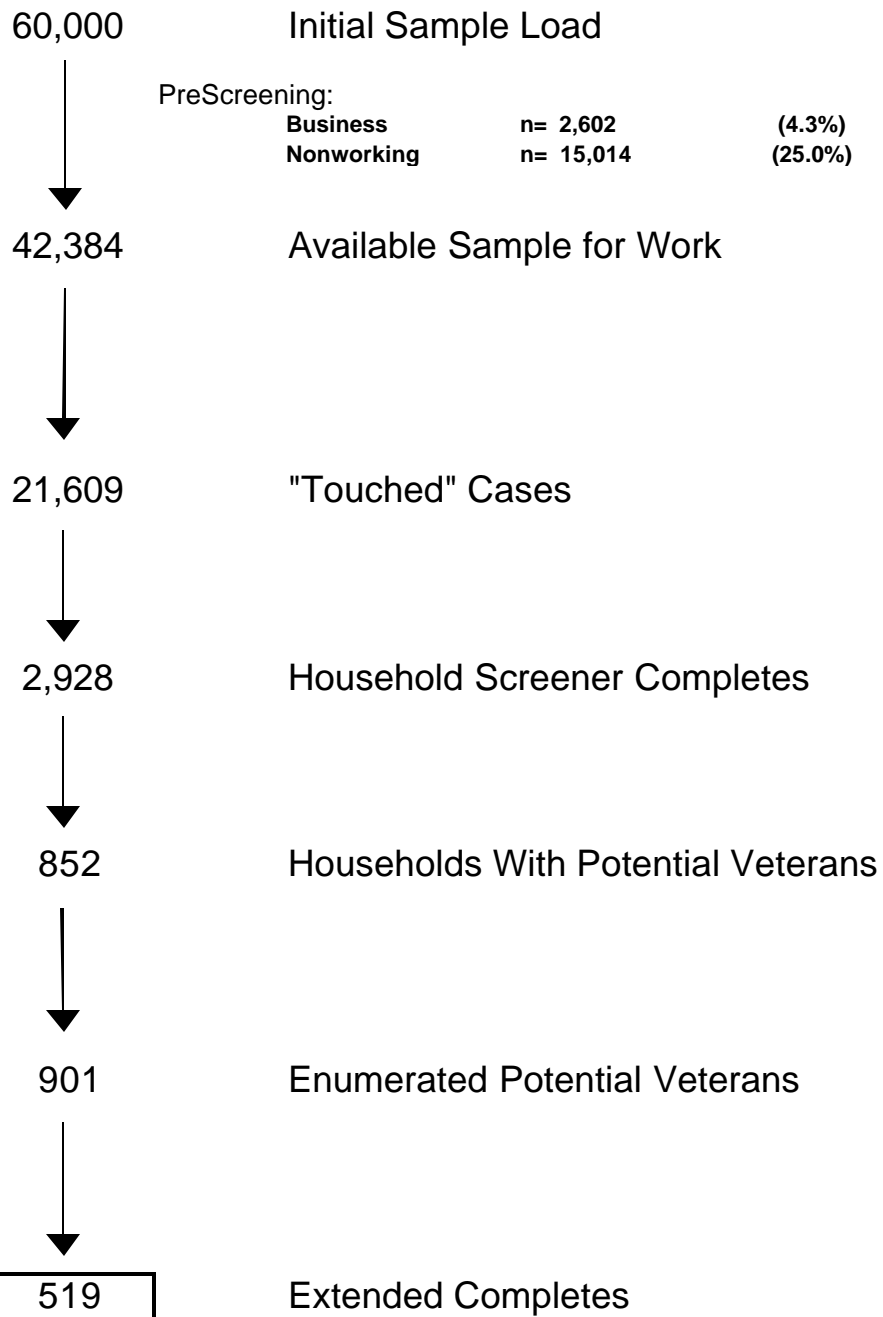


Figure 7-1. Flowchart of NSV 2001 pretest sample activities

reports, reviewed and recoded open-ended responses as necessary, and reviewed interviewer comments. This section describes each of these methods in turn, along with the additional steps we took to prepare the database for delivery after the close of data collection.

It is important to note that we did not impute missing data during this phase of the project. We did, however, impute a few variables that were required for the weighting effort. (See Chapter 6 for details about data weighting for the NSV 2001.)

**Automated Data Checking.** To reduce survey administration time, we programmed item MB18 to automatically determine and store the veterans' service eras based on their responses to questions about when they began and were released from active duty. The programming logic for this item was complicated, using dozens of "if-then" statements. Because it would have been difficult to identify errors by doing a manual review of this item, not to mention unreasonably time consuming, we created a SAS<sup>®</sup> program that evaluated the various sets of variables defining each respondent's entry and exit dates from the military, and used that information to derive a corresponding date range. This date range was then compared to the preestablished date range of each service era to determine whether the respondent was in the military at that time. Finally, the program compared its findings to the service era flags that were set automatically by the CATI system and noted any discrepancies. We ran this program every 3 to 4 weeks during data collection. On those rare occasions when errors were found, they were easily corrected because they were simply the result of one or two variables being missed during a manual update.

**Review of Question by Question Response Frequency Reports.** Several times throughout data collection we created a question by question response frequency report. Using this report, we checked that responses fell within the allowable range for each question, verified that the relationships of responses across individual questionnaires were logical, and ensured that the data accurately reflected the instrument's skip patterns. We produced this report after the first 100 cases were completed, at the end of the pretest, about halfway through data collection, and after data collection closed.

The set of questions that asked veterans about the dates they were on active duty comprised a complicated, and somewhat lengthy, series of items. Veteran responses to those questions then determined which categories would be displayed in survey item MB20 ("Now I'm going to read a list of places. Please tell me if you served in, sailed in, or flew missions over each of the following while on active duty."). To check that the proper categories were displayed, and

that data were correctly recorded at MB20, we had to rely on more than just the frequency report. We generated detailed cross-tabulation reports that allowed us to examine the logical relationships among these items. Any discrepancies we found in the expected relationships were corrected by updating the affected variables.

The frequency report allowed us to conduct range checks of every variable. The range checks compared the specified ranges in the CATI data dictionary with the responses entered during interviewing or derived by the CATI programs. The few out-of-range responses we found were verified as representing valid answers that happened to be outside the range we had anticipated.

In a few instances, our review of the frequency report revealed the need to update the CATI program. In the first instance, responses entered at the Gender verification screen as “F” for “female” rather than “2” (the usual code for “female”) resulted in eight cases in which the respondents were not asked whether they served in branches reserved for women (survey items MB23f – MB23l). The variables that should have contained valid responses in those eight cases were set to the value “Not Ascertained.”

We also updated survey item SD19 (“What is the market value of your primary residence?”) to allow a value as low as zero, made minor wording changes to survey item SD13 (“During the year 2000, how many children depended on you for at least half of their support?”) and survey item SD17 (“Excluding your primary residence, what is the total amount of assets your family owns?”), and added a category to survey item BB3 (“What do you want done with your ashes?”). At the request of the VA, we changed survey items SD10 and SCD10 (“I am going to read you a list of racial categories. Please select one or more to describe your race.”) to ensure that it matched the approach to collecting race items used by the U.S. Bureau of the Census.

**Recoding of “OTHER (SPECIFY)”**. Because all response data were written to the CATI database directly from the interviewer data entry function, and the instrument included no open-ended questions that required post-coding, the NSV 2001 data posed no significant coding requirements. However, 24 closed-ended questions gave veterans an opportunity to add an “OTHER (SPECIFY)” response when they felt that the precoded response categories did not adequately cover their situation. We reviewed the “OTHER (SPECIFY)” responses weekly for the first 6 weeks of the survey period. In subsequent weeks we reviewed them less frequently. When data collection was complete, we conducted a final review of all “OTHER (SPECIFY)” responses. We checked whether the “OTHER (SPECIFY)” responses duplicated a precoded

response already given by the veteran, provided amplified or more specific information about a precoded response already given by the veteran, or duplicated a precoded response that had not been given by the veteran. Only in the last instance did we reset the response category to reflect the additional information. Overall, very few of the “OTHER (SPECIFY)” responses required that we reset them to precoded response categories.

**Review of Interviewer Comments.** NSV 2001 interviewers had the opportunity to record their assessments of irregular cases in four places. Interviewers could type their reports directly into the CATI Comments, Non-Interview Report, or Message Files. Or, they could fill out a paper problem sheet. Sometimes, their remarks qualified or, in some cases altered, a veteran’s responses in some way that did not fit the categorical values of the survey variables. Other times, these remarks had more general ramifications that affected the handling of the case or interpretation of the data.

**Comments File.** The Westat CATI system permits an interviewer to record a comment at any point in the interview. The comment is automatically linked to the specific case and item at which it was recorded. Interviewers used this function to record veteran comments that clarify or correct information entered into the questionnaire. On a daily basis, we reviewed Comments File output for potential changes or updates to the data. We permanently retained all comments in the file.

The majority of the comments were simply the veteran’s elaboration on answers that had been properly recorded, and therefore required no further action. However, in some cases the comments clearly affected responses in ways that required data updates. For example, as veterans cycled through questions, a few realized they had not understood the original question and had replied incorrectly. When this occurred at a point in the interview where it was impractical to back up and change the original response, interviewers would make a comment to change the previous response. The database managers would then correct the original response and, where appropriate, update all related responses.

**Non-Interview Report File (NIRF).** For all refusal cases, language or hearing problems, or other special problem cases, interviewers recorded their assessment of the situation in the NIRF. While it is primarily interviewers who use the NIRF, it is an additional source of information for database managers and project management staff to use when deciding how to handle cases that require attention beyond that of the telephone supervisory staff. For the NSV 2001, we periodically checked the NIRF, but seldom had occasion to use it for this purpose.



**Message File.** At the end of every telephone contact, the Message File gives interviewers one more opportunity to record their observations about cases that require special attention. Interviewers and supervisors used this file to communicate more detailed information about cases to other interviewers, database managers, and project management. Database managers used the file to interpret inconsistent responses and pose general questions about the cases. As with the NIRF, we checked this file regularly but rarely had to use it.

**Paper Problem Sheets.** Interviewers and telephone center staff used paper problem sheets to record data entry errors and request updates to specific cases. Problem sheets were reviewed daily and handled by database managers in the same way as the electronic Comments File. Interviewers also used problem sheets to note instances where they thought the CATI system had not functioned as they expected. These notes were checked daily and resolved by database managers and programming staff as needed.

**Final Review of the Database.** After the close of data collection, we conducted a final, comprehensive review of the entire database of completed interviews. We again ran the MB18 automated data check, the detailed cross-tabulations, and the question by question response frequency report described earlier.

In addition, we verified that every final result code had been assigned according to project definitions. (See Appendix H for a list of NSV 2001 final result codes.) We did this through a series of cross-tabulations that examined sample type and responses to specific survey questions. The majority of result codes were correct. We changed a few extended interview result codes to reflect that a proxy rather than the sampled veteran had completed the interview.

## **Weighting**

Statistical staff prepared detailed specifications for developing the weighting systems for the List and RDD Samples. To validate the implementation of these specifications, the systems staff and the statistical staff independently checked the computer codes used to calculate and assign weights. We checked the calculation of nonresponse adjustment factors cell by cell. We also verified the calculation of nonresponse adjusted weights, in part by examining the variation in the nonresponse adjustment factors and the nonresponse adjusted weights. For the key veteran population subgroups, we designed, developed and checked cross-tabulations of

appropriate variables during each step of the weighting process. Additionally, we compared the weighted total from the RDD Sample to the estimate from the VA population model, Vetpop2000, to ensure consistency between the two. For construction of the composite weights for the combined List and RDD Samples, we prepared detailed specifications and checked their implementation by reviewing the computer code used to calculate and assign the composite weights. To ensure that the weight file created for delivery to the VA was without error, we produced several summary statistics from it and compared those statistics to the summary statistics produced from the original Westat file.