MAIN SURVEY INSTRUMENT

BUILDING THE INSTRUMENT

We began construction of the survey instrument with the acknowledgment that loss of memory over time would be an issue of consideration, particularly in light of the fact that data collection would be conducted approximately seven years after the end of Gulf War. Of greatest concern was recall bias,[1] which can have a significant effect on the integrity of collected data in a retrospective study relying on self-reported information (such as self-reports of pesticide use and exposure). Our first course of action, therefore, was to review and summarize existing recall bias literature to learn what methodologies had been developed or used previously to mitigate this effect.

Using this review, we assembled a list of methods that have been shown to reduce recall bias and used them for guidance in the development of the instrument:

- The use of data-fixed events as landmarks to help place queried events accurately;

- The use of cues to increase retrieval time and force respondents to search their experiences extensively, rather than form answers based on their current evaluation of health status (e.g., write longer introductions and include explanatory material);

- A mail-out of visual aids with letters in advance of phone interviews;

- Inclusion of context-setting introductions before questioning;

---

[1]Recall bias is defined as the error that appears when people cannot remember whether or when certain events occurred, resulting in either an omission of events or their incorrect placement in time.

- Anticipation of autobiographical sequences that may have surrounded pesticide exposure, using these events to develop cues to trigger recall of the sequence as a whole; and

- Scenarios built into the instrument that ask respondents to imagine the details of their life at the time, helping to draw out context-dependent memories.

Not all previous findings regarding cues to aid recall were supported by our pretests. One such item suggested that questions examining series events be developed in a transversal manner rather than chronologically, to avoid date error correlation that was found to occur in longitudinally structured recall. In the pretests originally designed under this organizational style, however, respondents continually strayed from the path of questioning and proceeded to recount events in chronological order, often becoming frustrated because the questions and sections themselves did not flow in this manner.

In addition, the instrument was originally designed to sample based on respondents' random locations during their tour, rather than on a random month in their tour as the final version now reflects. Sampling according to random location became a problem because it was not possible to reliably link individuals to locations. Further, it was difficult to query individuals about locations because of problems describing and categorizing locations, the high mobility of some units, and the different lengths of time respondents were stationed in an area.

The other primary change made to the instrument because of feedback from the pretest rounds was a reorganization of how certain data were collected and at what point in the survey. Originally, respondents were asked first what types of pesticides they saw others use, then what they personally used—all organized around location scenarios (sleeping tent, mess hall latrines, etc.). However, we found two problems with this format: (1) Veterans' memories of what pesticides others used were often triggered by recall of their own use of and exposure to pesticides; and (2) respondents' natural inclination was to first list all the places around their base that they used pesticides, going back later to fill in the details of how and what pesticides were used. Once we reorganized the questions, the interviews progressed much more smoothly. The data were collected as they most naturally flowed from the respondent, reducing respondent burden and frustration with the process.

ITERATIVE QUESTIONNAIRE PRETESTS

Once the initial draft instrument was ready, we conducted multiple pretests over a period of 10 months to refine and test the draft questionnaire. We conducted rigorous, iterative pretesting of the survey instrument with respondent

groups of fewer than 10 individuals at a time and revised the instruments accordingly between each pretest administration, starting in July 1997 and continuing through April 1998. The testing was carried out in two major phases, first testing the paper-and-pencil versions of the questionnaire and then the computerized version of the instrument. Altogether we completed pretest interviews with approximately 100 Gulf War veterans. Respondents of different age, gender, race, military rank, and military branch were included in the pretesting efforts. The test groups were chosen to be representative not in a statistical sense but rather in the range of experience that the main survey respondents were expected to demonstrate.

RECRUITMENT PROCEDURES

Our first pretest respondents were acquired through local recruitment methods via friends, relatives, and coworkers (such as RAND military fellows and current and former RAND staff). However, the majority of these respondents matched our sample characteristics in only a limited way, being mostly high-ranking white male officers, though of varying branches of the military.

In an effort to diversify our pretest sample, we next moved to Internet recruitment. Many Gulf-War-related Websites were mined and reviewed for content (e.g., the political nature of site or Gulf War illness biases displayed), in search of one that would attract a fair number of Gulf War veterans while presenting site information in a fairly open-forum, objective manner. Our goal was to post an ad calling for pretest volunteers on a site message board, in the hopes that a more diverse respondent group would be attracted than achieved through the previous local recruitment.

The Website that was selected was www.gulfweb.org, where our call for veterans was placed on their Guestbook page in August 1998. Through this, we obtained a tremendously diverse response in terms of military branch, rank, status, age, gender, and geographic location, as well as a great diversity of opinion regarding Gulf War issues. Approximately 35 interviews were completed between August and September with this Internet sample, distributed as follows between status categories: 9 civilian, 11 retired, 11 reserves, and 9 active. The feedback we received was invaluable, and changes made in response to this first official round of pretests included issues surrounding length, wording, and general organizational style.[2]

The second round of pretests was conducted in two phases, using respondents from the overflow of the drawn sample, once OMB approval had been received.

_____

[2]See the section on Building the Instrument for further details.

This round of pretests was used in part to fine-tune the instrument, but of greater importance was its implementation on the CATI system to work out programming glitches in advance of officially fielding the instrument. The first round of CATI pretests yielded approximately 15 completed interviews, after which the instrument was sent back to the programmer for further adjustments. In the second round of interviews, approximately 35 cases were completed in a three-day period.

The initial round of pretest interviews with the respondents of both our local and Internet recruiting was conducted by experienced survey professionals (full-time, senior staff from within RAND's Survey Research Group(SRG)). This unusual use of high-level staff during pretesting was due to the complexity of the survey's content and structure. As the initial versions of the instrument were experimental in their organization, and in an effort to minimize recall bias to the greatest degree, we needed interviewing staff who would be capable of giving informed feedback to the project team about the type of structural changes that would be most helpful in enhancing this goal. Later, once the appropriate organizational structure of the instrument had been established, pretesting was conducted by trained SRG telephone interviewers. This iterative testing process was essential in helping us to pinpoint and fix instrument problems, streamline questioning, and brainstorm and implement a reduction in respondent burden. To characterize the magnitude of the changes that were made, we began with an instrument that took over an hour to complete. We ended with a final product that averaged approximately 20 minutes to complete, with over 97 percent of the participants agreeing to participate in the follow-up survey if selected.

TRACKING

Although our refusal rates were extremely low (approximately 4 percent), the percentage of those who initially were classified as unlocatable was fairly high (see Chapter Two for specific details), even though the sample had been run through several tracking efforts before survey fielding. The details of these processes are outlined below.

The original contact information for our sample came from the Defense Manpower Database Center (DMDC). DMDC used several databases from which to draw contact information, including DEERS files, the retired pay file, the reserved pay file, the Persian Gulf Hotline, and UIC (Unit Identification Code) addresses. These data were then sent to OSAGWI for tracking through Autotrak files. Once this was completed, the sample was sent back to RAND and set up for calling.

The interviewers were the next tracking resource used. The addresses of many cases were generally more accurate than the accompanying phone numbers, so interviewers were instructed to call Directory Assistance once the preliminary number was deemed incorrect. If no other number was found for the respondent, or the number found was deemed incorrect, the case was forwarded to the next level of tracking.

The tracking sources used at this next level were as follows: Autotrak™ Extended Search, Base Locator, Choicepoint, CMAT Post Locator Search, IFC, National Change of Address Service, OSAGWI Veterans Data Management (VDM) Team, Worldwide Locator, and Internet searches (Yahoo, Whowhere, Switchboard, Infospace, Anywho, and Worldpages). In total, this level of tracking generated 3,741 leads, of which only 2,735 were called because of time constraints and duplication of numbers. VDM proved to be the most efficient tracking source in terms of yielding the highest percentage of located and positively identified respondents per number of leads generated (hit ratio of 57.5 percent). The Autotrak Extended Search yielded the highest number of located and positively identified respondents across the board, producing 39.2 percent of all located and finalized cases (compared with 17.8 percent for VDM). However, Autotrak's hit ratio was only 7.6 percent.