Appendix C

Sampling Methodology

Data Collection

Sample Distribution

Sampling Methodology

The following material is provided by Survey Sampling, Inc. (SSI) to describe in detail the method used to select the sample. This study used a Random B sample with likely disconnects eliminated from the sample. However, all types of SSI sample are discussed below.

Creation of the Random Digit Database

SSI starts with a database of all directory-listed households in the USA. Using area code and exchange data regularly obtained from Telcordia and additional databases, this file of directory-listed telephone numbers is subjected to an extensive cleaning and validation process to ensure that all exchanges are currently valid, assigned to the correct area code, and fall within an appropriate set of zip codes.

Most SSI samples are generated using a database of "working blocks." A *block* (also known as a *100-bank* or a *bank*) is a set of 100 contiguous numbers identified by the first two digits of the last four digits of a telephone number. For example, in the telephone number 255-4200, "42" is the block. A block is termed to be *working* if one or more listed telephone numbers are found in that block.
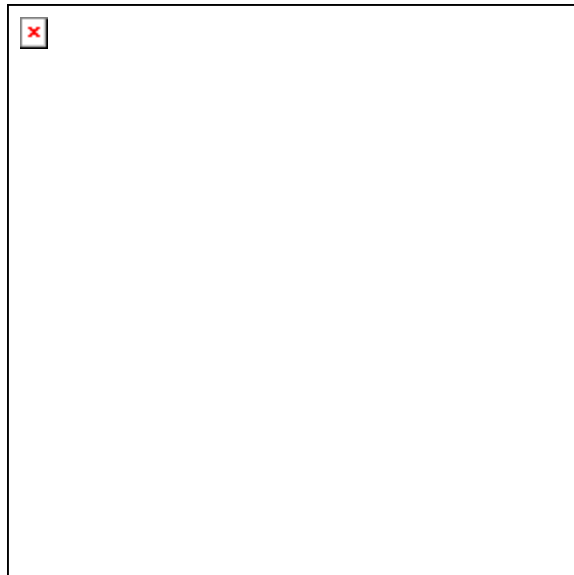
Each exchange is assigned to a single county. Nationally, about 72% of all exchanges appear to fall totally within single county boundaries. For those overlapping county and/or state lines, the exchanges are assigned to the county of plurality, or the county with the highest number of listed residents within the exchange. This assignment ensures known probabilities of selection for all telephone numbers.

Sample Stratification

All SSI samples are generated using stratified sampling procedures. Stratified sampling divides the population of sampling units into subpopulations called strata. A separate sample is then selected from the sampling units in each stratum. SSI's database has been stratified by county.

Measure of Size (MOS) Weights

Prior to sample selection, the sample is allocated proportionally across all strata in the defined geography using one of several frame adjustment options. The sampling frame determines the way a sample is distributed across geography at the county level. SSI offers five different measure of size (MOS) stratification frames for its random digit samples:

- *Estimated Number of Telephone Households*
  Estimates for telephone households are updated annually. The estimates are calculated by subtracting Census non-telephone household counts from current household estimates. Sample units will be allocated to each county in proportion to its share of telephone households. Estimated telephone households is the most commonly used sampling frame for Random B samples.

- *Number of Households*
  Estimates for households are updated annually. Sample units will be allocated to each county in proportion to its share of households in the defined geography.

- *Total Population*
  Estimates for population are updated annually. Sample units will be allocated to each county in proportion to its share of population in the defined geography.

- *Total Active Blocks*
  Sample will be distributed by county in proportion to the total eligible blocks in the exchanges assigned to that county. Rather than being an estimate of target population, all frame units are represented with equal probability across counties. The number of eligible blocks in an exchange is multiplied by 100 (the number of possible 10-digit telephone numbers in a block) to calculate the total number of possible phone numbers. Sample will be allocated to each county in proportion to its share of these possible 10-digit telephone numbers. This is the recommended frame for apportioning Random A samples.

- *Other*
  Sample allocation may also be based on special frames which may or may not result in equal probability samples. Such frames may be user-defined or based on incidence estimates and may be used singly or in combination with these or other sampling frames.

Samples are first systematically stratified to each county in the survey area in proportion to the sampling frame selected. After a geographic area has been defined as a combination of counties, the sum of the estimated telephone

households or requested frame value is calculated and divided by the desired sample size to produce a sampling interval.

The counties are ordered by alphabetically by state. A random number between zero and one is generated and multiplied by the sampling interval to calculate a random starting point between one and the sampling interval. A cumulative count of elements is calculated. At the point at which the accumulation reaches the random starting point, a specific county is selected and the next sampling point is one interval away. Accumulation continues in this fashion until the entire sample has been apportioned.

Sampling Frame Adjustments
- *Minimum Acceptable Block Size*
  Approximately 2.5 million blocks are identified as working (having one or more listed numbers). By raising the minimum acceptable block size from 1 to 3 (SSI's default) or more, further gains in efficiency can be achieved with only minimal reduction in coverage. Blocks with 1-2 listed numbers represent only 5.9% of all working blocks and only 0.3% of all listed telephone households. These listed numbers are far more likely to be keypunch errors or White Page business listings than the only listed number in a given block. SSI uses a default minimum block size of 3 listed numbers, but this minimum may be adjusted up or down based on the user's specifications. Users can even sample from blocks with zero listed numbers, but efficiency may fall as low as 16%.

Sample Selection

After the sample has been allocated, three methods of systematic sample selection are available.

Random A is an SSI term denoting samples of random numbers systematically selected with equal probability across all eligible blocks. All blocks within a county are organized in ascending order by area code, exchange, and block number. Once the quota has been allocated to all the counties in the frame, a sampling interval is calculated for each county by summing all the eligible blocks in the county and dividing that sum by the number of sampling points assigned to the county. From a random start between zero and the sampling interval, blocks are systematically selected from each county. Once a block has been selected, a two-digit number is systematically selected in the range 00-99 and is appended to the exchange and block, to form a 10-digit telephone number.

Random B is an SSI term denoting samples of random numbers distributed across all eligible blocks in proportion to their density of listed telephone households. All blocks within a county are organized in ascending order by area code, exchange, and block number. Once the quota has been allocated to all counties in the frame, a sampling interval is calculated by summing the number of listed residential numbers in each eligible block within the county and dividing that sum by the number of sampling points

assigned to the county. From a random start between zero and the sampling interval, blocks are systematically selected in proportion to their density of listed households. Once a block has been selected, a two-digit number is systematically selected in the range 00-99 and is appended to the exchange and block to form a 10-digit telephone number.

Epsem Samples *(equal probability of selection method)* are single stage, equal probability samples of all possible 10-digit telephone numbers in blocks with one or more listed telephone numbers. The Working Phones Rate (WPR) for an epsem sample is on average 50%, but can range from 30%-70% depending on the size and nature of the geographic area and local telephone number assignment practices. Epsem sampling uses a *total active blocks* frame and *Random A sampling methodology*. A sample of random numbers is systematically selected with equal probability across all blocks containing one or more listed numbers, which distributes the sample across counties in proportion to their share of total active blocks. Epsem samples have the following characteristics:

- Minimum block size is 1.
- Business numbers cannot be replaced, but can be flagged.
- Number protection is unavailable.

Selection Options
SSI's database and sampling software support a variety of other epsem and non-epsem sampling options designed to accommodate different sample specifications or study objectives:

Business Number Purge
To improve efficiency, SSI maintains a database of over 11 million business telephone numbers, compiled from Yellow Page directories and special directories (Standard & Poor's and industry specific directories). Once a 10-digit telephone number has been selected for a sample, the status of the number generated may be compared to SSI's list of known business numbers. If the RDD number matches a known business listing, two options are available:

- *Replace the number with the next number that is not a known business number.* This is a non-epsem procedure but ensures that the requested sample size is met. In order to prevent introducing additional sampling bias, this procedure operates within strict limits. During either Random A or B sample selection, the search will not go beyond the boundaries of the selected block.

- *Select the number but flag it as a business number.* This option preserves epsem sampling. Business numbers selected and flagged may be included as part of the sample or removed. If these numbers are retained in the final sample file, they may be sorted to the bottom of the file or the bottom of each replicate. If these numbers are removed, the final sample file will fall short of the requested sample size.

Number Protection
Virtually every SSI Random Digit Sample is marked on the database to protect against reuse for a period of six months. The SSI *Protection System* was designed to reduce the chance of selecting the same number for multiple projects or multiple waves of a single project conducted by a single research firm or by competing research firms.

Incorporating number protection during sample selection is only an option. Once a 10-digit telephone number has been selected for a sample, the "protected" status of the number selected is checked. If the number has not been selected for a sample in the previous six months, the selected number is marked as "protected" and sampling continues. If the number is identified as having been selected for a sample in the previous six months, two sampling options are available:

- *Replace the number with the next number that is not a "protected" number.* In order to prevent introducing sampling bias in areas which have been frequently sampled, this procedure operates within strict limits. In either Random A or B sample, the search for an eligible replacement will not go beyond the boundaries of the selected block. In the event that an eligible replacement cannot be found within these limits, the originally selected "protected" number will be taken.
- *Select the number anyway, preserving epsem sampling.*

Screen for Disconnected Numbers
The SSI *Sample Screening Service* is a stand-alone, post-production process that identifies non-working or unassigned numbers, as well as modem and fax numbers in random digit telephone samples. It employs a new and proprietary technology that recognizes almost half of these numbers, thereby improving the effective working phones rate of random digit telephone samples by an average of 10-15%. Once these numbers have been identified, two options are available:

- *Remove disconnects from the sample.* If these numbers are removed, the final sample file will fall short of the requested sample size. Number removal may be exercised either before replication (sample will have equal sized replicates) or after replication (sample will have unequal sized replicates but each replicate will contain exactly the same "good" telephone numbers as it would have if the sample had not been screened).

- *Include disconnects as part of the sample.* If these numbers are retained in the final sample file, they are flagged and may be sorted to the bottom of the file or the bottom of each replicate and printed on separate sample pages.

The sample for TDI used the option of removing the disconnects from the sample.

DATA COLLECTION

Recruiting and Training Interviewers

In order to complete the survey within the allotted time, PPRI employed 155 interviewers for the project. A large percentage of the interviewers were selected from among those who had extensive experience with other PPRI interviewing projects. PPRI's established pool of interviewers was composed of both university students and local community residents who were already participating in existing studies or who had worked on earlier projects. The availability of experienced interviewers facilitated the training process and contributed to a high-quality survey product.

Recruiting New Interviewers

New interviewers were recruited and selected according to PPRI's standard operating procedures. This process began with the announcement of new interviewer positions in local and university newspaper advertisements and at student employment offices. A new outlet for recruiting employees was utilized by posting an advertisement on the new student employment section of the Texas A&M University web page. A multi-step screening process required potential interviewers to telephone a Survey Research Laboratory supervisor. Candidates were initially screened through this first telephone conversation. Those who failed to present themselves well over the phone were eliminated from further consideration. Those who passed the preliminary screening were asked to visit the Survey Research Laboratory and to complete an application form. Lab staff, including the Operations Coordinator, Assistant Coordinator and Senior Supervisors, interviewed the more promising applicants. In addition to providing standard employee information, the candidate was required to conduct a brief, mock telephone interview with a supervisor, using the actual project questionnaire. Each applicant received a rating and the top applicants were selected.

Training Interviewers

Training sessions were designed to encourage active participation on the part of the trainees, to familiarize them with the different types of respondents they might encounter and, ultimately, to maximize interviewer effectiveness. A large portion of the training session was devoted to a question-by-question review of the survey instrument. In addition, much of the training session involved didactic classroom activities and practice conducting mock interviews.

During the training session, each trainee was observed and evaluated. Trainees who did not perform satisfactorily were either given additional individualized training or were replaced. Topics covered in the training included:
- General information about TDI & PPRI;
- Organization of the interviewing staff including the responsibilities of supervisors, interviewers, and other staff;
- Standard management procedures including scheduling, logging in and out, payroll, sickness, absences, and tardiness;

- Information on sampling procedures (how it works in general, how the TDI survey was derived, what the interviewer must do, why the procedures must be followed exactly);
- General instructions on interviewing including preparing the interviewer, establishing contact, maximizing response rates, and handling problems and objections;
- Discussion about interviewing techniques such as maintaining neutrality, encouraging responses, and probing;
- Mechanics of the survey including pronunciation, skips, and allowable clarifications;
- Discussion about specific problems (such as the purpose of the survey, what will be done with the data, or substantive questions about survey content); and
- Procedures for ensuring confidentiality;

Senior project staff and the Survey Operations Coordinator presented the training material. Supervisors worked with trainees both individually and in groups. Although some of the material was presented in a lecture format, most of it was illustrated by example, or through hands-on participation in exercises designed to simulate actual interviewing experiences. Finally, interviewers conducted mock interviews with one another, using the actual Computer-Assisted Telephone Interviewing (CATI) equipment, while supervisors made observations and provided feedback.

All interviewers received at least eight hours of training. The first four-hour session covered general interviewing issues and provided a conceptual foundation for the TDI uninsured project. A second two-hour session provided specific instruction on the survey instrument itself. The final two hours were devoted to practicing the interview using the CATI system.

Finally, new interviewers were carefully monitored during a trial period to identify and remedy problems. This on-the-job-training continued until the basic skills were mastered. At least five experienced shift supervisors were assigned to the project and were trained along side the interviewers.

Conducting Interviews

PPRI began collecting data via telephone interviewing on October 15, 2001 and concluded on December 6, 2001. Detailed procedures were established for conducting interviews. Prior to each week of scheduled interviews, the supervisory staff determined the requisite number of interviewers to be assigned to each shift. Typically, for a project of this magnitude, 30 to 40 interviewers were assigned to the project during evening (6:30-9:30) and weekend shifts (10:00-2:00 and 2:30-6:30 on Saturday, and 1:00-5:00 and 5:30-9:30 on Sunday). Additionally, four to five interviewers worked on the project during business hours to make daytime attempts and call-backs.

Survey Supervision

The Survey Research Laboratory supervisory staff oversaw daily preparations. The routine consisted of the following tasks:

- Use sample status reports (generated daily) to identify potential problems and establish priorities for interviewing during the shift;
- Use interviewer productivity reports (generated daily) to identify problems; and
- Determine the appropriate response to refusals (e.g., scheduling another attempt) and other special situations.

Prior to each shift, the shift supervisors:

- Allocated interview stations on the CATI to interviewers;
- Assigned interviewers to special tasks, such as refusal conversion; and
- Determined which interviewers to monitor (priority was given to new interviewers, interviewers with recognized problems, and the interviewers who had not been monitored during their last four shifts).

During an interviewing session, shift supervisors were responsible for:

- Answering interviewer questions;
- Resolving difficult situations posed by respondents;
- Monitoring interviews -- at least 20 percent of the interviewers in a shift were monitored and at least five percent of interviews conducted were monitored;
- Maintaining shift productivity; and
- Monitoring the CATI system to make sure that appropriate allocations of the sample were made.

Interviewers were carefully supervised. One supervisor was routinely on duty for every ten interviewers. Interviews were regularly monitored from a central phone. As noted above, supervisors monitored five percent of the interviews conducted on the CATI system by observing all screen and keyboard activity at a workstation from a central terminal.

Household Contacts

The standard PPRI procedure for attempting to contact a household is to place a call during each of five different shifts throughout the week. Numbers that had been disconnected were tried twice. Busy numbers were tried twice during the same shift, with repeated attempts during five different shifts. Once a household was reached, but the correct respondent was not available, as many as five more attempts were made to reach the correct respondent.

Attempts were made to convert virtually all refusals. When a refusal occurred, interviewers completed a special form that provided as much information as possible on the circumstances of the refusal. The respondent was then re-contacted by interviewers specially trained to convert refusals. These procedures maximized the response rate.

Confidentiality

There are a variety of procedures that ensure confidentiality in the interviewing process. PPRI is required to maintain confidentiality of records on a variety of projects, including ones in which records are maintained on identified individuals. The approaches include maintaining security, following specified procedures, and training and supervising employees.

The CATI system enables control to be maintained over all files and records. The computer handles all sample management and data collection. The computer system is secure and all areas where confidential material is stored is password protected and accessible only to a select group of staff. Floppy disks from the workstations contain data that are not readable in a meaningful way without access to computer programs available only to supervisory staff. Additionally, the premises and physical storage areas are secured.

The most important procedural consideration in maintaining security is to make sure that the anonymity of the telephone interviews is not compromised. In the CATI system, specific information (e.g., telephone number and name of someone to be called back) are in a file separate from the collected data. These files can be linked, but they are not maintained in a linked form.

All staff at PPRI is aware of the issues involved in confidentiality. Highlighting its importance is part of all new employee training as well as the monitoring and supervision processes.

Data Coding

A few of the questionnaire variables required coding of verbatim responses entered by the interviewers. Experienced coders used a program that displayed the response alternatives for each question and allowed a code to be entered. Data was coded for each question (rather than for each respondent) at a time, thereby increasing coding consistency. Randomly selected surveys were monitored.

Quality Control Measures

To ensure the quality of the survey data, PPRI used several internal checks to guide survey processing. Many aspects of the quality control plan were embedded in automated procedures of the CATI system. The CATI was programmed so only valid codes could be entered by the interviewers and all skips occurred automatically. The CATI also checked the internal consistency of responses during the interview, allowing corrections to be made at any point in the interview process. The design of the CATI system also prevented data loss by PPRI. Each workstation constituted a separate computer linked with others through a computer network. All files on the network were automatically backed up on tape every night.

There is, however, no replacement for human oversight. Monitoring procedures allowed supervisory staff to identify problems of inconsistency, interviewer practices that might affect response rates, and practices that affected the ability of the respondents to understand some interviewers. The CATI also allowed project data to be constantly accessible to researchers. This data provided information about cooperation rates, number of calls made, and other characteristics of interviewers that were monitored constantly. Any problems were identified and handled immediately by the supervisory staff. In addition, a random five percent of the interviews were verified.

Sample Distribution

Finding the respondents involved a very large effort. The survey found that 4.6% of all successfully screened respondents fit the qualifications. The detailed summary of the status of all calls is presented in the table below. PPRI started with 61,702 telephone numbers. Over 20,000 of those were found to be bad numbers (not working, not residential, and so on). We were able to successfully determine the age appropriateness, insurance status, and poverty status of 18,030 households. From those, we identified 830 meeting the requirements of the survey and completed interviews with 598 of them. The table shows the counts and percentage distributions for the entire 61,702 sampled numbers, for the 40,691, excluding the bad numbers and the 18,030 of those that were successfully asked the screening questions.

Distribution of Sample Disposition

	Count	Excluding Bad Numbers	Those Screened
Screened-Over 64	4097	10.1%	22.7%
Screened-Have health insurance	11508	28.3%	63.8%
Screened-Have low income	1595	3.9%	8.8%
Unknown income or number in HH	110	0.3%	
Not reached before screen	11365	27.9%	
Bad number*	21011		
Refused/Terminated before Screen	11186	27.5%	
Higher income/no insurance-completed survey	598	1.5%	3.3%
Not reached after screen	14	0.0%	0.1%
Refused/Terminated after screen	218	0.5%	1.2%
Total	61702	40691	18030
		100.0%	100.0%

*Not reached-No answers, continual busy, answering machines, and callbacks not resolved by the end of the survey.

The completion rate for the screening and full survey was 44%. That is, 44% of the numbers not identified as bad resulted in a completed final interview and/or screening interview. Of those screened, 72% percent resulted in a completed interview.

The amount of effort made to complete the calls is reflected in the data reported below. An average of 4.8 calls were attempted on each of the 61,702 telephone numbers. The efforts for the uncompleted categories are considerably greater. For example, eleven attempts were made on average for those not completed. These are a combination of situations, but reflect a number that has an indication that it may be a working number. It also includes numbers that are designated as "callbacks" indicating that the interviewer has made contact with someone in the household, although they have not yet consented (or declined) to do the interview. The 11 average attempts reflects the effort made to complete these calls.

Number of Call Attempts by Disposition

	N	Average Number of Attempts
Screened-Over 64	4097	3.1
Screened-Have health insurance	11508	3.4
Screened-Have low income	1595	3.3
Unknown income or number in HH	110	3.4
Not reached before screen	11365	11.2
Bad number	21011	2.1
Refused/Terminated	11186	6.0
Higher income/no insurance-completed survey	598	3.6
Not reached after screen	14	3.6
Refused/Terminated after screen	218	6.5
Total	61702	4.8

The productivity of the sample was less than we estimated prior to beginning the survey. Only 4.6% of the respondents completing the screening questions were eligible for the survey. Further, 28% of them declined to participate in the longer interview.