



## RDD Landline Sample Methodology

According to the Census Bureau's 2007 American Community Survey estimates, 95% of households (106.3 million households) in the United States have telephone service. To represent these telephone households in a sample is a challenge for several reasons: households without landline telephone service, unlisted by choice, and unlisted by circumstance (mobility). The National Center for Health Statistics reported that in 2008, roughly 20% of all households had only wireless or cell phone service and only 76% of households could be reached on a landline telephone. Approximately 30% of the landline telephone households in the U.S. have unlisted numbers. Each year, about 20% of American households move, so that 12-15% of the residential numbers in a typical directory are disconnected over the life cycle of the directory. Samples drawn entirely from directories, and "plus-one" techniques based on directory seed numbers, often significantly under-represent unlisted households.

To address these representation issues, SSI developed a variety of random digit (RDD) methodologies. (Refer to the SSI Wireless Methodology-US information sheet for more detail on supplementing landline RDD samples with Wireless/Mobile RDD sample.) SSI's random digit landline methodology is unsurpassed in efficiency. You'll experience fewer dialings, faster completion, and improved interviewer morale.

### I. 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 203-567-7200, "72" is the block. A block is termed to be working if one or more listed telephone numbers are found in that block.

### Calculating Sample Size

The sample size is the number of cases selected from the universe to use as the sample. In this case, the number of telephone numbers required to meet the desired number of interviews is the sample size. The basic equation used to derive the correct sample size is as follows:

$$\text{Size} = \frac{\text{Completed Interviews}}{\text{Working Phones Rate} \times \text{Incidence} \times \text{Contact/Cooperation Rate}}$$

If, for example, 800 completed interviews are needed, and the working phones rate equals 0.46, the incidence is 0.70, and the completion rate is estimated at 0.23, a sample size of 10,802 numbers would be required ( $800 / (0.46 \times 0.70 \times 0.23)$ ).

#### Working Phones Rate (WPR)

The WPR for a Random B sample is, on average, 45%. Because of geographical differences in phone distribution, individual samples may show a higher or lower rate. Lower rates could range as low as 35%. The listed sample WPR averages 80%, EPSEM sample averages 32%.

#### Incidence

Population incidence is the percentage of the type of respondent you wish to interview that you can expect to find in the sample universe. Product incidence is the percentage of qualified respondents who use a certain product or qualify for your survey in some way. Both population and product incidences may be elements in the sample size calculation equation.

#### Contact/Cooperation Rate: 20-25%

Contact rate includes: No Answers, Answering Machines, and Busy Signals assuming three attempts for each number. Cooperation rate is: Immediate refusals and getting the qualified person to complete your interview.

*(Refer to the SSI White Paper 8: Declining Working Phone Rates Impact RDD Efficiency for more detail on WPR for landline RDD samples.)*

*Continued on next page*



info@surveysampling.com • surveysampling.com



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.

## II. 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. These estimates do not take into account those households that have only wireless telephone service since estimates of wireless-only households are currently not available for states or counties.

- **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.

*Continued on next page*



info@surveysampling.com • surveysampling.com



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 and determine the amount of sample to be allocated to each county in the sample.

The counties are ordered by alphabetical state and county within 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 2.8 million 100-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 7.4% of all working blocks but only 0.4% of all listed telephone numbers. These listed numbers are far more likely to be keying 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 10%.

### III. Sample Selection

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

**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 eligible blocks 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.

**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.

**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 32%, but can range from 20-60% depending on the size and nature

*Continued on next page*



info@surveysampling.com • surveysampling.com



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.

#### IV. 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:

##### The Landline RDD Method to Fit Your Needs

**Random B** ... provides the most efficient random digit sample available. Each exchange and working block will have a probability of selection equal to its share of listed telephone households in eligible blocks. Numbers are protected against reuse for a period of six months. Business numbers are eliminated.

**Random A** ... provides an extremely representative random digit sample. Each exchange and working block will have a probability of selection equal to its share of eligible working blocks. As an option, numbers selected in the past six months may be eliminated. (See Number Protection for further details.) Also as an option, business numbers may be eliminated.

**EPSEM** (Equal Probability Selection Method) ... every possible telephone number (including business and protected numbers) in a working block with at least one directory-listed telephone number has an equal probability of selection. Telephone numbers are not protected. Business numbers may be identified, but not removed.

#### 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

*Continued on next page*



info@surveysampling.com • surveysampling.com



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 any way, 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 more than half of these numbers, thereby significantly improving the effective working phones rate of random digit telephone samples. 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.

*(Refer to the SSI Sample Screening Service information sheet for more detail.)*

### **About Survey Sampling International**

Survey Sampling International is the premier global provider of sampling solutions for survey research. SSI offers access to more than 6 million consumer and business-to-business research respondents in 54 countries via internet, telephone, and mobile. Additional client services include customer profiling, survey programming and hosting, data processing, sampling consulting, and survey optimization. SSI serves more than 1,800 clients, including nearly three-quarters of the top researchers, worldwide. Founded in 1977, SSI has an international staff of 400 people representing 50 countries and 36 languages. The company has 16 offices around the world for local client support. For more information, contact SSI at +1.203.567.7200 or info@surveysampling.com.