March 2010

New collection entitled: FCC Wired and Wireless Internet Survey

Part B: Collections of Information Employing Statistical Methods:

1. *Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons)in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.*

The target population for this study is the adult population of the 50 states of the United States and the District of Columbia.

The study will focus on both on households estimated to total approximately 129 million as of July 1, 2008[1], as well as individuals. The survey will seek reliable national estimates of performance, pricing, and other terms and conditions of broadband and related service offerings, including individuals' understanding of those characteristics.

In addition, the survey will attempt to provide a reliable national estimate of how often American adults change providers of broadband and related services and the barriers they face to changing providers. The estimated civilian non-institutional population age 18 and older as of 2008 is 227 million.[2]

Based on recent experience, this survey, calling landline and cell phone numbers, will achieve a response rate of between 15 and 25 percent using American Association of Public Opinion Research (AAPOR) response rate definition number 3. This response rate estimate is based on the severely constrained interviewing period of the survey.

2. *Describe the procedures for the collection of information including:*

- *Statistical methodology for stratification and sample selection,*
- *Estimation procedure,*
- *Degree of accuracy needed for the purpose described in the justification,*
- *Unusual problems requiring specialized sampling procedures, and*
- *Any use of periodic (less frequent than annual) data collection cycles to reduce burden.*

As directed by the FCC, the survey is designed to achieve appropriate national estimates of population parameters of plus or minus 2 percentage points at a 95 confidence level.

[1] US Census Bureau, Population Estimates Program.
[2] U.S. Census Bureau, 2006-2008 American Community Survey

The consumer survey will be conducted using one questionnaire.

The survey, designed to provide rigorous estimates of broadband-related characteristics of all American households at the national level, will consist of 3,000 completed interviews with adults age 18 and older in the 50 states and the District of Columbia. This portion of the survey will provide national estimates subject to sampling errors of no more than plus or minus 2 percentage points at the 95 percent confidence level.

One should note that this design provides the requested level of accuracy for the total main survey. Estimates of the characteristics and behaviors of sub groups – including broadband adopters as one subgroup and non-adopters as another subgroup – will be subject to higher sampling error margins based strictly on the smaller number of interviews in a specific subgroup.

The second segmentation of the population that is critical to this study is the Urban-Suburban-Rural (U/S/R) continuum. The basic classification of the data will be based on an analysis of the location of the telephone exchange of the landline telephone number. The classification is subject to a certain degree of error, especially since the definition of suburban is not widely accepted. However, for cell phone numbers the U/S/R determination is much more difficult since the location reported to the telephone sample provider is related to where the cell phone was activated which may or may not have relevance to current home address for the respondent. Using techniques developed by PSRAI, each cell phone interview will be categorized in terms of U/S/R.

The average expected length of the interview will be no more than 20 minutes. The survey will be conducted in English or Spanish, at the respondent's discretion.

The survey will be conducted using two samples:

- A standard Random Digit Dialing (RDD) sample of landline telephones; and
- An RDD sample of cell phones.

Consumer telephone Samples

Two samples will be used for the consumer data collection - a random digit dial (RDD) landline sample and an RDD cell sample. The landline sample frame will be an equal probability sample across all active blocks.[3] All blocks within a county will be sorted in ascending order by area code, exchange and block number. A sampling interval will be computed for each county in our sample by summing all eligible blocks in the county and dividing that sum by the quota 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 random number is appended to the block to create a phone number. Business numbers will not be excluded at the sampling stage. Rather we

[3] Active blocks are defined as 100 contiguous phone numbers (e.g., 609-924-9200 to 609-924-9299) with at least one residential directory listing.

will purge the numbers flagged as business before the numbers are dialed.[4] Additionally, we will not exclude protected numbers from our sample.[5]

The cellular sample will not be list-assisted because no list of cellular numbers exists. Rather, cellular phone numbers will be systematically sampled from dedicated wireless 100-blocks and shared service 100-blocks with no directory-listed landline numbers.

Sample Weighting

The landline and cell samples will be combined in analysis. The data will be weighted to correct for three sample elements that could bias sample estimates - [1] different probabilities of selection based on phone use and household composition, and [2] disproportionate non-response.

Different Probabilities of Selection based on Phone Use: The people and households we interview will have different probabilities of being in our sample based on three variables: [a] the number of phone lines in the household; [b] the number of adults in the household; and [c] the number of adults in the household who have a cell phone. For example, a person who lives alone in a house with one landline phone will have twice the probability of being in our landline sample as an adult who lives with one other adult in a household with only one landline telephone. Likewise, an adult with both a landline and a cell phone has twice the probability of being sampled as an adult with only one kind of phone.[6]

The probability of a household being included in the sample is a function of the number of phone lines into the household (n_L) and the number of adults with cell phones (n_C). We will assume independence in the sampling of the two frames so the probability that a household is sampled will be the probability it is sampled by landline plus the probability it is sampled by cell ($P_L n_L + P_C n_C$). Then the appropriate weighting adjustment to correct for unequal probabilities of household selection would be $1/(P_L n_L + P_C n_C)$. If we assume that $P_L = P_C = P$, then the probability of a household being selected would simplify to $P*(n_L + n_C)$ and the adjustment would simplify to $1/(n_L+n_C)$.

The probability that a person will be sampled will be a function of the number of phone lines in the household (n_L), the number of adults in the household (n_A) and whether or not the person has a cell phone. So the probability of being sampled would be:
$P_L n_L/n_A + P_C$ if the person has a cell phone, or
$P_L n_L/n_A$ if the person has no cell phone (i.e., $P_C=0$)
If we again assume that $P_L = P_C = P$, then the probabilities simplify to:

[4] The reason we do not have SSI purge the business numbers is that their sampling program automatically replaces a purges business number with another RDD number from the same block. This will, in effect, oversample numbers in blocks with business numbers.
[5] Phone numbers in SSI random digit database are flagged as "protected" if they have recently been pulled for any sample order. While excluding protected numbers will resulting in a "fresher" sample (i.e., one that consists of numbers that have not been recently pulled), we feel that excluding these numbers could potentially bias our sample.
[6] We will assume that the probability of being sampled in each frame is equal.

$P \times n_L/n_A + P = P \times (n_L/n_A + 1) = P \times (n_L+n_A)/n_A$ if the person has a cell phone
$P \times n_L/n_A$ if the person has no cell phone
So the proper adjustment to make the probabilities of selection constant would be:
$n_A/(n_L+n_A)$ if the person has a cell phone, or
n_A/n_L if the person has no cell phone

Differential non-response: It is well established (at least in landline samples) that certain kinds of people are easier to reach and more cooperative than others and therefore end up over-represented in RDD samples. One example of this is that older people are much easier to reach by landline phone and are almost always over-represented in landline samples. In order to correct for potential non-response bias in our sample we will rake the final sample demographics to match population parameters. We will rake the data to match population parameters for: sex by age; sex by education; age by education; race/ethnicity; census region; population density and telephone use. All but two of the parameters will be derived from the Census Bureaus' 2009 ASEC data. The population density parameter will come from an analysis of Census 2000 data at the county level. The phone use parameter will come from the most recently available statistics from the National Health Interview Survey. [7]

[7] Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June 2009. National Center for Health Statistics. December 2009.

Effect of Weighing on Sample Estimates

Post-data collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. PSRAI calculates the effects of these design features so that an appropriate adjustment can be incorporated into tests of statistical significance when using these data. The so-called "design effect" or *deff* represents the loss in statistical efficiency that results from disproportionate sampling and systematic non-response. PSRAI calculates the composite design effect for a sample of size n, with each case having a weight, w_i as:

$$deff = \frac{n\sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2} \qquad \text{formula 1}$$

In a wide range of situations, the adjusted standard error of a statistic should be calculated by multiplying the usual formula by the square root of the design effect (√deff). Thus, the formula for computing the 95%

$$\hat{p} \pm \left(\sqrt{deff} \times 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) \qquad \text{formula 2}$$

confidence interval around a percentage is:

where $\hat{p}$ is the sample estimate and n is the unweighted number of sample cases in the group being considered.

We estimate that the total sample design effect will be between 1.20 and 1.30 and that the final total sample margin of error will be 2.0 percentage points.

3. *Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.*

The survey will be conducted in accordance with best practices of the survey industry. Trained and experienced interviewers will be briefed on the project, its goals and the specifics of the questionnaire. A full 15-call design will be implemented for the landline survey with a 7-call design for the cell phone numbers. This means that the field house will make up to 15 calls to each landline telephone number in the sample in an effort to complete an interview. Recent experience suggests that calling cell phone numbers repeatedly in an attempt to complete an interview beyond 7 calls is not productive. A telephone number will be removed from the active sample only when a terminal disposition has been achieved for that number (completed interview, non-working number, business number, hard refusal, etc.) Refusal conversion will be attempted at least once on all soft refusals.

Plan To Examine Potential Non-Response Bias

In order to quantify and measure potential non-response bias on survey results, we propose investigating two types of non-response. First, we will look at unit non-response. This is non-response at the survey level caused by not contacting potential respondents or not gaining cooperation once a potential respondent is reached. We will also investigate item non-response. This is non-response at the question-level and is caused by people not answering specific questions during the interview. This analysis tracks with the analysis being conducted on other FCC surveys by PSRAI in the past 12 months.

Unit non-response

The first step in our unit non-response analysis will be to use an approach, in line with Keeter, et. al. and previous PSRAI analyses, that will segment the completed interviews by the level of effort that was required to get the interview.

We will segment the completed interviews into four categories based on the amount of "effort" it took to complete the interview. "Effort" will be measured as a function of the number of calls and refusal conversions necessary to gain cooperation.

We will start with the following definitions that we have used in a past non-response analysis, but which may be modified slightly.

- Least effort. Five or fewer calls and no refusal conversions.

- Five or fewer calls with one refusal conversion or six to ten calls with no refusal conversions.
- Six to ten calls with at least one refusal conversion.
- Most effort. More than ten calls.

We will also experiment with second way to measure effort which will simply be to assign an effort measure E to each completed interview defined as E = C + R, where C is the number of calls made and R is the number of refusals. For example, if a number was called three times with one refusal, then E=3+1=4.

These effort measures will then be used to compare and contrast key survey results. For example, we will see if the harder to reach respondents are more or less likely to have broadband access in their home.

Where population parameters are available, we will see what effect extra effort has on the quality of our data. For example, from Census data we know the demographic profile of the adult population (e.g., sex, age, education, and race/ethnicity). We will be able to see what effect more effort has in matching sample demographics to population parameters.

This effort will draw on the 2007 work done by PSRAI. As part of a project for the University of Illinois and the Pew Internet and American Life Project that was funded by the Institute for Museum and Library Studies,[8] a study of possible non-response bias in a large national telephone survey was conducted at the request of OMB. A version of the analysis is available online.[9]

[8] IMLS grant award number LG-06-05-0398-05, OMB Clearance Number, 3137-0070, expiration date 06.30.2010.

[9] *How Different Are People Who Don't Respond to Pollsters*?**,**Witt and Best, April 21, 2008, pewresearch.org/pubs/807/how-different-are-people-who-dont-respond-to-pollsters

Item non-response

Item non-response is due to respondents no answering specific questions during the interview. We will investigate unit non-response by identifying which questions have the highest and lowest levels on "DK/Ref" response. We will compare item results with item non-responders both included and excluded and see what effect this has on reported results. We will also compare item non-response for the two sample frames. We will identify questions that have different levels of non-response by sample frame.

We can also see if there is a link between item and unit non-response. We will investigate whether people who are harder to reach have different levels of item non-response than those who are easier to reach. This analysis will involve comparing levels of item non-response across different kinds of respondents we identify in the unit non-response analysis.

4. *Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.*

If the timetable permits, a field test will be held to confirm the questionnaire design, question wording and CATI programming of the surveys.

If the field test is held after OMB approval has been provided, the survey will be field tested in the following manner:

1. The questionnaire will be programmed in the CATI system and the programming checked for accuracy.
2. A small group of interviewers will be briefed on the study.
3. These interviews will then complete up to 15 interviews with actual respondents from the appropriate populations.
4. The interviews will be closely monitored by field house supervisors and debriefed at the end of each shift.
    a. A focus will be on question clarity and assuring that the respondent understands each question.
    b. In addition, the flow of the questionnaire will be judged, based on how respondents deal with changes in topics throughout the survey.
    c. The interviewers will be checked for proper performance and adherence to the survey instructions.
    d. The proper functioning of the CATI system will be checked in real time and the appropriate log files examined for any problems.

Should the field test take place in advance of OMB approval, the field house will conduct the field test using its employees as test respondents. Other than field house staff, the only other individuals who may be contacted for the field test are staff members of the

contractors or the FCC. Other than the change in source of the respondents and a reduction in the total number of interviews, the test will proceed as if members of the public were being interviewed.

5. *Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.*

John B. Horrigan, PhD
Consumer Research Director
National Broadband Task Force
202-418-1553 (office)
443-904-2709 (cell)