

New collection entitled “National Broadband Plan Survey: Demand for Broadband”

**B. Collections of Information Employing Statistical Methods:**

The agency should be prepared to justify its decision not to use statistical methods in any case where such methods might reduce burden or improve accuracy of results. When item 17 on the Form OMB 83-I is checked, "Yes," the following documentation should be included in the Supporting Statement to the extent that it applies to the methods proposed:

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.*

Total Sample available from KN	50,000
Total Sample for this project	4,500

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

The study will focus on both on households estimated to total 129,065,264 as of July 1, 2008, as well as individuals.

Based on recent experience, this survey will achieve a response rate of **between 15 percent and 20 percent** using the appropriate American Association of Public Opinion Research (AAPOR) response rate definitions. This response rate is consistent with other Knowledge Network online surveys that have been approved by OMB. Knowledge Networks is a member of AAPOR and follows its guidelines for response rate reporting. AAPOR has developed standards for the calculation of response rates that must be followed by its members, and these are the response rate standards that to the best of our knowledge are also employed by OMB. However, the AAPOR response rate standards were not established for web panels.

Knowledge Networks has provided technical leadership for the industry in the calculation of web panel response rates. Two of its survey scientists, Charles DiSogra (Chief Statistician) and Mario Callegaro (Senior Survey Methodologist), recently published a path-breaking article in *Public Opinion Quarterly* to provide for the first time a standard for web panel response rate calculations. The citation is:

Callegaro, Mario & DiSogra, Charles (2008). Computing Response Metrics for Online Panels. *Public Opinion Quarterly*. 72(5) pp. 1008-1031.

The below discussion of response rates is based on the Callegaro and DiSogra article. Response rates for surveys conducted using KnowledgePanel must be calculated taking into account all stages of participation. For this reason, response rates appear significantly lower than for other surveys that involve only one stage of sampling (e.g., a simple cross-sectional sample survey). The cumulative or overall response rate is the product of the response rates obtained at each stage of participation. For KnowledgePanel, there are four stages for which response rates must be calculated:

**Stage 1: RDD Panel Recruitment.** The rate used for calculating the response rate for households recruited by Random Digit Dialing (list-assisted sampling of 1+ directory-listed telephone banks) is AAPOR Response Rate No. 3, the response rate formula approved by AAPOR (2006):

$$RR3 = I / (I + R + NC + O + e(UH + UO))$$

where I is the number of household-level recruits to the panel, R is the number of refusals or break-offs, NC is a non-contact, O is other, UH denotes an unknown final disposition for a known household, and UO is an unknown final disposition for a number of unknown household status. The term e is the estimated proportion of cases of unknown eligibility that are actually eligible households. The term e is calculated as:

$$e = I / (I + IE)$$

where IE is the number of ineligible households.

At present for most Knowledge Networks-conducted panel surveys, the AAPOR Response Rate No. 3 for recruitment is approximately 30%. The actual response rate depends upon which RDD sample replicates donate the actual survey sample selected for any given web panel survey.

This Knowledge Networks panel recruitment response result is comparable to rates reported by high-quality RDD surveys. The panel recruitment response rate for the opt-in web panels cannot be calculated as the initial sample is not drawn probabilistically from a defined sample frame.

**Stage 2: Connection Rate.** This is the number of households that “connect” to the panel by completing their first Knowledge Networks profile survey divided by the number of RDD recruited households. At present for most Knowledge Networks-conducted panel surveys, the connection rate is approximately 60%. The actual response rate depends upon which RDD sample replicates donate the actual survey sample selected for any given web panel survey.

**Stage 3: Retention Rate.** This is the percentage of adults in connected households who are active and available for panel sampling at the time of drawing a given sample for a web panel survey. The current, within-household recruitment rate is approximately 50%. At any point in time, panel members on temporary leave (e.g., vacation) and panelists that have already completed four surveys in a given month are considered ineligible and are excluded from the sample.

**Stage 4: Survey Completion Rate.** This is the percentage of the panelists sampled for a given survey that actually complete the interview. The completion rate varies by subject type, questionnaire length, use of incentives, and by length of the field period. This response rate ranges can vary widely depending upon the study protocol and the study population. The range can be between 50% and 90% depending upon study-specific factors including the study topic, the length of the data collection field period, the use of incentives, the length of the survey, and sample composition. Surveys conducted using KnowledgePanel Latino sample have had a survey completion rate of 45% to 60%. This range is dependent upon survey topic, time in field, and use of survey incentives. For the calculation of a cumulative response rate, we recommend and Public Opinion Quarterly has also adopted the formula of taking into account these stages: Stage 1 Recruitment; Stage 2 Connection Rate; and Stage 4 Survey Completion Rate.

*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.*

Statistical methodology for stratification and sample selection.

In September 2007, Knowledge Networks was assigned a U.S. Patent (U.S. Patent No. 7,269,570) for its unique methodology for selecting online survey samples. The selection methodology, which has been used by KN since 2000, ensures that KN panel samples will closely track the U.S. population.

The selection methodology was developed by KN in recognition of the practical issue that different surveys target different subpopulations. Often, only panel members with certain characteristics are selected for a survey. This can skew the remaining panel sample and affect the sample representativeness of later surveys. The patented KN methodology also was developed to attempt to adjust or correct for nonresponse and noncoverage error in the panel sample.

In our patented solution, a survey assignment method uses a weighting factor to compensate for members which are temporarily removed from a panel because of an earlier draw of sample. This weighting factor adjusts the selection probabilities of the

remaining panel members. The sample is drawn using systematic probability-proportionate-to size sampling (PPS) where the panel post-stratification weights will be the measure of size (MOS). If the user requirements call for independent selection by stratum, the panel weights (MOS) are adjusted as follows: (1) sum the MOS for each stratum, call this sum  $S_h$  for stratum  $h$ ; (2) consider the user-specified or system-derived target sample size for stratum  $h$  to be  $n_h$ ; (3) multiply each MOS for members in stratum  $h$  by  $n_h/S_h$ ; and (4) use an interval of  $k=1$  and apply systematic PPS sampling to achieve the desired yield per stratum.

The above solution allows for representative samples to be drawn from the panel, even when earlier surveys oversampled different subpopulations. For an illustration, consider the following example. Suppose Study A requires a 100% oversample of Hispanics from the panel. At the beginning of the time period, each panel member will have an original selection weight making the panel selection distributions match the demographic benchmarks from the U.S. Census. After the sample draw for Study A is made, the new and temporary selection weights are calculated, making the panel selection distributions match the demographics of the general public. Consequently, the sample draw for Study B will yield a representative sample. Each demographic category in the remaining panel is monitored to ensure that there are enough members in each category to produce representative survey samples. The process is repeated for each study.

The implicit stratified systematic sample design has the additional benefit of correcting, in part, for nonresponse and noncoverage error introduced at the panel recruitment, connection, and panel retention stages of building and maintaining the panel. This correction is made possible by the fact that the selection weights are calculated using the latest Census Bureau (Current Population Survey) benchmarks for age, gender, race/ethnicity, and educational attainment. The samples are drawn using systematic PPS sampling where the panel post-stratification weights are the MOS, in an attempt to correct for under- and over-representation of certain demographic segments on the panel.

#### Estimation procedure.

This is a national probability sample and will be weighted according to our standard practice. This overview of our protocol was prepared by our Chief Statistician, Charles DiSogra. In a fashion similar to other probability-based samples, interview study samples drawn from KnowledgePanel® require a series of statistical adjustments to account for the probabilities of selection and also to render the sample more representative of the population of interest. These weights support analysis of the survey data for the purpose of estimating accurate and generalizable measures. These adjustments are accomplished through the application of statistical weights that function as a unique multiplier for each case in the study's interview data file. The computation of these weights is a complex task and addresses several features that are the necessary result of construction of the panel from recruitment through the assigned study-specific sample and finally the response to the study survey itself (see Callegaro and DiSogra, 2008). KnowledgePanel weights thus take into account what are in effect two distinct sampling segments: 1) panel recruitment and construction which are handled in a panel base weight and 2) elements related to the study sample incorporated in study-specific weights.

## Panel Base Weight

The recruitment of all panel members is based on a probability sample that had been traditionally sourced through a random-digit dial (RDD) sample frame but is now more recently sourced through an address-based sample frame (ABS) (see DiSogra, Callegaro, Hendarwan, 2009; Dennis, 2009). The national ABS recruitment sample provides superior coverage of the U.S. population including the rapidly growing proportion of households that do not have landline telephone service (almost entirely replaced by cell phones). Current KnowledgePanel members available for study samples had either been recruited from RDD methods or more recently from ABS mail methods. This makes the existing panel a dual-frame composition.

The weighting computation takes these differing frames and their related design features into account. This includes the different selection probabilities associated with the RDD- and ABS-sourced samples since a number of efficiency strategies are used in the respective recruitment samples, such as the oversampling of minority communities and of Spanish-language dominant Hispanic areas.

Additionally, there are several sources of error that are an inherent part of any complex survey process, for example, non-response to panel recruitment and panel attrition among recruited panel members. These and other sources of sampling and non-sampling error are corrected with a post-stratification adjustment using demographic distributions from the most recent data from the Current Population Survey (CPS). Other reliable benchmarks (not available in CPS) are employed for Spanish language usage and for household Internet access adjustments. The demographic variables used in this post-stratification task are gender, age, race/ethnicity, education, U.S. Census region, metropolitan area, Internet access, and language spoken at home (English/Spanish).

The results of all these adjustments go into the panel base weight that is used to make KnowledgePanel representative of the U.S. population of adults. This panel base weight is then employed in a probability proportional to size (PPS) selection method for drawing specific study samples from KnowledgePanel.

## Study-Specific Weights

Once a study sample has been drawn, fielded and the data compiled from all KnowledgePanel respondents, a sample-specific post-stratification process is carried out to adjust for survey non-response and for any elements related to the study-specific sample design (such as subgroup oversamples). An iterative raking procedure starting with the panel base weight is used for this task. Demographic and geographic distributions for the population ages 18+ from the most recent CPS provide the majority of the benchmarks for this adjustment. The demographic variables used are again gender, age, race/ethnicity, education, U.S. Census region, metropolitan area, Internet access, and, if appropriate for the study sample, language spoken at home (English/Spanish).

The distribution of these calculated weights are generally examined to identify and trim outliers at the extreme upper and lower tails of the weight distribution. These trimmed weights can be scaled to the nominal sample sizes to optimize statistical analyses as well as scaled to the generalizable population size for population count estimates.

## References

Callegaro, Mario and Charles DiSogra. 2008. Computing Response Metrics for Online Panels. *Public Opinion Quarterly*. 72(5) pp. 1008-1031.

Dennis, J. Michael. 2009. Summary of KnowledgePanel® Design. Available at Other methodological papers related to KnowledgePanel are available at <http://www.knowledgenetworks.com/ganp/reviewer-info.html>.

DiSogra, Charles, Mario Callegaro, and Erlina Hendarwan. 2009. [Recruiting Probability-Based Web Panel Members Using an Address-based Sample Frame](#): Results from a pilot study conducted by Knowledge Networks. Presented at the 2009 Joint Statistical Meetings. Full paper available at <http://www.knowledgenetworks.com/ganp/reviewer-info.html>.

Degree of accuracy needed for the purpose described in the justification.

The survey is designed to achieve appropriate national estimates of population parameters of plus or minus 3 percentage points at a 95 confidence level.

Unusual problems requiring specialized sampling procedures

There are none expected since this is a straightforward national probability sample.

Any use of periodic (less frequent than annual) data collection cycles to reduce burden.

This is a one-session national survey and there will not be any additional data collection undertaken for this specific project.

Additional information and resources can be found at:

<http://www.knowledgenetworks.com/ganp/reviewer-info.html>

Design Summary - Attachment A

**The consumer survey will be conducted using one questionnaire. The initial portion of the survey, designed to provide rigorous estimates of broadband characteristics of all American households at the national level, will consist of approximately 4,500 completed interviews with adults age 18 and older in the 50 states and the District of Columbia.**

*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.*

### Maximizing Response Rates

To maximize response rates, Knowledge Networks will send up to email reminders. Email reminders, in addition to allowing the respondent to choose the time of their response, have proven to achieve within-survey cooperation rates of 65% and greater.

Knowledge Networks employs its best practices for maximizing response rates. These measures can provide a survey completion rate or up to 70% and higher. If the study schedule, budget, and design are supportive of maximizing the survey completion rate, the following procedures are followed:

- Field period of 2 to 4 weeks
- Respondent incentives of \$5 to \$10 for participation, especially for surveys requiring 25 or more minutes of survey respondent time
  - Use of the Federal agency or University/College name in the email invitation
  - Email reminders
  - Telephone reminder calls to nonresponders

### Accuracy and Reliability

The recent paper from Stanford University professors Yeager & Krosnick demonstrate that the Knowledge Networks methodology provides greater accuracy-to-benchmarks than other online methodologies:

[http://www.knowledgenetworks.com/insights/KN\\_Breakfast-Forum-2009.html](http://www.knowledgenetworks.com/insights/KN_Breakfast-Forum-2009.html).

### Non-Response Bias Analysis

The Knowledge Panel contains more than 2,500 variables on each member, including a comprehensive set of demographics. These data enable us to know a great deal about those panelists that choose not to respond to any specific survey and these data form the core of the information used in creating non-response bias analyses. The key element in determining bias is whether or not the responders are very different from the non-responders on some important characteristic. These data make such detailed analyses possible.

If an analysis of the demographics of responders and nonresponders warrants, we will rigorously estimate model parameters corrected for non-response bias, along the lines of the econometric sample selection literature (see Heckman, 1979; Ozog and Waldman, 1996; Waldman, 1981). The steps for this analysis are:

- Fit a probit model of responders/nonresponders as a function demographic variables
- From the estimates of the probit model, construct a Heckman-type correction term (inverse Mill's ratio)

- Add the Heckman correction in the specification of the second stage regression of the utility function.

*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.*

Small scale tests of the questionnaire will be conducted with 8 individuals in a focus group setting in Boulder, CO. In addition, informal field tests will be conducted with those working with the principal investigators at the University of Colorado.

*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.*

Scott Wallsten, PhD  
Economics Research Director  
National Broadband Task Force  
Federal Communications Commission  
202-418-3632 (office)

Charles DiSogra  
Chief Statistician  
Knowledge Networks  
(650) 289-2000  
[cdisogra@knowledgenetworks.com](mailto:cdisogra@knowledgenetworks.com)



## ATTACHMENT A

1350 Willow Road, Suite 102, Menlo Park, CA 94025 • Phone 650.289.2000 • Fax 650.289.2001

# KnowledgePanel® Design Summary

This document was prepared at the effort and expense of Knowledge Networks. No part of it may be circulated, copied, or reproduced for distribution without the prior written consent of Knowledge Networks

© 2009 Knowledge Networks, Inc. 2

## KNOWLEDGEPANEL® OVERVIEW

KnowledgePanel®, created by Knowledge Networks (KN), is an online Non-Volunteer Access Panel, in which potential panel members are chosen via a statistically valid sampling method and using known published sampling frames that cover 99% of the U.S. population. Sampled non-internet households are provided a laptop computer and free internet service. KnowledgePanel consists of about 50,000 U.S. residents, age 18 and older, including cell phone-only households and those who are of Hispanic origin that were selected probabilistically. In addition, the KN panel also includes approximately 3,000 teens age 13 to 17 whose parents or legal guardians have provided consent. The panel size fluctuates because of the addition of panelists from the on-going recruitment and because of voluntary withdrawals and retirements of panelists reaching the end of their panel tenure.

## DUAL-FRAME SAMPLE RECRUITMENT METHODOLOGY

The first RDD recruitment to KnowledgePanel was conducted in 1999. At that time, all households recruited were given a WebTV to use for answering surveys. In August 2002, KN began allowing households to use their own computers connected to the Internet for taking surveys. Starting in January 2009, new Windows-based laptops were provided to non-Internet households instead of WebTV units.

Until recently, KnowledgePanel's probability-based recruitment had been based exclusively on a national RDD frame. In 2009, KN added address-based sample (ABS) frame to supplement the RDD frame in response to the growing number of cell-phone-only households that are outside of the RDD frame and in response to declining RDD response rate. ABS involves probability-based sampling of addresses from the U.S. Postal Service's Delivery Sequence File. Randomly sampled addresses are invited to join KnowledgePanel through a series of mailings (English and Spanish) and in some cases by telephone refusal conversion calls when a telephone number can be matched to the sampled address. Invited households can join the panel by one of several means: by completing and mailing back a paper form in a postage-paid envelope; by calling a toll-free hotline maintained by KN; or by going to a designated KN Web site and completing the recruitment form at the website.

For the RDD-based sampling, KN uses list-assisted RDD sampling techniques on the sample frame consisting of the entire U.S. residential telephone population. Knowledge Networks excludes only those banks of telephone numbers (each consisting of 100 telephone numbers) that have zero or one directory-listed phone number. Two strata are defined using 2000 Census Decennial Census data that has been appended to all telephone exchanges. The first stratum has a higher concentration of Black and Hispanic households, while the second stratum has a lower concentration of these groups relative to the national estimates. Telephone numbers are selected with equal probability of selection for each number within each of the two strata, with the higher concentration Black and Hispanic stratum

being sampled at approximately twice the rate of the other stratum. Sampling is done without replacement.

□ 2009 Knowledge Networks, Inc. 3

For the RDD recruitment, the households for which there is an address-matched telephone number are sent an advance mailing (in English and Spanish) informing them that they have been selected to participate in KnowledgePanel. Seven to nine days following the advance letter, the telephone recruitment process begins for sampled numbers. Cases sent to telephone interviewers are dialed up to 90 days, with at least 10 dial attempts on cases where no one answers the phone. Extensive refusal conversion is also performed. Experienced interviewers conduct all recruitment interviews. The recruitment interview, which typically requires about 10 minutes, begins with the interviewer informing the household member that they have been selected to join Knowledge Panel Latino or KnowledgePanel®.

In addition, in 2008 KN constructed KnowledgePanel Latino<sup>SM</sup> to provide researchers a capability to conduct representative online surveys with the U.S. Hispanic community. The sample for KnowledgePanel Latino is recruited by a hybrid telephone recruitment design, based on a national random-digit-dialing sample of U.S. Latinos and Hispanic-surname sample. It is a geographically balanced sample that covers areas that, when aggregated, encompasses approximately 93% of the nation's 45.5 million Latinos.

For all new panel members, demographic information such as gender, age, race/ethnicity, income, and education are collected in an online "profile" survey. This information is used to determine eligibility for specific studies and eliminates the need for gathering basic demographic information on each panel survey. Once this survey is completed, the panel member is regarded as active and ready to be sampled for other surveys. Recruits to KnowledgePanel Latino are asked in a separate survey about language usage and proficiency, language spoken at home and at work, media usage in Spanish and English, country of birth, and other topics.

### **PANEL SURVEY SAMPLING**

Once panel members are profiled, they become eligible for selection for specific surveys. The sample is drawn from eligible members using an implicitly stratified systematic sample design. Customized stratified random sampling based on profile data is also conducted, as required by specific studies.

In September 2007, KN was assigned a U.S. Patent (U.S. Patent No. 7,269,570) for its unique methodology for selecting online survey samples. The selection methodology, which has been used by KN since 2000, assures that KN panel samples will closely track the U.S. population.

The implicitly stratified systematic sampling methodology was developed by KN in recognition of the practical issue that different surveys target different subpopulations. Often, only panel members with certain characteristics are selected for a survey. This can skew the remaining panel sample and affect the representativeness of later survey samples. The methodology was also developed to attempt to correct for nonresponse and noncoverage error in the panel sample; see U.S. Patent No. 7,269,570 for more information.

□ 2009 Knowledge Networks, Inc. 4

### **SURVEY FREQUENCY & BURDEN**

To minimize panel attrition, surveys are usually kept short (from 5 to 20 minutes in length). For surveys requiring 16 or more survey minutes, survey participation is rewarded with a variety of incentives (small cash awards, gift prizes, raffle opportunities).

Further, steps are taken to ensure that panel members are not overburdened with survey requests. The primary sampling rule is to assign no more than one survey per week to members. This level of survey frequency helps to keep panelists engaged as part of the panel. On average, most KN panelists participate in about two surveys a month.

KN operates a Panel Relations program to encourage participation and member loyalty. Members can enter special raffles or be entered into special sweepstakes with both cash and other prizes to be won.

### **RESPONSE RATES**

As a member of AAPOR, KN follows the AAPOR standards for response rate reporting. However, the AAPOR standards were not established for web panels. KN survey scientists, Mario Callegaro and Charles DiSogra, recently published an article in *Public Opinion Quarterly* to provide a standard for web panel response rate calculations. See Callegaro and DiSogra (2008) for examples of response rates calculated for KnowledgePanel surveys and for details on the formula.

### **STATISTICAL WEIGHTING**

KnowledgePanel® sample begins as an equal probability sample that is selfweighting with several enhancements incorporated to improve efficiency. Since any alteration in the selection process is a deviation from a pure equal probability sample design, statistical weighting adjustments are made to the data to offset known selection deviations. These adjustments are incorporated in the sample's base weight.

There are also several sources of survey error that are an inherent part of any survey process, such as non-coverage and non-response due to panel recruitment methods and to inevitable panel attrition. These sources of sampling and nonsampling error are addressed using a panel demographic post-stratification weight as an additional adjustment.

However, prior to this adjustment, a separate sample of Spanish-speaking Latino panel members are weighted so as to be merged into the overall panel. This language-specific group is recruited through a geographically targeted dual frame sample that is screened for Spanish-language dominant households. The weighting of this unique sample involves a Spanish-language base weight that incorporates several adjustments including ones that address geographic frame and home language usage. The panel demographic post-stratification weight is then calculated for all panel members and proportionally adjusts for the merged Spanish-speakers.

© 2009 Knowledge Networks, Inc. 5

### **REFERENCES**

Callegaro, Mario & DiSogra, Charles (2008). Computing Response Metrics for Online Panels. *Public Opinion Quarterly*. 72(5) pp. 1008-1031.

Other methodological papers related to KnowledgePanel are available at <http://www.knowledgenetworks.com/ganp/reviewer-info.html>.