Field Report

Election 2008 & Beyond
Baseline Survey

Conducted for
University of Chicago

Submitted to Cathy Cohen, Ph.D.
November 12, 2008

1350 Willow Road, Suite 102
Menlo Park, CA 94025
P: 650-289-2160
F: 650-289-2001
www.knowledgenetworks.com

Knowledge Networks Methodology

Introduction

Knowledge Networks has recruited the first online research panel - KnowledgePanel[SM] - that is representative of the entire U.S. population. Panel members are randomly recruited by telephone and households are provided with access to the Internet and hardware if needed. Unlike other Internet research which covers only individuals with Internet access who volunteer for research, Knowledge Networks surveys are based on a sampling frame which includes both listed and unlisted numbers, and is not limited to current Web users or computer owners.

Knowledge Networks selects households using random digit dialing (RDD). Once a person is recruited to the panel, they can be contacted by e-mail (instead of by phone or mail). This permits surveys to be fielded very quickly and economically. In addition, this approach reduces the burden placed on respondents, since e-mail notification is less obtrusive than telephone calls, and most respondents find answering Web questionnaires to be more interesting and engaging than being questioned by a telephone interviewer.

KnowledgePanel[SM] Recruitment Methodology

Beginning recruitment in 1999, Knowledge Networks (KN) has established the first online research panel based on probability sampling that covers both the online and offline populations in the U.S. The panel members are randomly recruited by telephone and households are provided with access to the Internet and hardware if needed. Unlike other Internet research that covers only individuals with Internet access who volunteer for research, Knowledge Networks surveys are based on a sampling frame that includes both listed and unlisted phone numbers, and is not limited to current Web users or computer owners. Panelists are selected by chance to join the panel; unselected volunteers are not able to join the KN panel.

Knowledge Networks initially selects households using random digit dialing (RDD) sampling methodology. Once a household is contacted by phone and household members recruited to the panel by obtaining their e-mail address or setting up e-mail addresses, panel members are sent surveys over the Internet using e-mail (instead of by phone or mail). This permits surveys to be fielded quickly and economically, and also facilitates longitudinal research. In addition, this approach reduces the burden placed on respondents, since e-mail notification is less obtrusive than telephone calls, and allows research subjects to participate in research when it is convenient for them.

Knowledge Networks' panel recruitment methodology uses the quality standards established by selected RDD surveys conducted for the Federal Government (such as the CDC-sponsored National Immunization Survey).

Knowledge Networks utilizes list-assisted RDD sampling techniques on the sample frame consisting of the entire United States residential telephone population. Knowledge Networks excludes only those banks of telephone numbers (consisting of 100 telephone numbers) that have zero directory-listed phone numbers. 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 and the second stratum has a lower concentration relative to the national estimates. Knowledge Networks' telephone numbers are selected from the 2+ banks with equal probability of selection for each number within each of the 2 strata, with the Black and Hispanic stratum being sampled at a higher rate than the other stratum . Note that the sampling is done without replacement to ensure that numbers already fielded by Knowledge Networks do not get fielded again.

Telephone numbers for which Knowledge Networks is able to recover a valid postal address is about 60%-70%. The telephone phone numbers for which an address is recovered are selected with certainty; between one-half and one-third of the remainder were subsampled randomly depending on the recruitment period up until July 2005. In May 2007 subsampling was resumed at a rate of 0.75 of non-address households. The address-matched telephone numbers are sent an advance mailing informing them that they have been selected to participate in KnowledgePanel[SM].

Following the mailing, the telephone recruitment process begins for all sampled phone 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, and on phone numbers known to be associated with households. 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 KnowledgePanel[SM]. If the household does not have a PC and access to the Internet, they are told that in return for completing a short survey weekly, the household will be given a WebTV set-top box and free monthly Internet access. All members in the household are then enumerated, and some initial demographic variables and background information of prior computer and Internet usage are collected.

As of August 2002, those RDD households that inform interviewers that they have a home computer and Internet access have been recruited to the panel and asked to take their surveys using their own equipment and Internet connections. Points, which can be redeemed for cash at regular intervals, are given to respondents for completing their surveys and take the place of a free WebTV and monthly Internet access provided to other panel households. Additional incentive points may be added to specific surveys to improve response rates or to compensate for longer surveys.

Prior to shipment, each WebTV unit is custom configured with individual email accounts, so that it is ready for immediate use by the household. Most households are able to install the hardware

without additional assistance, though Knowledge Networks maintains a telephone technical support line and will, when needed, provide on-site installation. The Knowledge Networks Call Center also contacts household members who do not respond to e-mail and attempts to restore contact and cooperation. PC panel members provide KN with their email account and their weekly surveys are sent to that email account.

All new WebTV panel members are sent an initial survey to confirm equipment installation and familiarize them with the WebTV unit. For all new panel members, demographics such as gender, age, race, income, and education are collected in a follow-up survey for each panel member to create a member profile. This information can be used to determine eligibility for specific studies and need not be gathered with each survey. Once this survey is completed, the panel member is regarded as active and ready to be sampled for other surveys. Parental or legal guardian consent is also collected for conducting surveys with teenagers age 13-17 as part of the first survey.

Latino Panel Recruitment

The sample for the Knowledge Networks Latino Panel is recruited by a hybrid telephone recruitment design based on random-digit dial (RDD) probability-based landline sample of US Latinos and Hispanic-surname sample. It is a geographically balanced sample that covers areas that, when aggregated, encompass approximately 93% of the nation's 45.5 million Latinos.

The Latino Panel covers the entire U.S. 50 States and the District of Columbia. National coverage of Latinos is obtained via bi-lingual telephone recruitment using the KnowledgePanel RDD-based probability sampling. The KnowledgePanel recruitment, therefore, captures the full spectrum of assimilated and unassimilated Latinos, including Latinos whose primary language is English and others for whom it is Spanish. The Latino Panel, as a consequence, supports research whose results must be projectable to the full Latino population. The KnowledgePanel recruitment is conducted using time-tested methodology for gaining respondent cooperation: pre-notification letters, professional interviewers, and refusal conversion.

To increase the sample size of Latinos that are less assimilated or so-called "unassimilated," we use oversampling of Latinos residing in 70 U.S. DMAs that have relatively large Latino populations. The DMA-oversampling approach is dedicated to the recruitment of Spanish-Language-Dominant adults that are categorized as "unassimilated" on the basis of Hispanic self-identification, Spanish-language TV viewing frequency, and primary spoken language.

The 70 DMAs are grouped into 5 regions (Northeast, West, Midwest, Southeast, and Southwest). Each region is further divided into two groupings of census tracts, those that have a "high-density" Latino population and the balance made up of all the "low-density" census tracts. The threshold percent for "high density" varies by region. The 5 regions each divided into 2 density groups constitute 10 unique sample frames (5 x 2).

Using a geographic targeting approach, an RDD landline sample is generated to cover the high-density census tracts within each region. Due to the inaccuracy of telephone exchange coverage, there is some spillage outside these tracts and some smaller degree of non-coverage within these

tracts. About 32% of the Latino population across these five regions is theoretically covered with this targeted RDD landline sample. All the numbers generated are screened to locate a Latino household.

The remaining 68% of the Latinos in these five regions are addressed with a listed-surname sample. Listed surnames only include households where the telephone subscriber has a surname that has been pre-identified to likely be a Latino name. It is important to note that excluded from this low-density listed sample frame are: a) the mixed Latino/non-Latino households where the subscriber does not have a Latino surname, and b) all the unlisted landline Latino households. The percent of listed vs. unlisted varies widely at the DMA level. The use of the listed surname is intended to lower screening costs to locate a Latino household in these low-density areas since the rate of finding a Latino household from this list although not 100% is still very high.
Finding Latino households only means that the households are qualified (i.e., they have one or more Latino residents) but this does not yet mean that the household is eligible for DMA-based oversample of Spanish-Language-Dominant Latino Panel. To be eligible for the panel, the household member(s) must pass two screening tests. First, they need to speak Spanish in their home at one of three response levels: All the time, Most of the time, Equally with English. Next, they need to be able to read Spanish since all our communications and online surveys require a Spanish-reading ability. It is a complex task to identify eligible persons in households with multiple adults and especially in households where English and Spanish are spoken equally. In the latter case, for a household to further be eligible, that household must watch Spanish-language television for 5 or more hours per week.

This screening is designed to achieve a panel of Spanish-speaking and Spanish-reading Latinos who are a subset of all Latinos but who are unique due to the dominance of Spanish language as their mode of communication in their home environment. English-speaking Latinos are recruited by the existing RDD-based KnowledgePanel recruitment.

Based on the hybrid sample design of RDD and surname-listed sample and one the eligibility screening criteria, the DMA-based recruited sample is not 100% precisely representative of the Spanish-speaking Latino population (a most difficult standard to achieve under any design), but it is an excellent approximation of this population with good geographic balance across the nation.

Survey Administration

For client-based surveys, a sample is drawn at random from active panel members who meet the screening criteria (if any) for the client's study. The typical sample size is between 200 and 2000 persons, depending on the purpose of the study. Once selected, members can be sent an advance letter by email several days prior to receiving the questionnaire through their WebTV appliance or personal computer to notify them of an important, upcoming survey.

Once assigned to a survey, members receive a notification email on their WebTV or personal computer letting them know there is a new survey available for them to take. The email notification contains a button to start the survey. No login name or password is required. The

field period depends on the client's needs, and can range anywhere from a few minutes to two weeks.

Email reminders are sent to uncooperative panel members. If email does not generate a response, a phone reminder is initiated. The usual protocol is to wait at least three days and to permit a weekend to pass before calling. Knowledge Networks also operates an ongoing incentive program to encourage participation and create member loyalty. To assist panel members with their survey taking, each individual has a personalized "home page" that lists all the surveys that were assigned to that member and have yet to be completed.

Survey Sampling from KnowledgePanel[SM]

Once Panel Members are recruited and profiled, they become eligible for selection for specific surveys. In most cases, the specific survey sample represents a simple random sample from the panel. 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.

The primary sampling rule is not to assign more than six surveys per month to members with the expectation that on average four surveys a month will be completed by a panel member. In certain cases, a survey sample calls for pre-screening, that is, members are drawn from a sub-sample of the panel (e.g., females, Republicans). In such cases, care is taken to ensure that all subsequent survey samples drawn that week are selected in such a way as to result in a sample that is representative of the panel distributions.

Sample Weighting

The design for a KnowledgePanel[SM] sample begins as an equal probability sample that is self-weighting 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. We address these sources of sampling and non-sampling error using a **panel demographic post-stratification weight** as an additional adjustment.

Lastly, a set of **study-specific post-stratification weights** are constructed to adjust for sample design and survey non-response.

A description of these types of weights follows.

The Base Weight

In a Knowledge Networks panel sample, there are seven known sources of deviation from an equal probability of selection design. These are corrected in the Base Weight and are described below.

1. **Under-sampling of telephone numbers unmatched to a valid mailing address**

   An address match is attempted on all the Random Digit Dial (RDD) generated telephone numbers in the sample after the sample has been purged of business and institutional numbers and screened for non-working numbers. The success rate for address matching is in the 60-70% range. The telephone numbers with valid addresses are sent an advance letter, notifying the household that they will be contacted by phone to join KnowledgePanel. The remaining, unmatched numbers are under-sampled as a recruitment efficiency strategy. Advance letters improve recruitment success rates. Under-sampling stopped between July 2005 and April 2007. It was resumed in May 2007 with a sampling rate of 0.75.

2. **RDD selection proportional to the number of telephone landlines reaching the household**

   As part of the field data collection operation, information is collected on the number of separate telephone landlines in each selected household. A multiple line household's selection probability is down weighted by the inverse of its number of landlines.

3. **Some minor oversampling of Chicago and Los Angeles due to early pilot surveys**

   Two pilot surveys carried out in Chicago and Los Angeles when the panel was first being built increased the relative size of the sample from these two cities. With natural attrition and growth in size, the impact is disappearing over time. It remains part of our base adjustment weighting because of a small number of extant panel members from that nascent panel cohort.

4. **Early oversampling the four largest states and central region states**

   At the time when the panel was first being built, survey demand in the four largest states (California, New York, Florida, and Texas) required over-sampling during January-October 2000. Similarly, the central region states were over-sampled for a brief period. These now diminishing effects still remain in the panel membership and thus require weighting adjustments for these geographic areas.

5. Under-sampling of households not covered by the MSN® TV service network

Certain small areas of the U.S. are not serviced by MSN®, thus MSN®TV units cannot be used. We under-sample households in these areas and use other Internet Service Providers for their Internet access.

6. Oversampling of African- American and Hispanic telephone exchanges

As of October 2001, we began over-sampling telephone exchanges with a higher density of minority households (uniquely African American and Hispanic) to increase panel membership for those groups. These exchanges are oversampled at approximately twice the rate of other exchanges. This over-sampling is corrected in the base weight.

7. Selection of one adult in a household with two or more adults

For some samples, adult members are selected in two stages: households in the first stage and one adult per household in the second stage. A base weight selection correction is made by multiplying the selected adult by the inverse of the number of adults residing in the household. For many other samples, panel members can also be selected from the entire pool of members regardless of household affiliation. In this latter case, a base weight adjustment is not applied.

The Panel Demographic Post-stratification Weight

Once the study data are returned from the field, the final qualified respondent data are subjected to an additional post-stratification process to adjust for any non-response and non-coverage as a result of the study-specific sample design. Demographic and geographic distributions representing the study population within KnowledgePanel℠ are used as benchmarks for this adjustment.

The primary purpose of this post-stratification adjustment is to reduce the sampling variance for any characteristics highly correlated with the representative study population's demographic and geographic totals (these are referred to as the population benchmarks). This adjustment also helps reduce bias due to survey non-response. The following benchmark distributions were utilized for the current project:

- Gender: male/female
- Age: 18-34, 35-44, 45-59, 60+
- Race/ethnicity: White (non-Hispanic), Black (non-Hispanic), Asian (non-Hispanic), Hispanic
- Education: less than high school, high school graduates, some college, college graduates

For Hispanic respondents, language spoken in the home (Spanish only, mostly Spanish, Spanish and English equally, mostly English, English only) was also used in post-stratification.

Comparable distributions are calculated using all completed cases from the field data. Since study sample sizes are typically too small to accommodate a complete cross-tabulation of all the survey variables with the benchmark variables, an iterative proportional fitting is used for the post-stratification weighting adjustment. This procedure adjusts the sample data back to the selected benchmark proportions. Through an iterative convergence process, the weighted sample data are optimally fitted to the marginal distributions.

After this final post-stratification adjustment, the distribution of the calculated weights are examined to identify and, if necessary, trim outliers at the extreme upper and lower tails of the weight distribution. The post-stratified and trimmed weights are scaled to the sum of the total sample size.