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 at which point the subsampling was discontinued. 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.

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.

Weighting and Estimation

Whereas in principle the sample design is an equal probability design that is self-weighting, in fact there are several known deviations from this guiding principle. Furthermore, despite our efforts to correct for known sources of deviation from equal-probability design, there are several other sources of survey error that are an inherent part the process. We address these sources of survey error globally through the poststratification weights, which we describe below.

Sample Design Weights

The eight sources of deviation from an equal probability sample design are:

1. Subsampling of telephone numbers for which we could not find an address,
2. Random Digit Dial sampling rates proportional to the number of phone lines in the household,
3. Minor oversampling of Chicago and Los Angeles due to early pilot surveys in those two cities,
4. Short-term double-sampling the four largest states (CA, NY, FL, and TX) and central region states,
5. Under-sampling of households not covered by MSN® TV,
6. Oversampling of minority households (Black and Hispanic),
7. Oversampling of households with personal computers and access to the Internet, and
8. Selection of one adult per household.

A few words about each feature:

1. Once the telephone numbers have been purged and screened, we address match as many of these numbers as possible. The success rate so far has been in the 60%-70% range. The telephone numbers with addresses are sent an advance letter, notifying the household that they will be contacted by phone to join the Knowledge Networks panel. The remaining, unmatched numbers were subsampled in the past in order to reduce costs. As of July 2005, we no longer subsample unmatched numbers.

2. As part of the field data collection operation, we collect information on the number of separate phone lines in the selected households. We correspondingly down-weight households with multiple phone lines.

3. Two pilot surveys carried out in Chicago and Los Angeles increased the relative size of the sample from these two cities. The impact of this feature is disappearing as the panel grows, but we still include it as part of our correction process.

4. Since we anticipated additional surveying in the four largest states, we double-sampled these states during January-October 2000. Similarly, the central region states were over-sampled for a brief period.

5. Certain areas of the U.S. are not serviced by MSN®. We select a smaller sample of phone numbers in those areas and use other Internet Service Providers for Internet access of recruited households in those areas.

6. As of October 2001, we began oversampling minority households (Black and Hispanic) to increase panel capacity for those subgroups.

7. As of August 2002, we began oversampling households with personal computers and access to the Internet to reduce the panel costs associated with distributing, set-up and maintenance of MSN®TV units.

8. For most of our surveys, we select panel members across the board, regardless of household affiliation. For some surveys, however, we select members in two stages: households in the first stage and one adult per household in the second stage. We correct for this feature by multiplying the probabilities of selection by $1/a_i$ where a_i represents the number of adults (18 and over) in the household.

To reduce the effects of potential nonresponse and noncoverage bias in panel estimates, a cell post-stratification adjustment is applied to panel weights after accounting for sample design factors using demographic distributions from the most recent data from the Current Population Survey. The post-stratification variables include age, race, gender, Hispanic ethnicity and education are applied prior to selection of any client sample from the Knowledge Networks panel. These weights constitute the base starting weights for any client survey selected from the Knowledge Networks panel.

Statistical 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 for each study's completed data to adjust for sample design and survey non-response.

A description of these types of weights follows.

The Base Weight

In a KnowledgePanel[SM] sample, there are six 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[SM]. The remaining, unmatched numbers are under-sampled as a recruitment efficiency strategy. Advance letters improve recruitment success rates. This 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 of 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 our MSN®TV units cannot be used for recruited non-Internet households. In some of these cases, we use other Internet Service Providers for Internet access via the member's personal computer. Overall, the result is a small under-sample of these geographic areas thus requiring a minor weighting adjustment.

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 (specifically 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.

The Panel Demographic Post-stratification Weight

To reduce the effects of any non-response and non-coverage bias in panel estimates, a post-stratification raking adjustment is applied using demographic distributions from the most recent data from the Current Population Survey (CPS). The post-stratification variables include age, race, gender, Hispanic ethnicity and education. This weighting adjustment is applied prior to the selection of any client sample from KnowledgePanel[SM]. These weights constitute the starting weights for any client survey selected from the panel.

Study-Specific Post-Stratification Weights

Once all the data from a client study are returned from the field, we proceed with a post-stratification process to adjust for any survey non-response and also any non-coverage due to the study-specific sample design. Demographic and geographic distributions representing the study population within KnowledgePanel[SM] are used as benchmarks for this adjustment.

The following benchmark distributions are generally utilized for this type of post-stratification adjustment:

- Gender (Male, Female)
- Age (18-29, 30-34, 35-44, 45-59, 60+)
- Race/Hispanic ethnicity (White/Non-Hispanic, Black or African American/Non-Hispanic, Other/Non-Hispanic, 2+ Races/Non-Hispanic, Hispanic)
- Education (Less than high school, High school, Some college, College graduates)
- Census Region (Northeast, Midwest, South, West)
- Metropolitan Area (Yes, No)
- Household Internet Access (Yes, No)

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.