

Knowledge Networks Methodology

Introduction

Knowledge Networks has recruited the first online research panel that is designed to be representative of the entire U.S. population. The panel is representative because it is recruited using high quality probability sampling techniques, and is not limited to current Web users or computer owners. Knowledge Networks selects households using random digit dialing (RDD) and provides selected households with free hardware and Internet access. This allows surveys to be administered using a Web browser and enables the inclusion of multimedia content. 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.

Panel Recruitment Methodology

Knowledge Networks' panel recruitment methodology uses the quality standards established by the best Random Digit Dialing (RDD) surveys conducted for the Federal Government.

Knowledge Networks utilizes list-assisted RDD sampling techniques on the sample frame consisting of the entire United States telephone population. The sample frame is updated quarterly. Knowledge Networks excludes only those banks of telephone numbers (consisting of 100 telephone numbers) that have zero directory-listed phone numbers. Knowledge Networks' telephone numbers are selected from the 1+ banks with equal probability of selection for each number. Note that the sampling is done without replacement to ensure that numbers already fielded by Knowledge Networks do not get fielded again.

Having generated the initial list of telephone numbers, the sample preparation system excludes confirmed disconnected and non-residential telephone numbers. Next, the sample is screened to exclude numbers that are not in the WebTV Internet Service Provider network. This process results in the exclusion of approximately 6% to 8% of the United States population. This percentage is diminishing steadily and as of July 2001, we will begin to include a small sample from the out of WebTV Internet Service Provider network in the panel to represent these areas and reduce coverage error.

Telephone numbers for which Knowledge Networks is able to recover a valid postal address (about 50%) are sent an advance mailing informing them that they have been selected to participate in the Knowledge Networks Panel. In addition to information about the Knowledge Networks Panel, the advance mailing also contains a monetary incentive to encourage cooperation when the interviewer calls.

Following the mailing, the telephone recruitment process begins. The numbers called by interviewers consist of all numbers sent an advance mailing, as well as 50% of the numbers not sent an advance mailing. The resulting cost efficiency more than offsets the decrease in precision caused by the need for sample weights. Cases sent to telephone interviewers are dialed up to 90 days, with at least 15 dial attempts on cases where no one answers the phone, and 25 dial attempts on phone numbers known to be associated with households. Extensive refusal conversion is also performed.

Experienced interviewers conduct all recruitment interviews. An interview, which typically requires about 10 minutes, begins with the interviewer informing the household member that they have been selected to join the Knowledge Networks Panel. 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.

To ensure consistent delivery of survey content, each household is provided with identical hardware, even if they currently own a computer or have Internet access. Microsoft's WebTV is the hardware platform currently used by the Knowledge Networks panel. The device consists of a set-top box that connects to a TV and the telephone. It also includes a remote keyboard and pointing device. WebTV has a built-in 56K modem that provides the household with a connection to the Internet. The base unit also has a small hard drive to accommodate large file downloads, including video files. File downloads do not require any user intervention and usually occur during off hours.

Prior to shipment, each 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.

All new panel members are sent an initial survey to confirm equipment installation and familiarize them with the WebTV unit. Demographics such as gender, age, race, income, and education are collected for each participant to create a member profile. This information can be used to determine eligibility for specific studies and need not be gathered with each survey.

Starting in August 2002, Knowledge Networks began oversampling households that have pre-existing, home-based Internet access, allowing these households to participate in the web-enabled panel using their own equipment and Internet access. These panel members are enrolled into a panel loyalty program intended to encourage long-term participation in the research panel.

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 mail several days prior to receiving the questionnaire through their WebTV appliance to notify them of an important, upcoming survey.

Once assigned to a survey, members receive a notification email on their WebTV 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 Panel

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 one survey per week to members. 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 sample drawn that week are selected in such a way as to result in a sample that is representative of the panel distributions. Furthermore, Panel Members are not assigned surveys on the same topic in a given three-month period.

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 of the process. We address these sources of survey error globally through the post-stratification weights, which we describe below.

Sample Design Weights

The seven sources of deviation from epsem design are:

1. Half-sampling of telephone numbers for which we could not find an address,
2. RDD 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. 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 50-60% range. The telephone numbers with addresses are sent a letter. The remaining, unmatched numbers are half-sampled in order to reduce costs. Based on previous research we suspect that the reduced field costs resulting from this allocation strategy will more than offset increases in the design effect due to the increased variance among the weights. We are currently quantifying these balancing features.
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.
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. Finally, 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.

Post-stratification Weights

The primary purpose of a post-stratification adjustment to survey weights is to reduce the sampling error for characteristics highly correlated with reliable demographic and geographic totals – called population benchmarks. To implement post-stratification, we used the following raking variables:

- gender: male, female
- age: 18-29, 30-44, 45-59, 60 and over
- race/ethnicity: white (nonhispanic), black (nonhispanic), other (nonhispanic), hispanic
- region: northeast, midwest, south, west
- education - highest level achieved: less than high school, high school, some college, college degree or more

In order to calculate final weights, we derive weighted sample distributions along various combinations of the above variables. Similar distributions are calculated using the most recent U.S. Census Bureau's Current Population Survey data and the Knowledge Networks panel data. Cell-by-cell adjustments over the various univariate and bivariate distributions are calculated to make the weighted sample cells match those of the U.S. Census and the Knowledge Networks panel. This process, known as raking, is repeated iteratively until there is convergence between the weighted sample and benchmark distributions (CPS distributions). Occasionally, collapsing of post-stratification cells is necessary. This is dependent on the size of the sample and topology of the sample universe.

The variable “weight” is the final post-stratification weight for all completed cases.