GfK Methodology

Introduction

The GfK Group (formerly Knowledge Networks) is passionate about research in marketing, media, health and social policy – collaborating closely with client teams throughout the research process, while applying rigor in everything we do. We specialize in innovative online research that consistently gives leaders in business, government, and academia the confidence to make important decisions. GfK delivers affordable, statistically valid online research through KnowledgePanel® and leverages a variety of other assets, such as world-class advanced analytics, an industry-leading physician panel, an innovative platform for measuring online ad effectiveness, and a research-ready behavioral database of frequent supermarket and drug store shoppers.

GfK has recruited the first online research panel that is representative of the entire U.S. population. Panel members are randomly recruited through probability-based sampling, and households are provided with access to the Internet and hardware if needed.

GfK recruits panel members by using address-based sampling methods [formerly GfK relied on random-digit dialing methods]. Once household members are recruited for the panel and assigned to a study sample, they are notified by email for survey taking, or panelists can visit their online member page for survey taking (instead of being contacted by telephone or postal mail). This allows surveys to be fielded very quickly and economically. In addition, this approach reduces the burden placed on respondents, since email notification is less intrusive than telephone calls, and most respondents find answering Web questionnaires more interesting and engaging than being questioned by a telephone interviewer. Furthermore, respondents have the freedom to choose what time of day to complete their assigned survey.

Documentation regarding KnowledgePanel sampling, data collection procedures, weighting, and IRB-bearing issues are available at the below online resources.

- http://www.knowledgenetworks.com/ganp/reviewer-info.html
- http://www.knowledgenetworks.com/knpanel/index.html
- http://www.knowledgenetworks.com/ganp/irbsupport/

The GfK Group

The GfK Group has a strong tradition in working with sophisticated academic, government, and commercial researchers to provide high quality research, samples, and analyses. The larger GfK Group offers the fundamental knowledge for governmental agencies, academics, industries, industry, retailers, services companies and the media need to provide exceptional quality in research to make effective decisions. It delivers a comprehensive range of information and consultancy services. GfK is one of the leading survey research organizations worldwide, operating in more than 100 countries with over 11,000 research staff. In 2010, the GfK Group's sales amounted to EUR 1.29 billion.

For further information, visit our website: www.gfk.com.

Panel Recruitment Methodology

When GfK began recruiting in 1999 as Knowledge Networks, the company established the first online research panel (now called KnowledgePanel®) based on probability sampling covering both the online and offline populations in the U.S. Panel members are recruited through national random samples, originally by telephone and now almost entirely by postal mail. Households are provided with access to the Internet and a netbook computer, if needed. Unlike Internet convenience panels, also known as "opt-in" panels, that include only individuals with Internet access who volunteer themselves for research, KnowledgePanel recruitment has used dual sample frames to construct the existing panel. As a result, panel members come from listed and unlisted telephone numbers, telephone and non-telephone households, and cell phone only households, as well as households with and without Internet access, which creates a representative sample. Only persons sampled through these probability-based techniques are eligible to participate on KnowledgePanel. Unless invited to do so as part of these national samples, no one on their own can volunteer to be on the panel.

RDD and ABS Sample Frames

KnowledgePanel members today may have been recruited by either the former random digit dialing (RDD) sampling or the current address-based sampling (ABS) methodologies. In this section, we will describe the RDD-based methodology; the ABS methodology is described in a separate section below. To offset attrition, multiple recruitment samples are fielded evenly throughout the calendar year.

KnowledgePanel recruitment methodology has used the quality standards established by selected RDD surveys conducted for the Federal government (such as the CDC-sponsored National Immunization Survey).

GfK employed list-assisted RDD sampling techniques based on a sample frame of the U.S. residential landline telephone universe. For purposes of efficiency, GfK excluded only those banks of telephone numbers (a bank consists of 100 numbers) that had fewer than two directory listings. Additionally, an oversampling was conducted within a stratum of telephone exchanges that had high concentrations of African American and Hispanic households based on Census data. Note that recruitment sampling was done without replacement, thus numbers attempted in earlier waves were excluded from subsequent recruitment waves.

A telephone number for which a valid postal address can be matched occurred in about 67-70% of each sample at the time RDD was being used for recruitment. These address-matched cases were all mailed an advance letter informing them that they had been selected to participate in the KnowledgePanel. For purposes of efficiency, the numbers without a matched-physical address were under-sampled at a rate of 0.75 relative to the address-matched numbers. Both the minority oversampling mentioned above and this under-sampling of non-address households are accounted for appropriately in the in the panel's weighting procedures.

Following the mailings, telephone recruitment by professional interviewers/recruiters began for all sampled telephone numbers. Telephone numbers for cases sent to recruiters were dialed for up to 90 days, with at least 14 dial attempts for cases in which no one answers the phone, and for numbers known to be associated with households. As occurs for most telephone interviews, for those participants who were hesitant or expressed a soft refusal, extensive refusal conversion was also performed. The recruitment interview, about 10 minutes in length, begins with informing the household member that the household had been selected to join KnowledgePanel. If the household did not have a computer and access to the Internet, the household member is told that, as reward for completing a short survey weekly, the household will be provided with free monthly Internet access and a laptop computer (in the past, the household was provided with a WebTV device, currently, netbooks are provided). All members of the household were enumerated, and some initial demographic and background information on prior computer and Internet use was collected.

Households that informed recruiters that they had a home computer and Internet access were asked to take GfK surveys using their own equipment and Internet connection. Incentive points per survey, redeemable for cash, are given to these "PC" (personal computer) respondents for completing their surveys. Panel members provided with a laptop computer and free Internet access do not participate in this per-survey points-incentive program. However, all panel members receive special incentive points for selected surveys to improve response rates and/or for all longer surveys as a modest compensation for the extra burden of their time and participation.

For those panel members receiving a laptop or netbook computer, each unit is custom-configured prior to shipment 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, although GfK maintains a toll-free telephone line for technical support. The GfK Call Center contacts household members who do not respond to email and attempts to restore both contact and participation. PC panel members provide their own email addresses, and we send their weekly survey invitations to that email account.

All new panel members receive an initial survey for the dual purpose of welcoming them as new panel members and introducing them to how online survey questionnaires work. New panel members also complete a separate profile survey that collects essential demographic information such as gender, age, race, income, and education to create a personal member profile. This information can be used to determine eligibility for specific studies and is factored in for weighting purposes. Operationally, once the profile information is stored, it does not need to be re-collected as a part of each and every survey. This information is also updated annually for all panel members. Once new members have completed their profile surveys, they are designated as "active," and considered ready to be sampled for client studies. [Note: Parental or legal guardian consent is also collected for the purpose of conducting surveys with teenage panel members, aged 13 to17.]

Once a household is recruited and each household member's email address is either obtained or provided, panel members are sent survey invitations linked through a personalized email message (instead of by phone or postal mail). This contact method permits surveys to be fielded

quickly and economically, and also facilitates longitudinal research. In addition, this approach reduces the burden placed on respondents, since email notification is less intrusive than telephone calls and allows research subjects to participate in research when it is convenient for them.

Address-Based Sampling (ABS) Methodology

When GfK first started panel recruitment in 1999, the conventional opinion among survey experts was that probability-based sampling could be carried out cost effectively through the use of national RDD samples. The RDD landline frame at the time allowed access to 96% of U.S. households. Due to the rapid rise of cell phone-only households, this is no longer the case. In 2009, GfK first used address-based sampling for panel recruitment to reflect the changes in society and telephony over recent years. Some of the factors that have reduced the long-term scientific viability of landline RDD sampling methodology are as follows: declining respondent cooperation in telephone surveys, the development of "do not call" lists to reduce unsolicited commercial calls, call screening with caller-ID devices, increased use of answering machines; dilution of the RDD sample frame as measured by the working telephone number rate (more fax lines and lines dedicated to other functions), and finally, the emergence of cell phone only households (CPOHH), since these households are typically excluded from the RDD frame because they have no landline telephone.

According to the Centers for Disease Control and Prevention (January-June 2011), approximately 33.6% of all U.S. households cannot be contacted through RDD sampling—31.6% as a result of CPOHH status and 2% because they have no telephone service whatsoever. Among some age segments, the RDD non-coverage would be substantial: 47% of young adults, ages 18–24, reside in CPOHHs, 58% of those 25–29 years old, and 46% of those who are 30–34.[1]

After conducting an extensive pilot project in 2008, GfK made the decision to move to use an address-based sample (ABS) frame in response to the growing number of cell phone only households that are outside the RDD frame. Before conducting the ABS pilot, we also experimented with supplementing RDD samples with cell phone samples. However, this approach was found to be much more costly, and raised a number of other operational, data quality, and liability issues (for example, calling cell phones while respondents were driving).

The key advantage of the ABS sample frame is that it allows sampling of almost all U.S. households. An estimated 97% of households are "covered" in sampling nomenclature. Regardless of household telephone status, those households can be reached and contacted through postal mail. Second, the GfK ABS pilot project revealed several additional advantages beyond expected improvement in recruiting adults from CPOHHs:

[1] Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, January–June 2011. National Center for Health Statistics. December 2011. Available from: http://www.cdc.gov/nchs/nhis.htm.

- Improved sample representativeness for minority racial and ethnic groups
- Improved inclusion of lower educated and lower income households
- Exclusive inclusion of the fraction of CPOHHs that have neither a landline telephone nor Internet access (approximately four to six percent of US households).

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 and, in some cases, telephone follow-up calls to non-responders when a telephone number can be matched to the sampled address. Operationally, invited households have the option to join the panel by one of several ways:

- Completing and returning a paper form in a postage-paid envelope,
- Calling a toll-free hotline maintained by GfK, or
- Going to a dedicated GfK web site and completing an online recruitment form.

After initially accepting the invitation to join the panel, respondents are then "profiled" online by answering key demographic questions about themselves. This profile is maintained through the same procedures that were previously established for RDD-recruited panel members. Respondents not having an Internet connection are provided a laptop or netbook computer and free Internet service. Respondents sampled from the ABS frame, like those sampled from the RDD frame, are offered the same privacy terms and confidentiality protections that we have developed over the years and that have been reviewed and approved by dozens of Institutional Review Boards.

Large-scale ABS sampling for KnowledgePanel recruitment began in April 2009. As a result, sample coverage on KnowledgePanel of CPOHHs, young adults, and minority population groups has been increasing steadily since that time.

Because current KnowledgePanel members have been recruited over time from two different sample frames, RDD and ABS, GfK implemented several technical processes to merge samples sourced from these frames. KN's approach preserves the representative structure of the overall panel for the selection of individual client study samples. An advantage of mixing ABS frame panel members in any KnowledgePanel sample is a reduction in the variance of the weights. ABS-sourced samples tend to align more closely to the overall demographic distributions in the population, and thus the associated adjustment weights are somewhat more uniform and less varied. This variance reduction efficaciously attenuates the sample's design effect and confirms a real advantage for study samples drawn from KnowledgePanel with its dual frame construction.

Survey Sampling from KnowledgePanel

Once panel members are recruited and profiled, they become eligible for selection for client surveys. In most cases, the specific survey sample represents a simple random sample from the panel, for example, a general population survey. Customized stratified random sampling based on profile data can also be conducted as required by the study design.

The general sampling rule is to assign no more than one survey per week to individual members. Allowing for rare exceptions during some weeks, this limits a member's total assignments per month to four or six surveys. In certain cases, a survey sample calls for pre-screening, that is, members are drawn from a subsample of the panel (such as females, Republicans, grocery shoppers, etc.). 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 remains representative of the panel distributions.

Survey Administration

Once assigned to a survey, members receive a notification email letting them know there is a new survey available for them to take. This email notification contains a link that sends them to the survey questionnaire. No login name or password is required. The field period depends on the client's needs and can range anywhere from a few hours to several weeks.

After three days, automatic email reminders are sent to all non-responding panel members in the sample. If email reminders do not generate a sufficient response, an automated telephone reminder call can be initiated. The usual protocol is to wait at least three to four days after the email reminder before calling. 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.

GfK also operates an ongoing modest incentive program to encourage participation and create member loyalty. Members can enter special raffles or can be entered into special sweepstakes with both cash rewards and other prizes to be won.

The typical survey commitment for panel members is one survey per week or four per month with duration of 10 to 15 minutes per survey. In the case of longer surveys, an additional incentive is typically provided.

Sample Weighting

The design for KnowledgePanel® recruitment begins as an equal probability sample 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 by using a **panel demographic post-stratification weight** as an additional adjustment.

All the above weighting is done before the study sample is drawn. Once a study sample is finalized (all data collected and a final data set made), a set of **study-specific post-stratification**

weights are constructed so that the study data can be adjusted for the study's sample design and for survey non-response.

A description of these types of weights follows.

The Base Weight

In a KnowledgePanel sample there are eight 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 to 70% range. Households having telephone numbers with valid addresses are sent an advance letter, notifying them 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 was suspended between July 2005 and April 2007. It was resumed in May 2007, using a sampling rate of 0.75. RDD recruitment ended in July 2009.

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. The probability of selecting a multiple-line household is down-weighted by the inverse of the number of landlines. RDD recruitment ended in July 2009.

3. Some minor oversampling of Chicago and Los Angeles in early pilot surveys

   Two pilot surveys carried out in Chicago and Los Angeles when the panel was initially being built increased the relative size of the sample from these two cities. With natural attrition and growth in size of the overall panel, that impact has declined over time. It remains part of our base adjustment weighting because of a small number of extant panel members from that initial 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) necessitated oversampling during January–October 2000. Similarly, the central region states were oversampled for a brief period of time. These now diminishing effects still remain in the panel membership and thus weighting adjustments are required 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 the MSN®TV units (Web-TV) distributed to non-Internet households prior to January 2009 could not be used for those recruited non-Internet households. Overall, the result is a small residual under-sample in those geographic areas which requires a minor weighting adjustment for those locations. Since January 2010, laptop computers with dial-up access are being distributed to non-Internet households, thus eliminating this under-coverage component.

6. RDD oversampling of African American and Hispanic telephone exchanges

As of October 2001, oversampling of telephone exchanges with a higher density of minority households (specifically, African American and Hispanic) was implemented to increase panel membership for those groups. These exchanges were oversampled at approximately twice the rate of other exchanges. This oversampling is corrected in the base weight. RDD recruitment ended in July 2009.

7. Address-based sample phone match adjustment

Toward the end of 2008, GfK began recruiting panel members by using an address-based sample (ABS) frame in addition to RDD recruitment. Once recruitment through the mail, including follow-up mailings to ABS non-respondents was completed, telephone recruitment was added. Non-responding ABS households where a landline telephone number could be matched to an address were subsequently called and telephone recruitment was initiated. This effort resulted in a slight overall disproportionate number of landline households being recruited in a given ABS sample. A base weight adjustment is applied to return the ABS recruitment panel members to the sample's correct national proportion of phone-match and no phone-match households.

8. ABS oversample stratification adjustment

In late 2009 the ABS sample began incorporating a geographic stratification design. Census blocks with high density minority communities were oversampled (Stratum 1) and the balance of the census blocks (Stratum 2) were relatively under-sampled. The definition of high density and minority community and the relative proportion between strata differed among specific ABS samples. In 2010, the two strata were redefined to target high density Hispanic areas in Stratum 1 and all else in Stratum 2. In 2011, pre-identified ancillary information and not census block data were used to construct and target four strata as follows: Hispanic ages 18-24, Non-Hispanic ages 18-24, Hispanic ages 25+ and Non-Hispanic ages 25+. An appropriate base weight adjustment is applied to each relevant sample to correct for these stratified designs. Also in 2011, a separate sample targeting only persons ages 18-24 was fielded across the year also using predictive ancillary information. Combined with the four-stratum sample, the base weight adjustment compensates for cases from this unique young adult over-sample. In 2012, a similar four-stratum design is used but the ages have been changed to 18-29 and 30+ for both the Hispanic and Non-Hispanic strata.

The Panel Demographic Post-stratification Weight

To reduce the effects of any non-response and non-coverage bias in the overall panel membership (before the study sample is drawn), a post-stratification adjustment is applied based on demographic distributions from the most recent data from the Current Population Survey (CPS). The benchmark distributions for Internet access among the U.S. population of adults are obtained from the most recent special CPS supplemental survey measuring Internet access (October 2010).

The overall panel post-stratification variables include:

- Gender (Male/Female)
- Age (18–29, 30–44, 45–59, and 60+)
- Race/Hispanic ethnicity (White/Non-Hispanic, Black/Non-Hispanic, Other/Non-Hispanic, 2+ Races/Non-Hispanic, Hispanic)
- Education (Less than High School, High School, Some College, Bachelor and beyond)
- Census Region (Northeast, Midwest, South, West)
- Household income (under $10k, $10K to <$25k, $25K to <$50k, $50K to <$75k, $75K to <$100k, $100K+)
- Home ownership status (Own, Rent/Other)
- Metropolitan Area (Yes, No)
- Internet Access (Yes, No)

The Panel Demographic Post-stratification weight is applied prior to a probability proportional to size (PPS) selection of a study sample from KnowledgePanel. This weight is designed for sample selection purposes.

Study-Specific Post-Stratification Weights

Once the sample has been selected and fielded, and all the study data are collected and made final, a post-stratification weight is computed to adjust for any survey non-response as well as any non-coverage or under- and over-sampling resulting from the study-specific sample design. Demographic and geographic distributions for the non-institutionalized, civilian population ages 18+ from the most recent CPS are used as benchmarks in this adjustment.

The following benchmark distributions are utilized for this post-stratification adjustment**:**

- Gender (Male/Female)
- Age (18–29, 30–44, 45–59, and 60+)
- Race/Hispanic ethnicity (White/Non-Hispanic, Black/Non-Hispanic, Other/Non-Hispanic, 2+ Races/Non-Hispanic, Hispanic)
- Education (Less than High School, High School, Some College, Bachelors and higher)
- Household income (under <$25k, $25K to <$50k, $50K to <$75k, $75K+)
- Census Region (Northeast, Midwest, South, West)
- Metropolitan Area (Yes, No)

- Internet Access (Yes, No)

Comparable distributions are calculated by 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, a raking procedure is used for the post-stratification weighting adjustment. Using the base weight as the starting weight, 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 then scaled to the sum of the total sample size of all eligible respondents.