

## KNOWLEDGEPANEL<sup>®</sup>

### Recruitment and Sample Survey Methodologies

#### Rigor:

KnowledgePanel is the largest online panel that is representative of the adult US population. Our recruitment process employs an addressed-based sampling methodology from the *Delivery Sequence File* of the USPS – a database with full coverage of all delivery points in the US. As such, samples from KnowledgePanel cover all households regardless of their phone status, allowing GfK provide fully representative online samples to the research community. This methodological rigor is backed by our experts in survey research methods and applications who work closely with our clients throughout the field period and data delivery. Relying on uncompromised statistical methodologies, survey results from KnowledgePanel samples are often used for government and academic research purposes and publications in peer-reviewed journals.

#### Expertise:

By all measures – clients, services, professionals, and industry recognition – GfK is a leading research organization. Relying on an impressive cadre of seasoned statisticians and survey research scientists, our tried and proven methodologies are supported by an organizationally mandated quest for remaining at the cutting edge of relevant technologies. Our experts have successfully transitioned a myriad of survey and market research projects from phone and mail modalities to an online data collection alternative via KnowledgePanel. These include the design and administration of longitudinal and tracking surveys with complex reporting needs.

#### Recruitment:

The recruitment process for KnowledgePanel was originally based exclusively on a national *Random Digit Dialing* (RDD) sampling methodology. In order to improve the representativeness of the panel, in 2009 GfK migrated to an *Address-Based Sampling* (ABS) methodology for recruitment purposes. This probability-based sampling methodology improves population coverage, particularly for hard-to-reach individuals such as young adults and minority subgroups. GfK continues to provide a web-enabled device such as a tablet computer and free Internet service for adults recruited from households without Internet access.

#### Address-Based Sampling (ABS):

The DSF-based sampling frame we use for address selection is enhanced with a series of refinements – such as the appendage of various ancillary data to each address – to facilitate complex stratification plans<sup>1</sup>. Taking advantage of such refinements, quarterly samples are selected using a disproportionate stratified sampling methodology across the following four strata:

- Stratum 1: Hispanic households with at least one 18 to 24 year-old
- Stratum 2: Remaining Hispanic households
- Stratum 3: Remaining households with at least one 18 to 24 year-old
- Stratum 4: All remaining households

Adults from sampled households are invited to join KnowledgePanel through a series of mailings, including an initial invitation letter, a reminder postcard, and a subsequent follow-up letter. Given that a subset of physical addresses can be matched to a corresponding landline telephone number, about 5 weeks after the initial mailing, telephone refusal-conversion calls are made to nonresponding households for which a telephone number is matched. Invited households can join the panel by:

---

<sup>1</sup> Fahimi, M. and D. Kulp (2009). “Address-Based Sampling – Alternatives for Surveys That Require Contacts with Representative Samples of Households.” *Quirk’s Marketing Research Review*, May 2009.

- Completing and mailing back a paper form in a postage-paid envelope
- Calling a toll-free hotline phone number maintained by GfK
- Going to a secure GfK website to complete the recruitment form online

During the initial recruitment survey, attempts are made to recruit every household member who is at least 13 years of age to become an active member. For teenage household members, consent is secured from a parent or legal guardian, and no direct communication with teenagers is attempted prior to obtaining consent. While surveys can be conducted with teens directly, in most instances teen surveys are conducted by first selecting a sample of active members who are parents. This parent route alternative, while slightly more expensive, makes it possible to reach a more representative sample of teens.

### **KnowledgePanel Latino:**

In 2008, KnowledgePanel Latino (KPL) was created to provide researchers with the capability to conduct representative online surveys with Hispanics, including those who speak only Spanish. KPL members are recruited using a dual-frame RDD sampling methodology targeting telephone exchanges associated with areas with a higher concentration of Hispanics. Households are screened in Spanish to recruit only those homes where Spanish is spoken at least half the time. These members supplement Hispanics recruited through general ABS recruitment sample who take surveys in English or Spanish.

### **Sampling from KnowledgePanel:**

Once panel members are recruited and profiled, they become eligible for selection for client surveys. For general population surveys, a representative sample is drawn from the entire panel. Customized stratified random sampling based on profile data can also be carried out as required by a study design. The general sampling rule is to assign no more than one survey per week to individual members, but to allow for occasional weeks when certain panel members receive more invitations.

### **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 custom link that sends members to the survey questionnaire without requiring any further login or password verifications. Each active member has a personalized *home-page* that lists all the surveys that have been assigned to that member and have yet to be completed.

The field period for surveys depends on the client's needs and can range from a few hours to several weeks. In order to secure high rates of response, however, email and telephone calls are used to encourage nonresponding panel members to partake in surveys they are assigned to. Typically, email reminders are sent after three days and phone calls are initiated about four days later. GfK provides a modest incentive to encourage participation and foster member loyalty. Members can enter special raffles or be entered into special sweepstakes with both cash rewards and other prizes. Most surveys take about 10 to 15 minutes, however, for longer surveys an additional incentive is typically provided.

### **KnowledgePanel Profile Data:**

One of the unique features of KP that provides a powerful tool for probability-based sample surveys of rare subgroups is a rich set of profile data for all 55,000 recruited panel members. Before recruited members are considered *Active* and eligible for participation in client surveys, they must take an in-depth *Core Profile* survey that gathers detailed geodemographic information about each member and his/her household. In addition, there are twelve *Topical Profile* surveys that panel members take on a rolling basis over the course of the calendar year. These surveys cover a wide-range of topics including shopping habits, media usage, health conditions, sports interests and participation, voting behavior and political ideology, and familiarity with and usage of technology. Newly recruited members are asked to complete these surveys soon after they become active, and existing panel members are asked to retake all profile surveys each year in order to maintain an up-to-date profile of panelists.

Our profile data provide a cost-effective approach for selecting representative samples from low-incidence populations based on geodemographic, behavioral, and attitudinal indicators. Any of the profile variables can be appended to surveys, eliminating the need for inclusion of such questions in client surveys. Such data can also be used to calculate incidence rates and weighting benchmarks for subpopulations that are not covered by government surveys, such as the Current Population Survey (CPS) or American Community Survey (ACS). As described below we now have passively collected digital profiles for a segment of KnowledgePanel.

### **KnowledgePanel Digital:**

In the 21<sup>st</sup> Century digital world, the type of data that can lead to actionable intelligence is evolving. Given the growing amount of time individuals spend online for work, school, and shopping, market and survey research endeavors can no longer rely only on structured data that can be collected via survey sampling. Some of the key behavioral and attitudinal nuances that cannot be measured via regimented questionnaires must be captured in a passive fashion with the adult members' permission and full knowledge of what data is being collected.

GfK is now able to collect and report a wide range of digital data obtained from a subset KP members who have mobile devices – smartphones and tablets- as they visit websites, use apps and search the Internet. By combining active survey data with passive digital tracking, KnowledgePanel Digital provides the type of nuanced data needed for a comprehensive assessment of consumers in the digital age. Through this GfK leads the way in applying behavioral and attitudinal linkages for social scientists, the Federal government and market researchers.

As with other KnowledgePanel -based surveys, those involving KnowledgePanel Digital can bring along a rich supplement of profile data available for all members. Furthermore, data from GfK's MRI<sup>®</sup> consumer marketing and media database can be appended, which include in depth view of the US adult consumers with respect to their media choices, demographics, lifestyles and attitudes, and usage of almost 6,000 products in 550 categories.

### **Sampling Design Weights:**

Significant resources and infrastructure are devoted to the recruitment process for KnowledgePanel so that the panel properly represents the adult population of the US. This representativeness is not only achieved with respect to a broad set of geodemographic distributions, but also hard-to-reach adults – such as those without landline telephone or Spanish language dominant individuals. Consequently, the natural distribution of KnowledgePanel mirrors that of the US adults fairly closely, barring disparities that emerge for certain subgroups due to differential attrition rates among recruited panel members.

For selection of general population samples a patented methodology has been developed that ensures the resulting samples behave as EPSEM (equal probability selection method). This methodology starts by weighting the entire KP to the detailed geodemographic benchmarks of US adults from the latest March supplement of the CPS. This ensures that the weighted distribution of KnowledgePanel perfectly matches that of the US adults.

Using the weights as the measure of size (MOS) for each panel member, a PPS (probability proportional to size) procedure is used to select study-specific samples. The application of this PPS methodology with the above MOS values produces fully self-weighting samples, for which each sample member can carry a design weight of unity. In instances where the study design requires any oversampling of specific subgroups, departures from an EPSEM design are corrected by adjusting the corresponding design weights, with the CPS benchmarks serving as reference distributions.

### **Study-Specific Analysis Weights:**

Virtually all survey data are weighted before they can be used to produce reliable estimates of population parameters. While reflecting the selection probabilities of sampled units, weighting also compensates for

practical limitations of a sample survey, such as differential nonresponse and undercoverage. Furthermore, by taking advantage of auxiliary information about the target population, weighting ensures that respondents to a survey better represent the corresponding target population.

Once the study sample has been selected and fielded, and all the survey data are cleaned, design weights are adjusted to compensate for any differential nonresponse that may have occurred during the data collection process. Final analysis weights are produced using an iterative proportional fitting (raking) procedure to ensure that the resulting weights are aligned with respect to all study benchmark distributions simultaneously. In the last step, calculated weights are examined to identify and, if necessary, trim outliers at the extreme upper and lower tails of the weight distribution. The resulting weights are then scaled to sum to the total sample size of all eligible respondents.

### KnowledgePanel Calibration:

Studies that need a large number of respondents, or those that focus on rare subpopulations, can require sample sizes that KnowledgePanel may not be able to provide. In such instances, a blended sample from both KnowledgePanel and nonprobability online panels can provide an effective alternative that raises the quality of the data. For such applications, our cutting-edge Calibration process is used to correct for biases due to systematic undercoverage associated with the nonprobability samples from online panels. These coverage problems include omission of non-internet households, as well as overrepresentation of hyper internet users. As compared to samples that exclusively rely on non-probability samples without any calibration, our calibrated weights enable the resulting blended sample to represent the target population more effectively and offer more robust inferential possibilities. This improved representativeness is not only with respect to geodemographic distributions, but also with respect to an important set of attitudinal/behavioral measures.

Based on repeated studies, it has been demonstrated that respondents from nonprobability and probability-based samples differ in several ways. For instance, respondents from nonprobability samples have a higher propensity for adopting new products and services. Moreover, our ongoing research has identified additional measures with respect to which significant differences exist. As such, our new calibration methodology aims to realign such respondents from nonprobability samples with respect to a multidimensional set of measures that are significant differentiators between the two types of respondents. The following chart shows a sampling of such differentiators.

