

Painting a truer picture

The Internet revolution taking place in the marketing information business is just hitting its stride. Today, about 10 percent of research is conducted online — but online market share is growing quickly. All indications are that, in the next few years, online will account for about 70 percent of research dollars. As this change occurs, the amount of core and high-stakes research migrating to the Internet will also rise. While there will always be roles for other media — particularly where they suit research needs better than the Internet — online custom research will certainly be the most common future methodology for informing business decisions.

At this pivotal time, as Internet research moves from a new method to a core method, important questions must be investigated:

Is data from a representative online panel more valid?

- How can online research be reliably integrated with business decisions?
- What are the opportunities and the pitfalls in applying online research to decision-making?
- What are the differences between higher- and lower-quality Internet research approaches?

All of these issues must be under-

stood by users as we go forward. Correspondingly, other factors that affect online marketing information quality — such as the environment for recruiting, maintaining and retaining respondents — are also at a tipping point, and changing even faster than the business uses of online research. Clutter, spam and slowing growth of computer and Internet penetration are all creating dramatic transformations in the online world — and more change is on the horizon. As a result, users of online research must not only be cognizant of how to choose and apply online research today, but also must know how to move through the chaotic future we are facing.

Through an ongoing research-on-research program, our firm is committed to understanding how, in the context of better decision-making, researchers, marketers and social scientists all must adapt to the evolving consumer and online environments. We must challenge ourselves to use the Internet revolution not just as a change in data collection method, but as an opportunity to maximize research's value for more informed consumer-based marketing decisions. But to make this important leap, we must first gain a thorough understanding of the variety of online research tools and their effects.

Quality dictates

Research quality should be defined by its impact on the business decision. Research outcomes are determined by the quality of the sample, the research design, or the analysis and interpretation of results. This assessment deals mostly with sample quality, which varies



By John J. Lewis

Editor's note: John J. Lewis is president and CEO of Knowledge Networks, a Menlo Park, Calif. research firm. He can be reached at 650 289-2050 or at jlewis@knowledgegenetworks.com.

widely in the online world. Simply put, sample quality dictates the soundness of results. So, while supporting decisions with good research requires more than high sample quality, it is a necessary pre-condition if one wishes to project findings beyond the specific set of respondents used in a given study. Thus, sample quality represents the most logical place to start assessing online studies and their business impact.

Most online research respondents come from one of two sources:

- Volunteer panels and e-mail lists: Consumers who have volunteered to participate in ongoing research; most are obtained via pop-up ads, e-mail blasts, word of mouth, list purchases.

- Volunteer rivers: Recruited by many of the same techniques as panels and lists, these respondents have volunteered for one-time research only.

Two things are common to both of the above types of sample. First, neither of the populations are defined by the researcher (anyone and everyone can sign up and participate in research conducted in this manner). This poses security as well as methodological risks. Second, neither of these methods is based on random selection — so respondents, in essence, choose the research(er) rather than the other way around.

An alternative to these methods is the representative approach, as embodied by the Knowledge Networks Panel. The panel is based on a representative sample of the full U.S. population — both Internet and non-Internet households. Sample members are initially recruited by phone; those who do not have Internet access are given it free of charge. The result is an online sample that represents all U.S. households.

In any research, there are three primary types of bias associated with sample source:

- coverage bias: excluding certain groups from the universe of potential respondents;
- self-selection bias: allowing respondents to be “self-chosen” (“volunteerism”), rather than using a pre-designated sample; and
- non-response bias: receiving survey responses from only some of those who were invited to participate.

Depending on the methods employed, these sources of bias can have a dramatic effect on survey results. The job of the

	Coverage of U.S. Population	Panel Recruitment Cooperation	Survey Response
KN Panel	96 percent (Non-telephone HH not covered)	39 percent	50 percent-90 percent (7- to 30-day fielding period)
Volunteer Internet	57 percent (Non-Internet HH not covered)	.02 percent	15 percent-35 percent (7- to 30-day fielding period)
RDD Telephone	96 percent	N/A	25-50 percent

researcher is to minimize, as much as possible, these biases. Thus, the “quality” of sample should be judged with reference to these types of bias. A brief description of each follows.

Coverage bias: Findings from Pew’s highly regarded “Internet and American Life” studies show that 61 percent of the U.S. population goes online from any location to access the Internet or to send and receive e-mail; this level has remained constant for well over a year. This means that, for the foreseeable future, any sample that is derived wholly from the Internet population — as volunteer online samples are — will exclude 40 percent of the population. The fact that there is no universal database of e-mail addresses also means that there is no way to create a probability sample of the 60 percent of consumers who are online. And, since Internet use is skewed to upper-income, white, highly educated people, this segment will be over-represented in online-only samples.

By contrast, 96 percent of the population has a telephone number; thus, a sample recruited from a universe of telephone numbers will exclude only 4 percent of the population.

Self-selection bias: Opt-in samples are, by definition, self-selected; and those who volunteer to participate in research represent a narrow sub-group of the Internet population — possibly more opinionated, and more interested in the topic of the survey or in answering surveys in general. When the researcher, via random sampling or other techniques, chooses samples then self-selection bias is eliminated.

Non-response bias: If a representative research sample has been chosen in advance, the quality of findings will depend on the ability to obtain answers from as many potential respondents as possible. In the world of telephone research, this means maximizing the number of surveys obtained from a pre-

selected sample of telephone numbers.

Volunteer Internet surveys do not have a predefined sample; their respondent pools consist of whoever has seen a given pop-up ad or received an e-mailed solicitation. Since this universe is almost never defined in volunteer surveys, it is impossible to prompt participation through follow-up messages.

Using a representative approach minimizes the three sources of bias as follows:

- Coverage: Because the sample for the representative panel is based on the telephone universe, only 4 percent of the population is excluded.
- Self-selection: Respondents are selected via random sampling; no self-selection is involved.
- Non-response: In our case, we make aggressive attempts to build a community — to recruit its full sample to the panel and to maintain their participation through incentives, newsletters and other techniques.

Table 1 compares average levels of success for typical volunteer Internet research, RDD telephone studies, and the Knowledge Networks Panel.

Some researchers have suggested that bias caused by opt-in samples can be essentially eliminated after the fact through weighting and balancing. But such adjustments can only be made according to a standard demographic profile that cannot reproduce the unpredictable ways that a truly representative sample would respond to a given set of survey questions. And our research has shown that, even with sample balancing, results from opt-in surveys can still differ widely from representative research — a fact we illustrate below.

Put it in context

The value of research quality cannot be established or understood in the abstract; it must be tied to specific interpretations and business decisions. Are the quality

variations from one survey to the next large enough to change actual decisions about products, ads, policies or commodities? Placing quality differences in this context demonstrates their significance — or lack thereof — and gives clients a sense of how best to deploy different types of research.

To shed light on this issue, we used two surveys as opportunities to compare volunteer research to that produced by a representative sample (the Knowledge Networks Panel). In both cases, after the proprietary surveys had been completed by KN, we obtained volunteer sample from prominent vendors and fielded the same instrument among those groups. (For more information on these studies, see Pineau & Slotwiner, “Probability Samples vs. Volunteer Respondents in Internet Research: Defining Potential Effects on Data and Decision-Making in Marketing Applications.” Knowledge Networks, 2003.)

In one case, a study of alcohol product potential among young men was fielded with two different opt-in samples, yielding three data sources. In the other, research on acceptance and liking of a new line of facial products was conducted among one volunteer group, as well as the KN Panel. In both cases, the volunteer samples produced data that was consistently different — often to extremes — from the representative KN data, and would have led to different marketing decisions. And in both cases, the client affirmed that the representative KN findings were more in line with expected levels and thus would be considered the definitive findings.

In the alcohol research, we first examined the data to see whether or not the different vendors produced comparable levels of on- and off-premise alcohol consumption among men ages 21 to 27. Second, we evaluated whether the data exhibited the same relative distributions across the vendors. Finally, we investigated the extent to which weighting the volunteer samples changed the results of the first two analyses. In all cases, we found that the data from the volunteer groups differed significantly from the KN data.

Chart 1 summarizes the estimates derived from the first question in our survey: “Which of the beverages listed below have you consumed in your own home or someone else’s home in the

past month?” Data are presented for the KN sample and from one of the volunteer list vendors. (We could not use the data from the second source for this chart because that vendor would only provide data for the qualified completes [hard liquor drinkers], and this table is based on all completes.)

The average difference between the representative and volunteer estimates is about 16 percent. Reproducing this chart for the second question — “Which of the beverages listed below have you consumed outside the home, for example in bars, clubs or restaurants in the past month?” — yields nearly identical results (the mean difference in the estimates was 12 percent). Chart 2 summarizes these differences by category for the measures of on- and off-premises alcohol consumption.

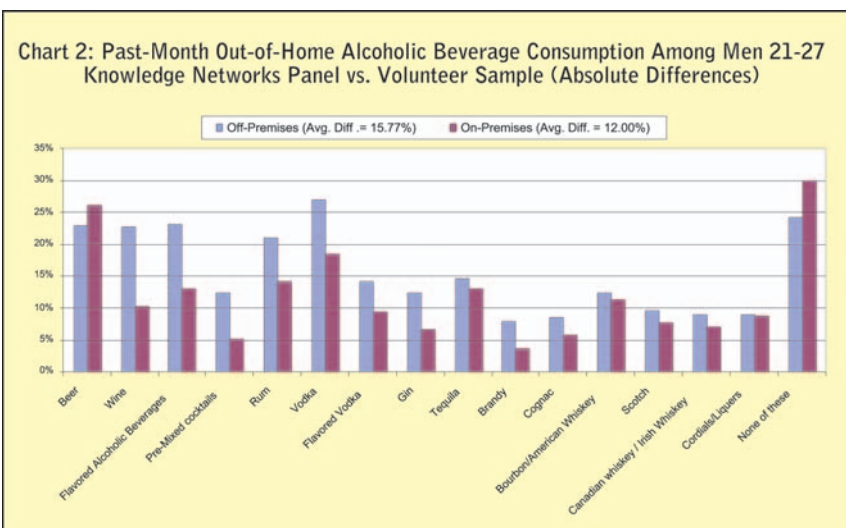
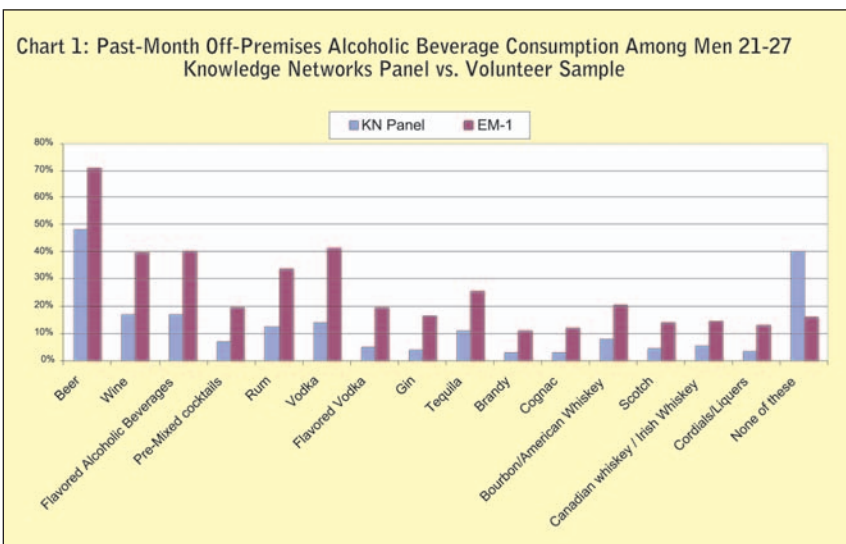
The measures of on- and off-premises consumption, then, show clear disparities across the two groups — differences

large enough to influence decision-making. The pattern of higher consumption in the volunteer group holds true for practically every type of alcoholic beverage examined in the study.

If the client were trying to estimate the overall size of the market for flavored alcoholic beverages and intended to invest heavily in product development if the size of the market exceeded 25 percent of males age 21 to 27, the final business decision would have been different depending on whether they relied on KN panel data or volunteer respondents. The result could have been development and launching of a product whose demand was significantly lower than expected.

Similar results

We obtained similar results — with equal potential to impact marketing decisions — when comparing volunteer and representative findings for the facial products



study. The study was intended to supply answers to important questions such as:

- Who are the early triers of the new product line (in terms of demographics and behaviors)?

- Does the profile of early users match initial marketing efforts for the line, or does the targeting need to be refined?

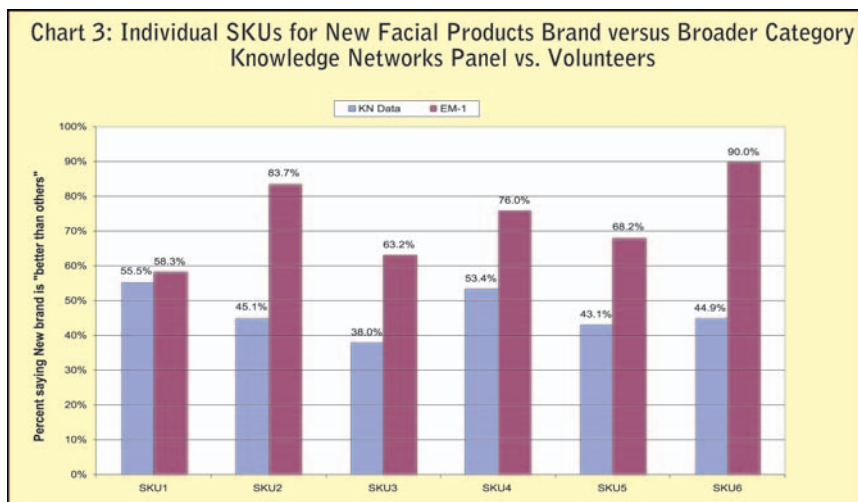
To answer these questions, it is critical to obtain accurate measures of incidence and reliable data on the demographic profiles of users. If the target group is not, ultimately, representative of all new users, then measures of product performance and benefits may be skewed as well.

The KN Panel data indicate that U.S. marketplace penetration for the line is about 1.2 percent; that is, 1.2 percent of those surveyed said they used the product one or more times per week. This figure is appropriate for an HBA product six months post-launch in a highly fragmented category. The volunteer group, however, yielded substantially different measures of brand usage; the line penetration number was 3.3 percent — nearly two-and-a-half times the representative panel data estimate.

Other differences were clearly evident between the two samples. For example, the profiles of brand users described by the two sources of data show notable deviations. Early brand triers identified via the volunteer study appear to be younger, more educated, wealthier, and have more children than their KN data counterparts; they are also less likely to be African-American or Hispanic. All of these skews match the classic profile of Internet users.

Chart 3 shows the data differences between representative and volunteer samples in terms of those liking the new products “better than others.”

The disparities between the estimates of brand usage could lead the manufac-



turer to draw dramatically different conclusions about the success of its product. As these data are used to make explicit decisions regarding advertising and marketing efforts, overestimation of the brand’s success could lead to decreased marketing support when, in fact, performance is average for the category. Such a mistaken decision could have serious consequences for the brand.

The extreme difference in the profiles of early adopters painted by the two data sources could also yield divergent business decisions. Understanding the basic characteristics of early adopters is key to assessing and refining marketing strategy. It affects all decisions related to the purchasing of targeted media for marketing and communications efforts. As a result, misunderstanding this group — as would likely happen with the volunteer data — could lead to an unwarranted change in course, or lack thereof.

Unlimited potential

The Internet offers almost unlimited potential as a marketing and public policy information source, and is fully suited to core studies if appropriate methods and quality checks are deployed. Studies like those discussed above

demonstrate that different online research resources and techniques can yield very different results and change business decisions; knowing the effects of these factors is essential to taking full advantage of all that the Internet has to offer.

We will continue to explore quality issues of all kinds related to research — an initiative that will take on growing significance as the online business and research environment continues to transform. Pop-up blockers, anti-spam technologies, the growing number of companies soliciting respondents via the Web, and government regulation all are poised to transform the Internet experience. Their effects demand monitoring and adjustment among those who wish to fulfill the Web’s potential as a market research medium.

The path to informed business decisions — by researchers and information suppliers alike — requires that the characteristics and business value of quality research become a high-profile issue in the industry. Understanding how to use and unlock the power of data obtained from the Internet is perhaps the most important opportunity — and challenge — in market research today. | Q

Reprinted with permission from Quirk’s Marketing Research Review, January 2004.

© Quirk Enterprises, Inc. All Rights Reserved. On the Web at www.quirks.com.

FosteReprints: 1-866-879-9144



1350 Willow Road, Suite 102
Menlo Park, CA 94025
(605) 289-2000

web: www.knowledgenetworks.com
email: info@knowledgenetworks.com