

# THE BUSINESS VALUE OF QUALITY ONLINE RESEARCH

by Daniel Slotwiner

Critical decisions demand attention to  
the methods behind consumer insights

With its unprecedented melding of interactivity, speed, and graphical complexity, the Internet was destined to become a rich vehicle for consumer insight. Now, as comfort with the medium grows, the online environment is being leveraged for marketing studies of greater complexity. Is Internet research ready for its big moment?

The quality of any research must be proportional to the importance of the decision; but too often, online research is viewed as being of one variety and quality level—potentially creating a false sense of either comfort or danger. In fact, “online research” is a general term that has been applied to a wide range of methodologies that may have little in common; being versed in those nuances is essential to

making sound marketing decisions with Internet-based learning. Because each business issue or decision is unique, finding an optimal solution will never be accomplished with a “one-size-fits-all” approach.

A research philosophy that aims to deliver the highest-quality insights to support business decisions must include a deep understanding of the various methodologies—choices that will have a profound impact on final results *and business decisions*. In this article, we detail exclusive Knowledge Networks research that demonstrates the value of high-quality techniques, proving that they can make the difference between a *yes* or *no* decision on core marketing questions.

**TYPICAL INTERNET RESEARCH: SUBSET OF A SUBSET**

Sometimes it's easy to forget that, for many consumers, the Internet is a nonfactor in their everyday lives. For nearly two years now, the proportion of households in the U.S. *without* Internet access has remained at about 40 percent. The online population is still quite different from the U.S. as a whole—more affluent, more Northeastern, and less ethnically diverse.

Consider, as well, that nearly all online research relies on consumers with online access who have *volunteered* to participate in either a single study or multiple studies (as part of a panel). Thus, typical Internet insights represent a small subset of a subset of all U.S. consumers. In some cases, this narrow focus may be desirable; in others, it can prove a decided handicap. Choosing an approach that is appropriate to the decision at hand is essential to wise marketing decisions.

**CHOOSING A RESEARCH APPROACH THAT IS APPROPRIATE TO THE DECISION AT HAND IS ESSENTIAL TO WISE MARKETING DECISIONS.**

There are three primary approaches used in Internet research: volunteer panels, volunteer “river” samples, and representative panels. Volunteer panels and “river” samples are recruited via pop-up ads, e-mail spam, and other mass-reach techniques; “river” respondents typically take just one survey,

while panelists are contacted repeatedly. In both cases, those answering surveys have “self-selected” into the sample and, consequently, have no clear relationship to the population as a whole (in the parlance of statistics, the sampling mechanism is unknown); thus, their answers cannot be projected to a wider group.

The Knowledge Networks panel uses the third technique—choosing a random sample of the full U.S. population and recruiting by telephone. If a given household does not have Internet access, it is provided so that they can participate. The result is Internet research that is truly representative of the U.S. as a whole.

**RESEARCH BIAS IN ACTION:  
A DIFFERENCE MEASURABLE IN DOLLARS**

To demonstrate the impact that volunteer methodologies can have on market research—and business decisions—Knowledge Networks has conducted several head-to-head comparisons. When the same study is conducted on a representative sample and a self-selected “volunteer” sample, are there significant differences? What

would be the impact on decisions of the different data?

One comparison involved research conducted by Knowledge Networks on behalf of a well-known alcohol manufacturer. This client commissioned an incidence study focusing on consumption

TABLE 1: STUDY OVERVIEW

	Rep (KN) sample	EM-1	EM-2
Total Completed	541	318	Unknown
Total Qualified	202	219	168
Rate of Qualification	37%	69%	Unknown

Due to the “river” methodology employed to collect the EM-2 sample, it is not possible to calculate a qualification rate.

of alcohol at home and away from home to assess market opportunities. In addition to the original study on the representative Knowledge Networks panel, two volunteer e-mail lists (EM-1 and EM-2) also received virtually the same questionnaire. These lists came from well-known suppliers in the marketing research industry, and from companies that typically supply other research companies with sample. All three samples were exclusively of males in the 21 to 27 age range—a demographic with a high level of Internet penetration.

Table 1 shows one of the most frequent side effects of volunteerism in research—those who are already interested in the subject are often disproportionately represented. Almost two out of three EM-1 respondents qualified for the survey (i.e., consumed at least one beer at or away from home in the past month), while less than two out of five Knowledge Networks panelists “made the cut.” It may seem that this makes the EM-1

sample more efficient—but what it really shows is that the e-mail sample was biased, a fact that will have important consequences for the survey results.

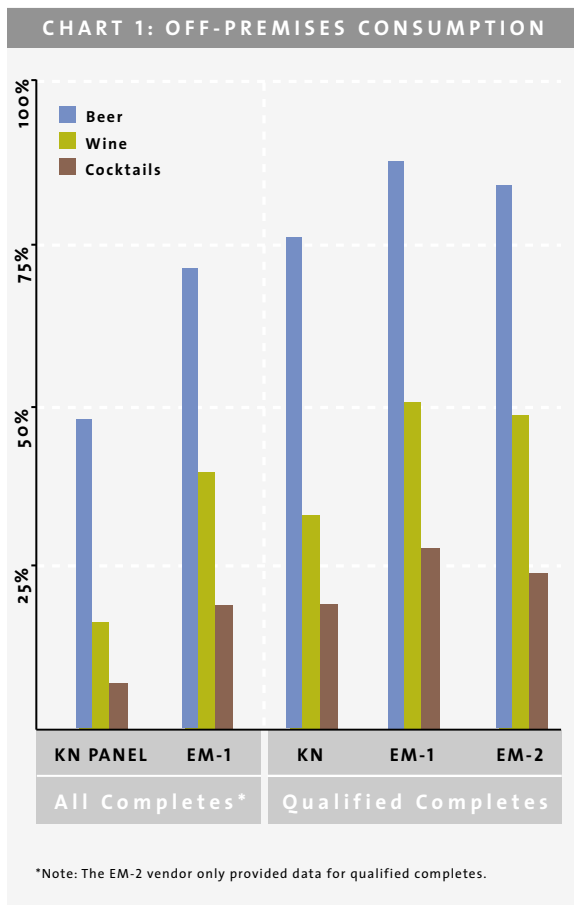
We compared the survey responses from these sample sources using three perspectives:

- levels of on- and off-premises alcohol consumption
- relative distributions across the vendors
- the degree to which after-the-fact “weighting” changed the answers from the EM-1 and EM-2 samples  
*(Some researchers assert that, through weighting, they can adjust for the differences between representative and volunteer samples.)*

In all the data we collected and the analyses we performed, the Internet volunteer samples yielded higher levels of alcohol usage and awareness than the representative Knowledge Networks sample. These results are consistent with the traditional notions that strictly volunteer samples reflect the bias of their volunteer participants.

CLIENT EXPERIENCE AFFIRMS REPRESENTATIVE FINDINGS

Chart 1 (following page) shows out-of-home alcohol consumption for the Knowledge Networks panel and the EM-1 sample, with an average difference of 16 percent between the two. The EM-1 levels are more than double the Knowledge Networks levels in several cases, and the Knowledge Networks



panelists were twice as likely to say they have consumed *none* of these beverage types away from home.

Imagine that you are a senior marketer who has only the EM-1 figures to go by; what you see is robust product demand that bolsters optimism for new product launches in several key categories. The problem is that, according to other sources available to the client, the representative data is more realistic; the EM-1 results are most likely skewed by volunteerism.

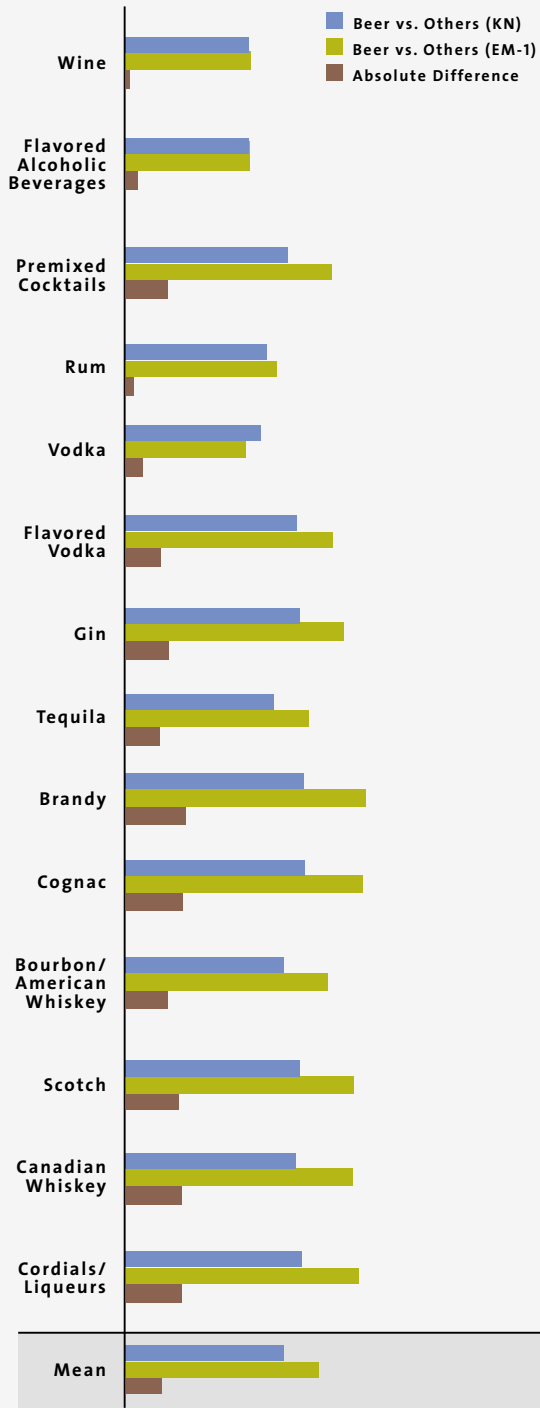
An examination of just the qualified completes yields similar results. Here the differences are seen across all three vendors. Again, it is important to stress that the differences seen here are, most likely, due to volunteerism. We can largely eliminate coverage (the lack of representation in the EM-1 and EM-2 samples of people without Internet access) since the level of Internet penetration among males in the 21 to 27 age group is very high. Similarly, the right portion of *Chart 1* is based on qualified completes only, which should further reduce any differences inherent in the sample.

Although the divergence in estimates is minimized when one restricts the analysis to qualified completes, it does not disappear, and the basic findings—that EM-1 and EM-2, on balance, provide higher levels of off-premises consumption than the representative Knowledge Networks data—hold.

#### DIFFERENT PERSPECTIVE, SAME RESULTS

It could be argued that these differences, however, are not of primary importance to the marketing issues at hand. This argument states that absolute values are of little consequence—that what actually matters is *relative differences* (i.e., “within sample variance”). By this measure, too (see *Chart 2*), there are marked differences between the Knowledge Networks panel and the other data sources—with the representative findings in line with client expectations from other research. For instance, the Knowledge Networks sample shows that beer consumption exceeds that of premixed cocktails by nearly 41 percent,

**CHART 2: RELATIVE DIFFERENCES BETWEEN BEER CONSUMPTION AND OTHER CATEGORIES BY VENDOR**



while EM-1 shows a 51 percent difference. Thus, there is a 10-percentage-point difference between the two estimates of the same statistic.

In fact, the estimates from the two sources (for off-premises consumption) are only correlated at .7, suggesting that they vary in concert but not perfectly. When one compares beer consumption to each of the other categories for the two samples and takes an average of the differences, the mean difference is still a significant 9 percent.

**SAMPLE BALANCING: NO SIGNIFICANT EFFECT**

Market researchers often use benchmarks (such as the U.S. Census) to “balance” or “weight” the data obtained from nonrepresentative methods. Like the “balance” control on a stereo system, this technique will put extra emphasis on data from groups underrepresented in the sample, and vice versa. Since the differences between the representative Knowledge Networks sample, EM-1 and EM-2 were so significant, we balanced the volunteer data to see how it would affect our comparisons. Can representative data really be “approximated” using nonrepresentative samples?

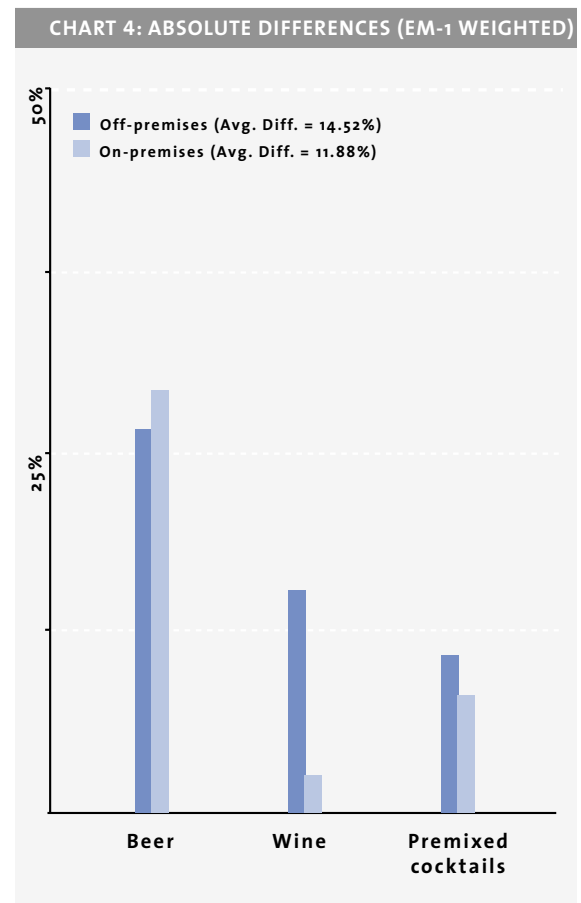
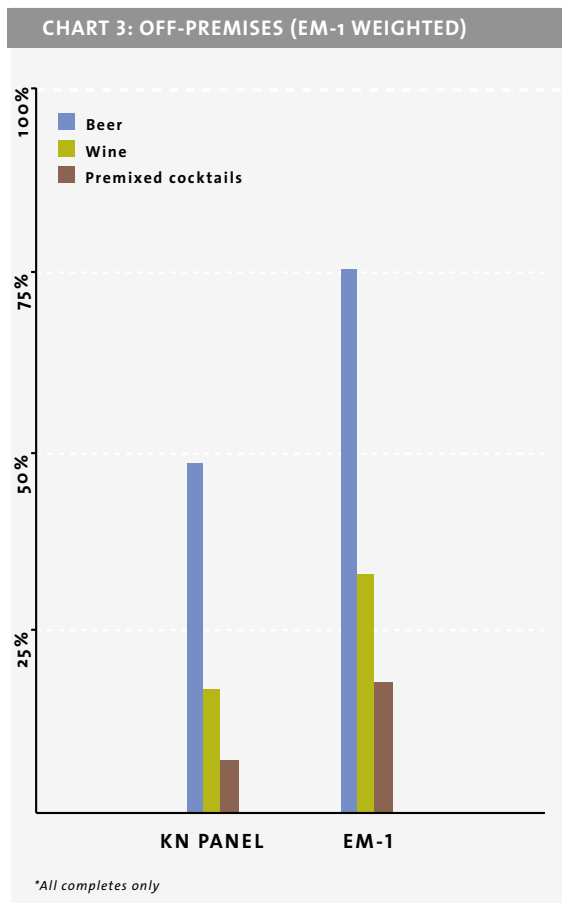
We weighted the EM-1 data so that the frequencies for race, age, education, and geographic region matched those of the U.S. Census. We then reproduced Chart 1 (as Chart 3) from above using the weighted results. The changes were negligible; significant differences were still evident between the Knowledge Networks panel sample and EM-1 data. The average differences between the two datasets for on-

and off-premises consumption (12 percent and 15 percent, respectively) were nearly identical to their unweighted counterparts (12 percent and 16 percent); and some estimates even moved further apart. *Chart 4* confirms that this was true of the absolute differences, as well.”

The results from this study show that the impact of the different methodologies is significant. If, for example, this manufacturer 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 Knowledge Networks panel data or EM-1.

Another way in which awareness and usage data are used is to further the research and market-intelligence-gathering processes. In this sense, the data from these studies are used to allocate research dollars for more detailed and specific studies. If a marketer wished to conduct research solely on hard-liquor drinkers, he or she would estimate wildly different field costs based on the EM-1 (69 percent) rather than the representative Knowledge Networks (37 percent) data.





It's not surprising, then, that this alcohol manufacturer found it prudent to use a representative sample. The study was important, and the potential for obtaining a biased sample on this topic was clear. If the goal had been to generate ideas, rather than to measure incidence levels, then a volunteer Web sample might well have proved more efficient.

Of primary importance, however, is simply understanding that there are critical business issues at stake in research methods—that different techniques can produce vastly different data for the same questions, and that those disparities can

easily affect business decisions. Knowing this simple fact is enough to increase your chances of success and help you make smarter choices about research and marketing.

*Daniel Slotwiner is Knowledge Networks' Vice President of Panel Management. He can be reached at [dslotwiner@knowledgenetworks.com](mailto:dslotwiner@knowledgenetworks.com).*