Sampling Bias Concerns in Web-Based Research

Leora E. Lawton, Ph.D.
Principal, TechSociety Research

The Internet opens up so many possibilities for communication of opinions, ideas, behavior and information that it was only a matter of time before market researchers and marketing analysts began to visualize benefits that they, too, could derive from Internet technology. Gathering data via the Internet is fast because it can be instantly coded into databases, because mailing time is cut dramatically, and because results are available immediately in real-time. The potential for speedy turnaround in research was tantalizing to the earliest visionaries of Internet business research.

The problem is that the Internet is subject to the laws of social behavior that govern physical space, in this case: sampling bias. In this essay, we discuss the development of web-based surveys and then how an understanding of the goals of the benefits and constraints of online methodologies applied to the goals of the intelligence campaign address and solve the problem of sample bias, while bringing additional benefits as well. The important thing to understand, above all, about the Internet is just like physical space, only different. By that we mean that the virtual world follows the same laws of social behavior as does physical space; except that the technology makes the application different.

Understanding best practices in Internet-based market research is no longer just an academic exercise: web-based studies became essential when customers began to use company websites to learn about products and services, whether as a prelude to picking up the phone, or as part of the e-commerce process.

Internet Surveys: a brief history

One of the earliest forms of surveying via the Internet employed by business researchers was presenting to web surfers who happened upon a site an invitation to click on a banner and take a survey. It became apparent that while this was a wonderful technology for fielding surveys, the banner invite was giving Internet surveys a bad name. This method creates a serious problem of selection bias: only those customers who happened to be online and who happened to be motivated to take the survey were included in the sample. In the early days of the Internet revolution (i.e., from 1995 when the web really began to take off), this bias was a major problem. Income was the greatest predictor of Internet access in those days (and still is, to a large extent), as well as skill in technology, which meant that the category of purchasers known as 'early adopters' were more likely to be voluntarily taking those surveys through sheer curiosity.

While that kind of sample might be reasonable for some Internet concerns, it certainly is not reasonable for most commercial enterprises. Obviously, product lines that appeal to a wider base than wealthy, highly educated curiosity-seekers would not be well served by this kind of data collection. But even those companies who would appear well poised can also be misled. For example, HotWired (www.hotwired.com) collected data from such voluntary web surveys

TechSociety Research
2342 Shattuck Avenue #362
Berkeley, CA 94704

Tel (510) 548-6174
Fax (510) 548-6175
lawton@techsociety.com
www.techsociety.com

linked to their site. For their early adopter tech-savvy readers in 1995-1996, such a sample was a reasonable population (Lawton, 1996). *Wired* (print) magazine has survived, but HotWired has had a difficult time identifying a growth market.

Other forms of data collection via the web are also of interest to Internet researchers. One can conduct ethnographic or 'participant observation' research, where the researcher participates in an activity while taking notes on patterns of behavior, interaction, deviance, norms and values (Lawton, 1995). On-line detailed open-ended interviews can replace face-to-face interviews. The added benefit that one can log conversation in real-time Internet environments provides an amazingly easy and accurate input into a database for text and content analysis (Lawton, 1995). Combining these two forms of interviewing facilitates an understanding of the growth in interest for online focus groups. While the software technology is still evolving, the potential for online focus groups to provide accurate, lightning-fast turnaround of data is high.

Email surveys are also possible, but coding data is often as difficult, or more so, than coding paper surveys because of varying formats of mail readers. In contrast, when web data is 'submitted' the form is parsed instantly into a database of choice, e.g., an ascii file, SPSS system file, an Oracle database, or into proprietary software analysis databases. With web-based systems providing instantaneous data entry, plus real-time display of results in password-protected environments, email surveys tend to be helpful only in small sample populations, usually within an employee sample, where the mail reader is known and standard.

Challenges for Web Surveys are Similar to Traditional Methods

One of the primary tasks of an Internet researcher is to identify potential sources of sample bias in web-based research design. Understanding sample bias is essential in traditional research as well. Indeed, technology is really the only difference between online and traditional methods of research. Sample design is essential in phone, mail, and face-to-face interviewing. Once the factor of access to online technology is considered, the means by which one deals with sample bias readily follows from traditional research methodologies. In other words, identify the universe of respondents, and then ascertain the best means for bringing them into the sample pool. There are limitations to all forms of data collection. Census bureaus worldwide seek 100% coverage of the respondent universe but do not attain it.

As with any methodology, high quality results derive from skilled personnel and carefully designed tools. With paper surveys you need to know you have good coders and coding techniques. In telephone surveys, you need a properly designed CATI survey with well-trained interviewers. In the Internet, you need someone who really understands both the programming of the software for electronic data collection and processing. For example, in web-based surveys the software developers need careful instructions from experienced researchers in how to save the data so that it will accessible to the analyst. [Before working with Informative's KnowledgeEngine, the author has personally had to clean data, respondent by respondent, cell by cell, due to poor data entry design in three different data collection tools.]

TechSociety Research
2342 Shattuck Avenue #362
Berkeley, CA 94704

Tel (510) 548-6174
Fax (510) 548-6175
lawton@techsociety.com
www.techsociety.com

Each form of data collection has its advantages and disadvantages, and its own particular form of sample bias. Consider mail surveys. The response rate is often as low as 5 to 10%, although a well-targeted respondent list can go considerably higher, especially when follow-up methods are employed. However, the cost of high response may be prohibitive for many research purposes. The response rate is a function of respondent interest in the survey, as well as accuracy of the list in terms of the names and addresses. In addition, for business-to-business surveying, the title and position of the respondent is key; short tenure in employment makes up-to-date lists difficult to maintain. For example, in a mail survey for an enterprise software product, only 40% of the list contacts were valid: about 10% of these responded.

Phone interviews offer other advantages and disadvantages. The method is quicker though more expensive than mail surveys. Random digit dialing (RDD), even when non-valid numbers are removed in advance, is still stymied by other sources of bias. Answering machines, unlisted numbers are obstacles to interviews. Cell phones and portable numbers have probably ended probability sampling. Changing work and lifestyles also affect the respondent pool. With most women in the labor force, even with young children, the kinds of people at home during the day is limited. The window of opportunity from 5 to 9 PM is in competition with telemarketers and the scourge of the researchers: 'suggers' (selling under the guise of research). In business-to-business researchers, the list may be dated, and secretaries act as gatekeepers to executives.

Knowledge and Experience Leads to Effective Web-Based Intelligence Campaigns

The problematic 'banner invite' survey discussed above can still be used effectively, but only if the researcher can manipulate either the flow of respondents or balance the sample post-data collection. Flow can be controlled by including a 'screener' as part of the survey. When the quota for each group is filled, leading to a representative sample, other potential respondents will be skipped out of the survey with warm thanks for their time. The use of cookies and email log-ins controls the possibility of repeat respondents. Thus, having a clear handle on the methodology and on research itself is key to high quality results.

The traditional methods for conducting surveys are challenged by sampling and response bias, just like the web surveys. Solutions for handling potential bias abound. The ultimate solution, however, is to recognize the fundamental macro changes occurring in economy and society. The Internet is as significant as the printing press for societal development and evolution. It follows that Internet technology is the next and most logical means to collect data in the 21st century. While traditional methods will always have their place, interactive methods of all kinds are likely to become the norm for market research as did phone and mail surveys in the 1970s and 1980s.

TechSociety Research
2342 Shattuck Avenue #362
Berkeley, CA 94704

Tel (510) 548-6174
Fax (510) 548-6175
lawton@techsociety.com
www.techsociety.com

References

Bowers, Diane. 1998. FAQs on Online Research. *Marketing Research.* Winter 1998/Spring 1999: 45-49.

Dillman, Don. 1978. *Mail and Telephone Surveys.* New York: John Wiley and Sons.

Lawton, Leora. 1996. HotWired. Why are people interested in HotWired content? Results from an on-line survey.

Lawton, Leora. 1995. Bellcore. Social Behavior in Internet Games

Mosley-Matchett, J.D. 1998. Leverage the Web's research capabilities. *Marketing News,* April 13, 1998: 6.

Weible, Rick and Wallace, John. 1999. Cyber Research: The impact of the Internet on data collection. *Marketing Research,* Fall, 19-24, 31.

TechSociety Research
2342 Shattuck Avenue #362
Berkeley, CA 94704

Tel (510) 548-6174
Fax (510) 548-6175
lawton@techsociety.com
www.techsociety.com