



## Using Internet Data Collection in Marketing Research

Yan Luo

School of Management

Tianjin University

#92 Weijin Road, Nankai District

Tianjin 300070, China

E-mail: [iamcarolloo@yahoo.com.cn](mailto:iamcarolloo@yahoo.com.cn)

### Abstract

With fast growth, internet has been used almost everywhere. For marketing research, internet's bi-directional, multimedia characteristics make marketing research more easily and efficient. In this paper, data collection in marketing research is discussed. It focuses on the application of internet in data collection. The strength and weakness are also mentioned.

**Keywords:** Internet, Online survey, Data collection, Marketing research

### 1. Introduction

#### *1.1 Marketing Research*

To be successful in the fiercer and fiercer competitive market, precise marketing researches are the basic and critical element. Formal studies of specific problems and opportunities are necessary for a marketer. Marketing research can help to realize these by market survey, product-preference test, sales forecast, and advertising evaluation, etc. Philip Kotler (2000) defined marketing research as "the systematic design, collection analysis, and reporting of data and findings relevant to a specific marketing situation facing the company."

#### *1.2. Data Collection*

Data collection is an elementary step in marketing research. Sampled data are collected through various means that provide a basis for analyzing the market behavior of a general population. Traditional methods for data collection include data recording and survey. Data recording refers to the process of collecting market-related data over time, such as sales data, advertising expenditure, customer information, etc. Usually, these data are collected for business purposes and are stored in the company's central database. A survey usually consists of three basic elements: target audiences (the group of selected people from the population), communication methods such as mail, fax, telephone, and person-to-person interviews, and survey content that contains a set of carefully-designed questions. Analyzing market data is crucial for understanding a consumer's buying habits and forecasting future market outcomes, and, accordingly, in assisting marketing decision-making.

### 2. Internet Data Collection

The growth of internet has revolutionized a lot of business and consumer behavior (Harter, 1999). One of the consequences is that it has been used as an alternative research tool for market research. Using of internet has made it less costly and more quickly and easily to get information than ever before. Internet also allows more detailed information to be collected than traditional surveys.

#### *2.1 Different Forms of Internet Surveys*

Internet survey can be done in different ways. In Comparing several forms of online surveying, MacElroy (1999) describes some of the common forms of online research being used for commercial applications, ranging from the most basic through highly sophisticated forms.

#### **E-mail**

The text-based e-mail survey is the most basic online research form. E-mail surveys are very like traditional mail surveys, the whole survey is sent to the audience, and can be seen by the respondent at once and respondents can change their answers at any time during the survey. The only difference is the way to transmit the survey. There is no need of special software and too much high technique, so it is considered as the least costly and simplest method. But at the same time, this simplicity leads to "boring" feeling when people complete the survey. Generally speaking, the ideal use of e-mail form is for internal corporate surveys because these audiences tend to follow instructions better than external

audiences (MacElroy 1999).

### **Bulletin boards**

Bulletin board researches are useful for a "modified Delphi method" for collecting responses over time (MacElroy 1999). Similar to a focus group, the researcher invites a group of respondents to a specific Web site, of a physically existed meeting room, where a discussion topic is posted. People can respond to the question(s) and express their opinions based on others' comments.

The cost of bulletin boards is also cheap, but somewhat higher than e-mail surveys because handling the comment transcripts and code responses is very time-consuming. The rich data it provides and relatively rapid reactions are its biggest advantages.

### **Web HTML**

The HTML form-based surveys allow the respondent to click buttons and boxes, to fill in text boxes, and eventually to submit the information all at once. The colorful design such as graphics, audio, video, and animation make the whole survey process more interesting and attractive. Nowadays, it is the most common form of on-line survey, accounting for almost 80 percent of all survey data being collected on-line (MacElroy 1999).

### **Web fixed-form interactive authoring tools**

Another new form of on-line research is survey authoring tools, which have been developed from previous generations of software used to conduct computer-assisted telephone interviews (CATI) or disk-by-mail (DBM) studies (MacElroy 1999). This technique allows the individual researcher to construct highly sophisticated studies for the on-line environment. However, these tools typically limit the range of options in which the survey can be displayed. Some software packages may limit the number of questions can be displayed on the Web page, while others may limit the design flexibility. The researcher has to choose between these two alternatives. Another disadvantage of this technique is that most of these tools usually cannot allow the researcher to control the whole research process (lack of self-hosting options). The cost for self-hosting is quite high (in the range of \$30,000 to \$60,000) if the researcher wants to host the survey they have created on their own internal servers (McElroy 1999).

### **Web customized interactive programming**

The most powerful and flexible, yet the most expensive online surveying option is custom programmed surveys. Custom programming not only provides all of the modern technical controls such as screening, skip-patterns, and error checking, but also offers many other functions that allow the researcher the highest level of flexibility.

The key benefit of this technique is its flexibility in layout. Unlike fixed-form tools, this type of online survey enables the researcher flexibly chooses many different design options such as question/response styles, backgrounds, and graphics. In addition, it provides much extended functionality such as running a live Web site for evaluation within the survey frames, and running procedures such as file downloads during the survey (McElroy 1999). However, compared to fixed-form tools, custom designed surveys take longer to program and can be more costly.

### **Downloadable surveys**

Another on-line research method is surveys that are downloaded from the Web and run on previously installed software provided by the researcher (MacElroy 1999). This technique shifts the computing tasks from the on-line server to the respondent's PC. Once the survey has downloaded on the respondent's PC, a data file is created which can then be opened the next time the Internet is accessed.

As an alternative, completely custom-programmed surveys can be used for single-use surveys. Of all the research options mentioned here, this technique allows the researcher to have the most control and flexibility. Because this type of survey requires respondents have certain level computer knowledge in order to install software and correctly handle the data upload process, this form of online research tend to be more costly and time-intensive than other forms. Typically, this technique is often used with panels or pre-recruited groups who regularly communicate with the survey organizer (MacElroy 1999).

### **Web Moderated Interviewing: Chat Interviews and Other Discussion Formats**

The last form of online research is the qualitative, real-time chat interviews. Although some people refer to this technique as on-line focus groups, MacElroy(1999) draws a few distinctions between these two methods. While people in traditional focus groups are highly influenced by the moderator's skill, in chat interviews, however, the logic and control mechanisms are supplied by a highly skilled human moderator. Just as the traditional moderator must control the overly enthusiastic participant, the on-line moderator must control the "tyranny of the fastest typist" (MacElroy 1999). The key benefit of chat interview is its flexibility. Respondents from different geographical regions can be brought together quickly and efficiently to participate discussion.

## 2.2 The Comparison of Internet Data Collection with Traditional Method

Comparing with other approaches, internet-based data collection methods are claimed: lower costs; faster turnaround; broader stimuli potential; and greater enjoyment (Forrest, 1999; Kehoe and Pitkow, 1996). In contrast, traditional marketing research is suffering from falling participation rate, rising costs, respondent fears concerning misuse of personal information and managerial issues resulting from the time taken to conduct postal surveys (Jarvis, 2002). But it also shows several disadvantages of internet survey such as sampling difficulties, the problems of response rates and quality.

### 2.2.1 Lower Costs

The cost of sending e-mail or setting up a Web questionnaire is much cheaper than other equivalents. In traditional research, the need for printing, stuffing, two-way postage, calling-expenses (especially, long-distance calls), data entry, handling and tracking, etc costs a lot of money for a survey. Using of internet survey can reduce or remove much of these burdens. Once a questionnaire is completed, the data are already in electronic form and can be downloaded immediately into a database for complex analysis (Lyer, 1996). In effect, once the last questionnaire for a survey is complete, researchers have all the data stored in a database at their fingertips instantaneously. If specific software doesn't need continually developing, the cost of internet-based survey is much lower than others. Weible and Wallace (1998) estimated that the variable cost of sending a questionnaire (via e-mail) or survey invitations (for a web survey) to 150 additional e-mail addresses was the equivalent of adding one extra contact to a mailing sample.

### 2.2.2 Faster Turnaround

Turnaround times of online survey are considered to be much faster. In the mixed mode e-mail invitation and HTML form study McDonald and Adam (2003) did, 40 per cent of responses were received in the first 24 hours, and over 85 per cent of responses were received within seven days. (Table 1)

### 2.2.3 Broader Range of Stimuli Potential

The fast development of computer technologies offer researchers richly tools to design and issue the surveys. Through the inclusion of color, graphics and sound, the boring questionnaire or survey will be interesting. Some of websites even use interactive windows during the survey. All of these stimulate the participation of surveys.

### 2.2.4 Sampling Difficulties

Although online survey is efficient at collecting a large data, several problems still exist making sampling difficult.

First of all, it may fail to develop a sample that is representative of the population as a whole. As previous research showed, regardless of the fact that the number of people who has internet access is increasing dramatically, the majority of internet population is still standing on younger and higher standard of education field than the public at large.

Secondly, the lack of a centralized database of e-mail addresses makes a major difficulty for e-mail survey. Dommeyer and Moriarty (2000) mentions that sampling of survey through e-mail is difficult because they frequently become outdated as internet users change their e-mail providers. In a study of Oppermann (1995), it was discovered that of the 500 e-mail addresses that were selected from the members directory of Association of American Geographers (AAG), 25 per cent were out of date.

### 2.2.5 Response Rate

Although the response rate of e-mail or/and HTML surveys is often claimed high, many researches showed that responses to internet surveys tend to vary according to the field of study. Table 2 shows the response rate of online and traditional data collection methods.

We can find from Table 2 that, with the exception of Parker's study (1992), response rates from surveys employing online methods (e-mail, HTML web, or the combination of these two) tend to be lower than those of traditional methods.

## 2.3 Data Mining

Data mining has a very close relation with data on-line collection. It's hard to separate them from each other. Internet data collection has created an increased demand for data-mining tools.

Data mining is defined as a class of database applications that look for hidden patterns in a group of data that can be used to predict future behavior. For example, data mining software can help retail companies find customers with common interests. The term is commonly misused to describe software that presents data in new ways. True data mining software doesn't just change the presentation, but actually discovers previously unknown relationships among the data. It is popular in the science and mathematical fields but also is utilized increasingly by marketers trying to distill useful consumer data from Web sites.

Data mining is rapidly becoming the mainstream in marketing research in recent years due to the need for consumer-centric marketing. The goal of customer-centric marketing is to seek to better understand the customers, thus, sophisticated data mining applications are required. Usually, companies are interested in collecting customers' data such as buying patterns, demographic, and psychological trends and other types of customer knowledge. Based on this detailed information, companies will be able to create appropriate targeted marketing plans. Then the marketer can correlate those data with information they have and find out who the most profitable customers are. However, sometimes it is difficult to collect data about people who have bogus information. In this case, the marketer needs to use data-cleansing tools to identify whether the data is valid and filter out junk data from the main database, so that the marketer will not make wrong decisions because of those errant data (Harter 1999).

#### 2.4 Barriers of Internet Surveys Development

Internet survey research also has some barriers on. Issues such as consumer privacy and non-representative of Internet samples are often- mentioned disadvantages. We will discuss these issues in detail in the following.

##### 2.4.1 Privacy issues

The growing application of the Internet in marketing research has enabled companies and marketers to obtain consumer information efficiently (Blattberg, Glazer, and Little 1994; Milne 2000). However, as companies are increasingly using sophisticated data-mining techniques to build extensive consumer databases, security on the Internet becomes a concern and consumer privacy is potentially violated during the process of information exchanges with marketers (Bloom, Milne, and Adler 1994; Milne 2000). As Milne indicates, information exchanges can occur in two different ways: "In some ways consumers volunteer the information, and in others the information it is gathered without consumers' knowledge and consent " (Milne 2000). As consumers become comfortable shopping online, filling out surveys, and giving their credit card information over the Internet, they have control over their personal information and are aware of what information is given out. An industry-sponsored poll indicates that "online shoppers do not mind their behaviors being watched if their shopping experience can be customized and they have an opportunity to opt out " (Greenberg 2000; Milne 2000). Unfortunately, by using cookies and tracking software, many organizations engage in data collection activities " that customers do not perceive as favorable" (Milne 2000). For example, companies such as DoubleClick can merge cookies to help online marketers to track consumers' online behavior such as click-and -viewing patterns. This information permits companies to sell advertising on their Website, which in turn, "contributes to consumers' concerns about profiling and unwanted e-mail solicitations" (Caudill and Murphy 2000; Petty 2000; Milne 2000). Another concern of consumer privacy is that consumer information on a database is potentially accessible to the entire Internet world (Milne 2000). As the result, whether consumers are willing to make or not to make a trade-off and permit marketers or organizations to collect their personal information become important issues for online marketing practices. Recently, the FTC unanimously supported a self-regulatory proposal developed by the Network Advertising Initiative, a coalition of third-party ad-servers such as DoubleClick, 24/7 Media and Engage to protect consumers' online privacy (AdAge 2000). In addition, technological solutions nowadays also enable consumers to protect their identity while providing information to marketers.

##### 2.4.2 Biased results

Non-representative of Internet samples is one of major disadvantages of online data collection. The expense and effort can be tremendous for correcting biased results. In *Research and the Internet: A Winning Combination*, Clarkson (1999) shows a few examples of how companies handle biased results. For example, instead of distributing surveys to its customers, Drugstore.com places the survey on its Website, which allows customers volunteer the information. The bottom line for this data collection method is that as long as the demographics of the Web are close to the demographics of the people the company tries to reach, the results are acceptable (Clarkson 2000). As Clarkson indicates, it is general belief that "the data collected would lead to good insights regarding the potential market, and the skewed demographics could be corrected by using statistical techniques to weight the actual responses to bring them in line with the demographics of the population at large" (Clarkson 2000). However, some companies such as Gallup challenges such a belief by conducting more rigorous research procedures combining traditional and Internet research methods in order to obtain scientifically accurate online data results. Gallup first conducts telephone interviews with all potential respondents. Gallup then mails letters, which include a personal identification number (PIN) and an Internet address, to the screened respondents. In this scenario, survey responses can be tied back to information already known about the individuals by comparing data with their original identification codes (Weissbach 1997). This way enables Gallup to maximize participation and representativeness as well as trace who has completed or has not completed the survey. After two weeks, e-mails are also used for reminding people to fill out their survey. In order to increase response rates, Gallup also provides \$5 cash reward for respondents who complete the survey. Clarkson also indicates that depending on the company and the respondents being surveyed, the Web-based survey is not always appropriate. The company must evaluate its problem and situation carefully before using Internet survey to gather the data.

Another reason for getting biased results in online data collection is due to limited usage of the Internet. Despite the rapid growth of personal computers, Internet usage penetration in every household is not as prevalent as television does. Besides, Internet users tend to be younger, more educated, and affluent. As a result, elders and people with less education and less income are more likely to be ignored by marketing researchers. Moreover, due to the nature of electronic questionnaires, Internet users can choose to discard and not to fill out the survey (Iyer 1996). In this case, the marketing researcher may need to use some incentives to attract people. For example, enticements could include entry in a contest, the chance to download some personal software, and participation in a game (Weissbach 1997).

### 3. Conclusion & Suggestion

The appearance and fast growth of internet benefit marketing research a lot. Comparing with traditional methods, internet data collection provides a cheaper and easier tool to collect data. Faster turnaround is offered, too. That allows business to record market-related data efficiently. However, there are a number of unfavorable natures. As for online-survey, uneven online population makes sampling difficult. That, furthermore, influences the quality of the survey results. Low response rate is another existing problem needs improving.

As some researches shown that, in current circumstance, internet based research should be used along with other methods as an additional supporting tool rather than as an alternative to traditional research approaches. However, for some organizations requiring specialist areas of research such as employee surveys and web site evaluation, an internet survey may be the most cost effective and appropriate way of carrying out the research.

### References

- Blattberg, R., Glazer R. & Little J. (1994). *The Marketing Information Revolution*. Boston: Harvard Business School Press.
- Bloom, P.N., Milne G.R. & Adler R. (1994). Avoiding Misuse of New Information Technologies: Legal and Societal Considerations, *Journal of Marketing*, 58, 98-110.
- Caudill, E. M. & Murphy P. E. (2000). Consumer Online Privacy: Legal and Ethical Issues, *Journal of Public Policy & Marketing*, 19, 7-19.
- Clarkson, B. (1999). Research and the Internet: a Winning Combination, *Quirk's Marketing Research Review*.
- Dommeyer, C. J. & Moriarty E. (2000). comparing two forms of an e-mail survey: embedded vs attached, *Journal of the Market Research Society*, 42, 39-50.
- Forrest, E. (1999). *International Marketing Research*, McGraw-Hill, Cincinnati, OH.
- Harter, B. (1999). Online Data Collection, *Wireless Review* 16 54-62.
- Jarvis, S. (2002), CMOR finds survey refusal rates still rising, *Marketing News*, 4 February, American Marketing Association, Chicago, IL, 4.
- Kehoe, C.M. & Pitkow, J.E. (1996). Surveying the territory: GVU's five WWW user surveys, *The World Wide Web Journal*, 1, 77-84.
- Lyer, R. (1996). The Internet: A New Opportunity for Marketing Research Firms, *Quirk's Marketing Research Review*.
- MacElroy B. (1999). Comparing Seven Forms of On-Line Surveying. *Quirk's Marketing Research Review*.
- McDonald, H. & Adam, S. (2003), A comparison of online and postal data collection methods in marketing research, *Marketing Intelligence & Planning*, 21, 85-95.
- Milne, G. R. (2000). Privacy and Ethical Issues in Database/Interactive Marketing and Public Policy: A Research Framework and Overview of the Special Issue, *Journal of Public Policy & Marketing*, 19, 1-5.
- Oppermann M. (1995). E-mail survey – potentials and pitfalls, *Marketing Research*, 7 28.
- Parker L. (1992), collecting data the e-mail way, *Training and Developing*, 7 135-139.
- Weible, R. & Wallace, J. (1998). Cyber research: the impact of the Internet on data collection, *Market Research*, 10 19-31.
- Weissbach, S. (1997). Using the Internet for quantitative survey research, *Quirk's Marketing Research Review*.
- Wilson, A. & Laskey, N. (2003). Internet based marketing research: a serious alternative to traditional research methods? *Marketing Intelligence & Planning*, 21 79-84.

Table 1. Cumulative Response for Online &amp; Postal Data Collection

Days from Invitation	Online Cumulative Response (%)	Postal Cumulative Response (%)
1	39.1	0.0
2	52.0	0.0
3	60.2	0.0
4	75.5	2.3
5	81.9	2.7
6	84.5	9.0
7	86.7	9.0
8	88.8	9.86
9	89.9	10.2
10	90.9	52.6
11	92.0	76.4
12	93.1	85.2
13	94.1	95.1
14	95.3	95.8
15	96.2	97.5
16	96.7	97.5
17	97.0	98.4
18	98.3	98.6
19	99.6	98.6
Cut-off day	100.0	100.0

**Note:** weekends and non-postal delivery days not included

**Sources:** Heath McDonald & Stewart Adam (2003)

Table 2. Response Rate of Online and Traditional Data Collection Methods

Author(s) and year	Online survey	Traditional survey	population
AFC (present study)	21w	46p	AFC members
Couper et al. (2001)	41w		Random sample of 1602 University of Michigan students
Adam and Deans (2000)	17w		Australian and NZ online business directors and managers
Bachmann et al. (2000)	19e	46p	Business school deans and chairs
Weible and Wallace (1998)	30e 34h	36p 31f	MIS faculty, mainly North America
Schaefer and Dillman (1998)	54e	58p	Washington State University faculty
Tse (1998)	7e	52p	Chinese University of Hong Kong teaching and admin staff
Noh (1998) <sup>b</sup>	31e		Public email directory <sup>b</sup>
Zelwetro (1998) <sup>b</sup>	38e	36p	na <sup>b</sup>
Besser (1997) <sup>c</sup>	20e	16p	Former members of the Rural Sociological society
Couper et al. (1997) <sup>c</sup>	43e	71P	Employees of US government statistics agencies
Smith (1997) <sup>c</sup>	8e 13e		Members of Web consultants association
Williams et al. (1997) <sup>c</sup>	27e	75p	Lowa State University students
Bachmann et al. (1996) <sup>c</sup>	53e	66p	Business School Deans in the USA
Comely (1996)	14e 47o	15p	Purchased list of UK Internet magazine subscribers
Hertz et al. (1996) <sup>a</sup>	69e	96t	Health care workers using CDC Wonder
Parks and Floyd (1996) <sup>a</sup>	33e		USENET newsgroups
Mehta and Sivadas (1995) <sup>c</sup>	40e	64p	Users of an electronic bulletin board
Opperman (1995) <sup>c</sup>	49e	26p 33p	American Association of Geographers
Kittleson (1995) <sup>c</sup>	28e	77p	International Directory for Health Educators
Tse (1995) <sup>c</sup>	6e	27p	Chinese University of Hong Kong teaching and admin staff
Schult and Totten(1994) <sup>c</sup>	19e		Marketing and MIS faculty at US universities
Parker (1992) <sup>c</sup>	68e	38p	AT&T employees
Walsh et al. (1992) <sup>c</sup>	76e		Subscribers to online user group
Liesler and Sproull (1986) <sup>b</sup>	67e	75p	Fortune 500 company department
Sproull (1986) <sup>c</sup>	73e	87t	Fortune 500 company department
Note: e = e-mail; h = HTML form; w = e-mail plus HTML form; o = postal plus HTML form; p = post; t = telephone; f = fax			
Sources: Cho and LaRose (1999); <sup>b</sup> Dommeyer and Moriarty (2000); <sup>c</sup> Schaefer and Dillman (1998)			