Internet marketing research: opportunities and problems

Olivier Furrer and
D. Sudharshan

The authors

Olivier Furrer is Lecturer in Marketing at the University of Birmingham, The Birmingham Business School, Birmingham, UK.
D. Sudharshan is Professor of Business Administration, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA.

Keywords

Internet, Market research, Surveys, Sampling

Abstract

The Internet is promised a brilliant future among the favorite tools of marketing researchers. Develops a typology of Internet marketing surveys showing the existence of eight different designs that can be used by marketers. However, researchers who plan to develop research using the Internet need to be aware of several problems related to this new tool. In particular we show that the nature of the Internet creates different sampling problems. To identify these problems, a seven-step procedure following the steps of the sampling process is proposed. Several practical problems are then discussed.

Electronic access

The research register for this journal is available at
http://www.mcbup.com/research_registers

The current issue and full text archive of this journal is available at
http://www.emerald-library.com/ft

Introduction

The Internet as a marketing research tool is relatively new and marketing researchers need to learn how to use it. In this paper we develop a framework to help researchers to identify potential opportunities and problems related to the use of the Internet for their marketing research.

We organize our discussion around three important themes facing any researcher:
(1) survey design;
(2) sampling; and
(3) field work.

We first develop a typology of Internet marketing survey designs based on the different combinations of the objectives of the survey, the selected sampling unit, and the data collection method to be used. The result of this combination leads us to discuss eight different designs. We then develop a seven-step methodology to identify the potential problems related to sampling and discuss specific problems with the fact the Internet is ever changing. Fieldwork problems are finally presented with a focus on the format of the questionnaire, the use of search engines to obtain a frame and problems related to the sampling units.

A typology of Internet marketing survey designs

The first question to be asked by a marketer planning to use the Internet for its research is: what do I want to do? What are the objectives of the research? Many different types of research may be conducted through the Internet or with the help of the Internet. A typology of Internet marketing research designs may be constructed around three dimensions:
(1) the objectives of the survey;
(2) the sampling units; and
(3) the data collection method.

The next paragraphs discuss these three dimensions and their content. We then present eight different research designs based on the combination of these three dimensions.

Objectives of the survey
Marketing researchers may use the Internet to meet three different types of research objectives:

(1) they can study how the Internet is used as a marketing tool;
(2) they can use the Internet as an alternative medium for traditional questionnaire surveys; or
(3) they can use the Internet to study Internet consumer behaviors.

The use of the Internet pertains to studies of the way firms use the Internet as a marketing tool. Pitts *et al.* (1996) noticed that Web sites have been used to achieve the objectives of a number of marketing strategies, for examples to:
• gain access to previously unknown or inaccessible buying influences;
• project a favorable corporate image;
• provide product information;
• foster and encourage consumer involvement with the product range;
• generate qualified leads for sales people;
• handle customer complaints, queries and suggestions; and
• serve as an electronic couponing device.

Internet marketing research may be used to study the effectiveness of these uses of the Internet.

The Internet could also be used to perform traditional questionnaire surveys. Except for technical particularities, there are no real differences between Internet questionnaires and traditional mail or phone questionnaires. The advantage of the Internet is that it can offer a richer support for the questionnaire and it also allows data to be recorded automatically (which is a very important advantage). Internet questionnaire surveys could be used to investigate any traditional marketing research question ranging from firms' marketing strategies and policies to consumer behaviors.

The third type of marketing objectives that could be reached by an Internet marketing survey is the study of Internet consumer behaviors, for example understanding how consumers shop in cyberspace and why they do what they do. Internet consumption includes:
• gathering information passively via exposure to advertising;
• shopping, which includes both browsing and deliberate information search; and
• the selection and buying of specific goods, services, and information (Goldsmith and Bridges, 2001).

These kinds of surveys are concerned about how consumers use the Internet, how they navigate, how they buy on the Internet, etc. Customer satisfaction related to Web sites could also be studied in the same way.

Sampling units
Three different sampling units could be used in an Internet marketing survey:
(1) Web pages;
(2) Web sites; and
(3) Internet users.

A sample of Web pages could be drawn to survey the design, the content, the advertising space on these pages, or the time these pages remain unchanged, but also the number of visitors that a page has (number of hits) or the time visitors spend on a particular page. A sample of Web sites could be drawn to survey firms' Internet communication and marketing strategies, the structure of a site, etc. Finally, a sample of Internet users could be drawn to study their general Internet behavior, or their behavior toward a particular Web site or page. A sample of Internet users could also be drawn to ask them to fill in an Internet questionnaire or to send them a questionnaire by e-mail. These questionnaires may be related to their Internet usage but not necessarily.

Data collection methods
The third dimension of our typology of Internet marketing survey designs is related to the data collection methods. Three different methods may be used:
(1) direct observation;
(2) a questionnaire; or
(3) an experiment.

The direct observation of the content of a Web site or page could either be done by the researchers or a sample of users. This method may be used to collect objective data such as the number of advertisements on a page or the existence of frames or the number of links inside a particular page; or to collect subjective data such as the satisfaction of users or their reaction to the content or design of a Web page.

Four kinds of format could be used for Internet questionnaire surveys:
(1) Web sites questionnaires;
(2) e-mail questionnaires;
(3) text formatted forms sent by e-mail as an attached file; or

(4) downloadable text formatted form posted at a designed ftp which can be returned by mail or fax.

Except for technical issues, there are no real differences between these forms of questionnaire and the traditional questionnaires by mail or by phone.

An experiment can be conducted by the creation of an experimental Web site and by studying users' behaviors towards some manipulated changes in the site. The content, the design, and/or the navigation flow may be manipulated to identify their effect on consumer behaviors.

Eight different designs

Eight different research designs may result from the combination of these three dimensions (Figure 1).

(1) The first and easiest design is to draw a sample of Web pages after the researchers have observed the content of the pages selected. This design may be used to study the use of the Internet as a marketing tool, especially the content of a page in terms of image or advertisement.

(2) A second design is to draw a sample of Web pages by a two-stage sampling procedure. The first stage consists of drawing a sample of Web sites and the second stage of drawing a sample of pages among the selected sites. This design allows the researchers to control the representativity of the sample in terms of types of site. With this design, non-commercial sites could, for example, be easily excluded from the sample. Observation may also be used to study the usage of the Internet as a marketing tool.

(3) A sample of Web sites may also be drawn to study the use of the Internet as a marketing tool by observation but at a level higher than the Web page. Compared with the previous design, this allows the researcher to study the structure of a site. Mottner *et al.* (2001) use a similar design to develop a typology of Internet retailers.

(4) In the previous two designs, the researchers were the observers, but to study some subjective characteristics of Web sites the observation could better be done by a sample of customers. In this case two samples need to be drawn: first, a sample of customers and then a second

Figure 1 Typology of Internet marketing research
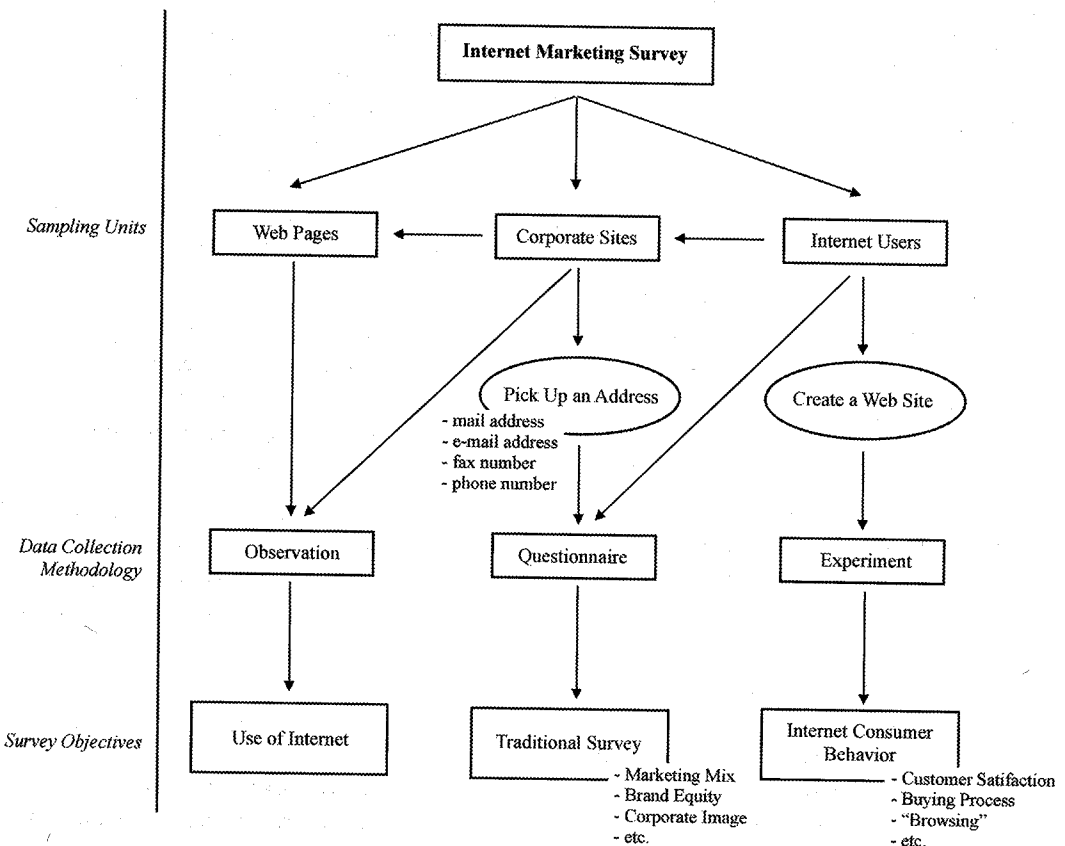
sample of Web sites or pages. This design may also be used to study the use of the Internet as a marketing tool. Griffith and Krampf (1998), for example, have used such a design to examine the impact of the retail marketing mix on consumer attitudes.

(5) A sample of Web sites could also be drawn to obtain the address (mail, e-mail, phone or fax) of firms and/or the name of managers. Then a questionnaire could be sent to these firms or managers. This design is particularly interesting for two main reasons: first, before sending a questionnaire, some characteristics of the firms to be sampled could be checked to be sure that these firms belong effectively to the population to be studied; second, Web sites can offer to the researchers information is not listed anywhere else, for example, an e-mail address. Fontenot and Vlosky (1998) used this design for their exploratory study of Internet buyer-seller relationships.

(6) Instead of drawing a sample of sites, a sample of Internet users could be drawn to send them a questionnaire by e-mail or to send them to an Internet address where a questionnaire is located. A similar design has been used by Comley (1996) to investigate the use of the Internet as a data collection method. But this design could also be used in various different questionnaire surveys related or not to the Internet.

(7) A sample of Web sites and a sample of Internet users could also be drawn. The Internet users are asked to visit selected Web sites and to fill in a questionnaire concerning their experience with the selected Web sites. Unlike the design where only the users are sampled, this design allows the researchers to control which sites are visited by the Internet users.

(8) An experimental design could also be used for an Internet marketing survey. An experimental Web site may be designed and a sample of Internet users could be drawn and asked to visit the site. Their behaviors could be studied after modifications of some characteristics of the site, such as its content, visual design, or its navigational design. Such a design has been used by Mosley-Matchett (1998) to study the effects of presentation latency on proficient and non-proficient users of Internet-based marketing presentations and by Luna *et al.* (2001) for the study of the use of the Web as a transnational communication tool.

Survey sampling problems

After having reviewed the different survey designs possible for an Internet marketing research survey and before presenting the practical problems of such surveys, we have to discuss the problems related to sampling. These problems are among the most difficult in the case of Internet surveys and actually some of these are not yet completely resolved.

The evolving nature of the Internet is the source of most of the problems related to selection of a probabilistic sample. To help marketing researchers to recognize these problems, we presented them in Table I following the seven steps of a classical sampling process (Kish, 1965). We then discuss in more detail the problems that are most important for Internet marketing research.

Moving universe

Among the more delicate sampling problems related with Internet survey is the fact that the Internet is a moving universe. The universe is the total group to be studied; it is the grand total of what is being measured. It must be defined so clearly that one will have no difficulty in deciding which unit belongs to it (Kish, 1965). But the Internet is moving, growing dramatically; many new sites are born every day and some others disappear. Consequently, the sampling universe changes every day and it is impossible to know the probability of selection assigned to each possible sample. If at a point of time this probability could be known, it could not be for a period of time. This leads to some problems of generalization. The results obtained from a survey at one point of time could only be generalized for this point of time. And problems are equally important for surveys that last for a period of time (month, year). The only solution to these problems of a moving universe is to substitute static indicators for dynamic ones and by surveying in several points of time.

Frame in a moving universe

A second important problem related with the Internet is the choice of a sampling frame. A

Table I Sampling problems

Steps	Description/procedure	Internet complications	Possible solution	General principles
1. Define the universe (population)	The universe is the total group to be studied. It is the grand total of what is being measured	Internet is an open (births and death) and growing universe	Multi-period sampling	The universe must be defined so clearly that one will have no difficulty in deciding which unit belongs where. It is important to note that the universe can only be defined operationally in terms of a real frame or a combination of frames and in the manner of using these frames
2. Define the sampling units	The sampling units are the units for which information is sought; they are the individuals, the elementary units comprising the population about which inferences are to be drawn	In Internet there is a choice of unit: Web site, Web page, etc.	Depends on the purpose of the survey	These units must cover the whole of the population and they must not overlap, in the sense that every element in the population belongs to one and only one unit
3. Identify the sampling frame	The sampling frame is literally a list from which the sample is selected	The different Internet research engines can be used as frame. But some problems of duplication must be resolved. Problem related with the fact that Internet is an open (births and death) and growing universe.	Multi-frame sampling Control of the duplication	No frame is perfect. It is impossible to work from a perfect frame, but good research planning requires knowledge of the shortcomings of the frame so as make adjustments in the sampling design
4. Select a sampling procedure	Selection of a sample procedure is inextricably intertwined with the identification of the sampling frame in that choice of sampling method depends largely on what the researcher can develop for a sampling frame	The choice of a procedure can be limited by the availability of supplementary information	Depends on the purpose of the survey and on the supplementary information available	Sampling procedures can be divided into the two broad categories of probability and non-probability samples. Probability samples are the only samples for which the sampling errors can be calculated, and for which the biases of selection, non-response, and estimation are virtually eliminated or contained within known limits
5. Determine the sample size	The sample size will be determined as a function with the degree of precision desired	Problem to evaluate the variance of the population	Increase the size of the sample. Control a posteriori the coefficient of variation of the estimates	The precision of the survey result is a function of the sample size Precision is related to the square of the number the sample and to the variance of the population
6. Select the sample elements	Randomization	Problem to control the randomization	Avoid Web automatic randomization	Each selected sample element must have a known probability of selection
7. Collect the data from the designed element	Different methods can be used to collect the information	Problem of non-response related with death	Depends on the purpose of the survey	Some problems of data collection must be controlled: getting the right respondent, the cooperation of the respondent, etc.

sampling frame is the list from which the sample will be selected. But resulting from the continuous births and deaths of Web sites, there is a difficulty in finding a sampling frame that is representative of the universe. If at a point of time, a universe could be defined, that is not the case for the frame. No frame is perfect, but with the Internet, the problem is particularly acute. Some search engines could be used to obtain lists of Web sites, but the results of the survey could only be applied to these lists. One solution to improve the validity of the results is to evaluate the "representativity" of the frame. This can be

done by the comparison of the sampled frame with other frames. The comparison with frames from other search engines creates information about the "spatial representativity", and the comparison with frames from the same engines, but at several points of time, provides information about the "temporal representativity". The bias related with the use of an incomplete frame can be evaluated by surveying a sub-sample of Web sites that are in the new frames but not in the sampled one and by testing the homogeneity of the results. The same tests could be done with a sub-sample of Web sites that are in the sampled frame but not in the new ones.

Non-referenced sites (non-coverage error)

Another important problem is the problem of non-coverage error in Internet research. The problem of partial frame could be stressed by the fact that many Web pages are not referenced by any search engines or other Web listings. This fact creates another limitation to the generalization of the survey results. In phone surveys, this problem of "non-coverage" could be solved by a random dialing of phone numbers. Could such a thing be possible with the Internet? That seems more difficult. A URL is different from a phone number in three ways:

(1) a URL is a combination of signs (letters, numbers, and special characters) not only numbers;

(2) the probability of occlusion of these different signs are not equal, contrary to the probability of occlusion of the different numbers in a phone number; and

(3) a URL has a variable length.

Respondent

In an Internet questionnaire survey, it is also difficult to control who is the respondent. If the questionnaire is located on a Web site, it is impossible to know who is visiting the site and answering the questionnaire. In some cases, it is possible to pick up the e-mail address of the respondent, but this address does not say anything about the respondent belonging to the sample. If the questionnaire is sent by e-mail, a control of the respondent could be better done, but the difficulty is to obtain the e-mail address. Unlike most phone numbers, e-mail addresses are only partially listed. And the available lists of e-mail addresses are far from being representative. When the questionnaire

has to be sent to a firm's manager, it is possible to find his/her e-mail address by exploring the firm's Web site. But often only the e-mail address of the Web master is available. In this case, the response will depend on the Web master's willingness to forward the questionnaire to the right respondent.

An international multilingual, multicultural environment

The Internet is a global worldwide network. It is not always possible to identify the nationality, the culture, or the language of a site only by its URL address: a .com address could be located anywhere in the world. This difficulty could create some problems when the goals of a survey are geographically or linguistically limited or when a geographical or a linguistic segmentation is needed. The language problem could equally reduce the ability of the respondents to answer the questions. In these cases, a frame obtained by a search engine, for example, needs to be cleaned of all the addresses that do not belong to the population studied.

Field work problems

Once the survey design has been defined, and once a sample has been drawn, some practical problems subsist. Among these problems, the most important are:

- the format of the questionnaire;
- the use of research engines to obtain a frame; and
- problems related to the sampling units.

Formats for the questionnaire

Three kinds of formats could be used for these questionnaires:

(1) Web site questionnaires;
(2) e-mail questionnaires; and
(3) text form sent by e-mail as an attached file.

Each one of these formats has some problems.

Quality of the frame

The most important problem of sampling on the Internet is to obtain a "representative" frame of the Internet universe from which we can draw a probabilistic sample. One solution to this problem is to use search engines as a sampling frame. But search engines are not perfect frames and their quality has to be evaluated. Their quality as a sampling frame depends on three criteria:

(1) the size of their index;
(2) the freshness of the index; and
(3) the ranking procedure and relevancy.

The "representativity" – i.e. the quality – of the frame is the most important problem for the validity of a survey. Some tests have to be done to estimate this quality. These tests consist of the control of the proportion of different defaults in the frame. A non-exhaustive list of the defaults could be:

- the proportion of dead sites in the frame – i.e. non-existing addresses;
- the proportion of shut down sites – i.e. non-responding addresses;
- the proportion of sites having changed their address but maintaining a link between the old and the new addresses;
- the proportion of sites (pages) appearing more than once in the frame – duplication.

The evaluation of these proportions has to be completed by the discussion of the importance of the bias that they create.

Problems related with the sampling units

The Internet is a network composed of sites, pages, links between different pages inside the same site, and links between different sites; also the sampling units could be a page, a site, or a group of sites. The frames generally available – search engines – give the address of one or more pages of sites. If the sampling unit is the site, a problem of redundancy appears, and if the sampling unit is the page, the problem is that not every page of a particular site is in the frame. The sampling unit could also be a group of sites belonging to the same firm – i.e. international firm – but having different addresses and related by several links. When the sampling unit is the site, before drawing a sample, it is important to clean the frame of all duplications. When the sampling unit is the page, a two-stage sampling is better. The first stage is to draw a sample of sites and the second is to draw a sample of pages among the sites sampled.

Conclusion

The Internet is a formidable tool for marketing researchers, which offers many opportunities; however it also has several difficulties. In this paper, we attempted to map the difficulties which lie on the road of the Internet marketing researcher. We first developed a typology of Internet marketing

survey design, which proposes eight different designs based on the objectives of the researchers, the sampling unit, and the data collection method. The second part of the paper identified, following the seven steps of the sampling process, the problems related to sampling the Internet. Finally, we presented some practical problems in conducting research on the Internet.

References

Comley, P. (1996), *The Use of the Internet as a Data Collection Method*, SGA Market Research, http://sga.co.uk/esomar.html

Deming, W.E. (1950), *Some Theory of Sampling*, Wiley, New York, NY.

Fontenot, R.J. and Vlosky, R.P. (1998), "Exploratory study of Internet buyer-seller relationships", paper presented at the AMA Winter Educators' Conference.

Goldsmith, R. and Bridges, E. (2001), "Attitudes toward e-commerce and buying online", in Krishnan, R. and Viswanathan, M. (Eds), *Marketing Theory and Applications: 2001 AMA Winter Educators' Conference*, Vol. 12, American Marketing Association, Chicago, IL, p. 58.

Griffith, D.A. and Krampf, R.F. (1998), "An examination of retail Web-sites: the impact of the retail marketing mix on consumer attitude", paper presented at the AMA Winter Educators' Conference.

Kish, L. (1965), *Survey Sampling*, Wiley, New York, NY.

Luna, D., Peracchio, L.A. and de Juan Vigaray, M.D. (2001), "The Web as a trans-national communication tool: flow, language, and Web site effectiveness", in Krishnan, R. and Viswanathan, M. (Eds), *Marketing Theory and Applications: 2001 AMA Winter Educators' Conference*, Vol. 12, American Marketing Association, Chicago, IL, pp. 131-2.

Mosley-Matchett, J.D. (1998), "The effect of presentation latency on proficient and nonproficient users of Internet-based marketing presentations", paper presented at the AMA Winter Educators' Conference.

Mottner, S., Thelen, S. and Karande, K. (2001), "A typology of Internet retailers: an exploratory study", in Krishnan, R. and Viswanathan, M. (Eds), *Marketing Theory and Applications: 2001 AMA Winter Educators' Conference*, Vol. 12, American Marketing Association, Chicago, pp. 56-7.

Pitts, L., Berthon, P. and Watson, R.T. (1996), "From surfer to buyer on the WWW: what marketing managers might want to know", *Journal of General Management*, Vol. 22 No. 1, Autumn, pp. 1-13.

Further reading

Blankenship, A.B. and Bree, G.E. (1992), *State of the Art Marketing Research*, American Marketing Association and NTC Business Book, Lincolnwood, IL.

Cochran, W.G. (1963), *Sampling Techniques*, 2nd ed., Wiley, New York, NY.

Quelch, J.A. and Klein, L.R. (1996), "The Internet and international marketing", *Sloan Management Review*, Vol. 37 No. 3, Spring, pp. 60-75.