

## Conducting Online Surveys

MARTINE VAN SELM<sup>1,\*</sup> and NICHOLAS W. JANKOWSKI<sup>2</sup>

<sup>1</sup>*Department of Social Science Research Methodology, Faculty of Social Sciences, Radboud University Nijmegen, 6500 HE Nijmegen, The Netherlands;* <sup>2</sup>*Department of Communication, Radboud University Nijmegen, The Netherlands*

**Abstract.** The World Wide Web (WWW) is increasingly being used as a tool and platform for survey research. Two types of electronic or online surveys available for data collection are the email and Web based survey, and they constitute the focus of this paper. We address a multitude of issues researchers should consider before and during the use of this method of data collection: advantages and liabilities with this form of survey research, sampling problems, questionnaire design considerations, suggestions in approaching potential respondents, response rates and aspects of data processing. Where relevant, the methodological issues involved are illustrated with examples from our own research practice. This methods review shows that most challenges are resolved when taking into account the principles that guide the conduct of conventional surveys.

**Key words:** online survey, Web based survey, email survey, research methods review, sampling procedures, response rate

### 1. Introduction

The Internet is increasingly used as a tool for and object of social scientific study. This may confront social scientists with the question whether or not new social science research methods need to be developed in order to enter the 'Internet arena'. With respect to this issue at least three positions can be distinguished, ranging from the idea that Internet research is 'business as usual' to 'the urgent need for entirely new methods'. We employ a third, middle-position in which, on the one hand, the usefulness of old ways of conduct are emphasized, and on the other hand, the development of principles and tools in order to deal with new challenges posed to researchers conducting Internet research. This research methods review is about one particular method increasingly employed in Internet research: online surveys. We discuss the advantages and liabilities with this form of survey research.

---

\*Author for correspondence: Martine van Selm, Department of Social Science Research Methodology, Faculty of Social Sciences, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands. Tel.: ++31-243611709; Fax: ++31-243612351; E-mail: M.vanSelm@maw.ru.nl

Frequent Internet surfers may encounter online surveys almost daily and be invited to visit specific sites where the questionnaires are located. A recently conducted search for Web based surveys<sup>1</sup> with a popular search engine resulted in a large amount of hits. The search produced many pages with information on software for the construction of 'do it yourself' Web questionnaires. In addition, a rough categorization of the first 100 pages suggests that online surveys are used for a wide range of topics, ranging from studies conducted among (medical) professionals, consumers, students, Internet users, job seekers, employees and more, and that (academic) discussion is prevalent about the 'ins and outs' of online surveys.

Schaefer and Dillman (1998) emphasize that the proportion of households with access to the Internet is too small to conduct general public surveys by email or via the WWW. Although the number of individuals and households with Internet access is increasing, the present penetration of the technology does not permit its use for such surveys (see, Dillman et al., 1998a; Van Dijk, 2000; Crawford, et al., 2001; Miller et al., 2002). Nevertheless, for some groups, such as company employees, members of professional associations, or college students, email access has reached nearly all and these groups can relatively easily be surveyed by email (Schaefer and Dillman, 1998: 378).

In this paper we address a multitude of issues researchers have to consider before and during the use of this method of data collection. First, we consider sampling issues related to these types of surveys. Here, the specific characteristics of sampling from an electronic sampling frame are presented. Further, attention is paid to the construction of an instrument for an online survey. Various ways of launching a questionnaire using email or the WWW are discussed as well as issues related to response rate, speed of return, and the quality of responses. The remainder of the paper is devoted to considerations of data processing and data cleaning, and to discussion of supportive software useful for performing tasks associated with survey research. Where relevant, the methodological issues involved are illustrated with examples from our own research practice.

## **2. Reasons for Conducting Online Surveys**

When do researchers conduct online surveys and what are their experiences with this research method? This question is frequently addressed in journals and academic forums.<sup>2</sup> One of the reasons for conducting online surveys has to do with the object of study. Online surveys are often employed in studies of Internet use in order to reach a population with Internet experience. An example of the latter is the well-known Web survey of Georgia Tech's Graphics Visualization and Usability Center (GVU), the purpose of which is to monitor changes in Internet user demographics and attitudes

towards technology and online commercial activity (Pitkow and Recker, 1995; Kehoe and Pitkow, 1996). The Nielson/NetRatings is another example of this kind of use. Inasmuch as the proportion of Internet users among most general populations is relatively small, securing a random sample by consulting postal addresses or telephone numbers would generate a large number of respondents who do not have access to the Internet (Kay and Johnson, 1999).

In addition to the object of study, particular characteristics of the population may also be a reason for conducting an online survey. For the recruitment of potential respondents with special interests, consultation of specific Internet environments can be valuable such as Usenet discussion forums, newsgroups, bulletin boards, chat rooms, and MUDs.<sup>3</sup> Various Internet environments serve as a meeting place for persons with similar hobbies, experiences, interests, and lifestyles. Also, support groups are frequently established in which personal experiences are shared (Mickelson, 1997). These self-organized groups can be valuable for researchers interested in designing a study of persons with particular characteristics. These selected groups enable, in other words, the conduct of "expert interrogations" (Swoboda et al., 1997: 243).

A third reason for conducting an online survey is to facilitate recruitment of respondents with deviant or covert behaviors. The anonymity possible on the Internet is believed to help in gaining access to respondents normally difficult to reach, and it may facilitate the sharing of their experiences and opinions. Therefore, online surveys are considered useful when the issues being researched are particularly sensitive (see Coomber, 1997).

A fourth point has to do with the attractiveness of computers to particular (age) groups. Using the Internet for survey purposes among young people may lead to higher response rates as compared to paper-and-pencil surveys. Beebe et al. (1997), in a discussion of design and development issues of computerized school surveys, suggest that the computer may have special relevance for younger groups. Use of the Internet in, for instance, school-based surveys is expected to improve response rate from such respondents.

There are other reasons for employing the Internet as a tool in survey research, such as economic advantage (Internet surveys are inexpensive as related to conventional paper-and-pencil surveys) and efficiency (data collection via the Internet is fast), reach, and simply gaining experience in using the Internet for survey purposes. With respect to *cost* reduction, Mann and Stewart (2000) warn that although computer-mediated communication (CMC) can cut production costs, the start-up expenses involved in Web based surveys, particularly expenses incurred to secure the necessary expertise for designing instruments, can be quite substantial.

A similar point is made regarding the *time benefits* offered by CMC use for survey purposes. Although Web based surveys can accelerate the rate of response, researchers often end up spending considerable time solving technical problems before and during implementation of an online survey. Time-consuming activities related to implementation of email surveys include: searching for email addresses, checking and replacing invalid email addresses, and explaining the form and procedures to less technically savvy respondents.

*Reach* refers to the ease by which potential respondents can be approached. Online surveys are particularly attractive in this respect when the population under study is distributed across a large geographic region. Mann and Stewart (2000), however, remind us about the risk of losing sight of who is responding to the questionnaire. The persons to whom an email survey is forwarded may not have the characteristics expected of respondents because of geographical (and cultural) distance. Also, access to the Internet in some workplaces and countries is limited or restricted. And, there is a strong bias against (survey) researchers in some computer-oriented subcultures. In the words of Mann and Stewart (2000: 73), "A stranger wanting to do academic research is sometimes seen as an unwelcome arbitrary intrusion."

This observation should temper the enthusiasm of researchers who feel that reaching Internet users is merely a matter of applying the technology. With respect to possible invasion of privacy, it seems that the identity of respondents in email surveys is less protected than in Web based surveys. When a respondent uses the 'reply' function of an email program to return the completed survey, the message carries at least the email address of the sender in the message header. Respondents can take measures in order to guarantee anonymity by sending messages through, for instance, a low profile remailer service. A remailer strips messages of their original headers and replaces them with other information. In this way the sender of a message becomes more difficult to trace. This increase in anonymity on behalf of respondents means a loss of control over the quality of data, however, as the remailer camouflages multiple responses from a single person. Web based surveys, in contrast, do allow for anonymity in as much as respondents are free to withhold their names. However, multiple responses are also possible in this form of online survey from a single person. Later we elaborate on possible interventions from a researcher concerned with controlling for multiple responses.

Other advantages to online surveys mentioned in the literature (Metha and Sivadas, 1995; Smith, 1997; Medlin et al., 1999; Brennan et al., 1999) include:

- absence of interviewer bias;
- removal of the need for data entry in as much as respondents directly enter data into an electronic file;
- convenience for respondents.

In summary, Sills and Song (2002) state that for particular populations that are “connected and technologically savvy”, the cost, ease, speed of delivery and response, ease of data cleaning and analysis weigh in favor of the Internet as a delivery method for survey research (Sills and Song, 2002: 28).

### 3. Sampling

The reasons for conducting online surveys mentioned above suggest that such surveys are best suited for studies among non-probability samples. This brings us to the issue of sampling bias, which is generally seen as the main objection to employing online surveys. Achieving a random sample of Internet users is problematic, if not impossible. Kay and Johnson (1999: 325) suggest that “By its nature, the Internet poses a unique set of problems in guaranteeing a random sample of respondents.” They consider the main problem with Internet surveys to be the absence of a central registration of users on the Web. Unlike telephone numbers and home addresses, constructing a random sample from email addresses seems impossible because of the variations in address construction.

Online surveys, one might argue, should be reserved for studies among non-probability samples because of these problems. Even though these surveys are not representative for the total population of Internet users, non-probability samples can be valuable as they may be representative for a subgroup of the total population. As Babbie (1990) recommends, non-probability sampling should meet the condition of sufficient response rate; otherwise the sample becomes no more than a self-selected group of respondents. Later we elaborate on this point and on the specific difficulties in determining response rate posed by online surveys.

Researchers who use online surveys generally invest considerable effort into ensuring an acceptable level of randomness and representativeness of non-probability samples. Kay and Johnson (1999) review a number of online studies on this point, and the authors suggest a variety of procedures that can be employed for achieving the desired level:

- random selection of email addresses from newsgroups;
- use of stratified random samples of proprietary bulletin board users;
- employment of a sampling frame from lists of users who have free access to the Internet;
- use of a stratified sample of individuals whose email-addresses are obtained from Usenet newsgroups.

As noted above, the Internet allows for different sampling frames in order to create (non-probability) samples. In newsgroups, for example, the forums are generally organized around specific topics, and messages posted

usually contain the email address of the sender. In a study by Swoboda et al. (1997), 200 newsgroups were selected from the groups available via the news server of the Technical University of Munich. In another study, concerned with developers of Web sites (Sheehan and McMillan, 1999), a sampling frame was constructed from the Yahoo directory on health-related sites and then email addresses were gathered from those sites. Extrapolating from this last example, the various directories of all search engines of the WWW could be considered a rough sampling frame. Sheehan and McMillan (1999) suggest a number of factors that led to the reduction of the size of their sample, constructed from the Yahoo directory on health:

- duplication of an URL or email addresses;
- malfunctioning URLs (statements, such as 'file not found' appear on the user's screen);
- absence of an email address on a particular Web site;
- 'return to sender' or 'bouncing' of email messages that are no longer current.

Medlin et al. (1999) distinguish three categories of samples: recruited samples, unrestricted and screened samples. In *recruited* samples, obtained by consulting and selecting from an existing sample frame, intended respondents can be given a password (by email) in order to control entry to a WWW survey. *Unrestricted samples* are generated by communicating the availability of a questionnaire widely and allowing anyone to visit the Web site and to complete the questionnaire. *Screened samples* can be constructed by collecting relevant screening data in the survey responses so that only responses from the required sample are analyzed and by using branching in the survey to filter out potential respondents.

Before a reliable estimate can be made about the distribution of a characteristic in a population based on a sample of a specified population, a number of sources of error must be overcome. Dillman et al. (1998a) recommend, first of all, employing careful sampling procedures to help avoid coverage and sampling errors. *Coverage errors* are the result of situations in which all units of a population do not have an equal probability of inclusion in the sample that is drawn to represent the entire population. *Sampling errors* are a consequence of only surveying a portion of the sample rather than all members. In online surveys increasing the number of respondents as a means of reducing sampling error is not particularly difficult. However, in online surveys the sampling error cannot be calculated since the underlying assumption behind this calculation requires knowledge about the probabilities of selection. Sills and Song (2002) discuss the issue of sampling error and recommend including entire populations of given groups in order to avoid this type of error. Inasmuch as sampling and

coverage errors are difficult to overcome in online surveys, researchers should, according to Dillmann et al. (1998b), direct effort at decreasing the occurrence of measurement and non-response errors through, for instance, designing respondent-friendly questionnaires.

#### **4. Designing the Online Questionnaire**

Sheehan and McMillan (1999) emphasize that researchers considering employment of online surveys should examine what is known about postal surveys as this constitutes the most valuable source of information for the development of standards for email and Web based surveys. A number of ways for increasing the response rate and turn-around time of postal surveys can also be applied to online surveys. An example of this is tailoring a questionnaire to the interests and style of the target audience. Knowing something about respondents' attitudes, perceptions, needs, decisions, behavior, lifestyle, and demographics may have implications for how (sections of) questionnaires are prepared. A problem here is, according to Manfreda et al. (2002), that Web questionnaires are often designed by persons with little or no training in survey methodology, resulting in poor questionnaire design.

Swoboda et al. (1997) recommend applying a number of principles, such as simplicity, cultural independence, completeness, relevance and neutrality when designing email surveys. Others (e.g., Perseus Development Corporation, 2000) remind us to respect the intelligence of respondents and their time constraints. These principles, we would argue, are not the special requirements of online surveys, but apply to all types of surveys. Length, however, is an aspect specifically relevant to surveys that are electronically distributed over the Internet. Sheehan and McMillan (1999) suggest that the length of a questionnaire is, more than in postal surveys, relevant to email surveys, as an average print page can take up the space of several computer screens. Hence, conventional surveys may give the impression of an unduly long questionnaire when sent by email, which can in turn negatively influence the response rate. A basic rule-of-thumb they propose is: the longer the questionnaire, the less likely people will respond. Dillman et al. (1998a) emphasize that response rate and accuracy of responses can be increased by attending to access problems, and the motivation and cognition of respondents during the design of the questionnaire. In a review of three email surveys Sheehan and McMillan (1999) found that respondents seemed to be much more willing to reply to open-ended questions in an email format than in traditional paper surveys. This suggests that researchers need not be reluctant to incorporate open-ended items in their questionnaires.

Medlin et al. (1999) comment on the function of dynamic graphics which can be incorporated in a WWW-based survey so as to reduce fatigue

on behalf of the respondents, allowing use of a longer questionnaire. However, Dillman et al. (1998b) in a comparison of the influence of plain versus fancy design on response rates and on the quality of data from online surveys found that a plain questionnaire provided better results in terms of response rate, completeness, and completion time than a fancy version of the same questionnaire. Researchers should realize, in other words, that the more sophisticated a questionnaire is, the greater emphasis that will be placed on the computer hardware and software respondents have at their disposal. This can, understandably, have a negative influence on the survey response rate. Similar conclusions were drawn by Manfreda et al. (2002) from an examination of the impact of three basic Web survey design features on response and satisfaction.

Couper et al. (2001) examined the systematic effects of Web questionnaire design on the behavior of respondents. They found that the presence of a progress indicator reduces respondent loss, and that the use of multiple item screens generate faster completion times and reduction in missing data. In addition, the employment of radio buttons resulted in less missing items, whereas employment of text box entries increased the quality of responses.

## 5. Data Collection

There are three main ways to electronically distribute questionnaires:

- sending respondents the entire questionnaire in an email message (the email survey, with or without attachment);
- emailing respondents an introductory letter with a hyperlink to the Web based survey;
- placing a general request for respondents in an electronic communication environment (e.g., a newsgroup) or on a Web page.

In this section the distribution procedures employed in email versus Web based surveys are discussed.

### 5.1. QUESTIONNAIRES SENT BY EMAIL

Questionnaires sent by email leave respondents with the relatively simple task of completing and returning the document. In the email version of online surveys, a questionnaire can be sent to respondents as text included within a conventional email message. To complete the survey, respondents use the 'reply' function of their email systems, having inserted their responses into the text of the message returned. Whereas emailed questionnaires may not be much of a burden to respondents, they can pose numerous difficulties for researchers. Pitkow and Recker (1995) note:



“... once the responses have been submitted, the collation of the data can become problematic, since consistent structure within responses can only be suggested, not enforced. For example, if the question is posed “How old are you?” the answer may appear on the same line as the question, two lines below, may contain fractions, an integer, or even a floating point number.”

Another type of email survey is based on HTML code.<sup>4</sup> This type of email survey requires that all potential respondents are equipped with email systems that can process HTML. There are two main advantages of this type of survey as compared to more conventional text-based email surveys. First, there is a large variety of options for producing an attractive instrument design. Second, there is increased researcher control over answers by respondents since the form can be constructed only to accept pre-coded responses (Mann and Stewart, 2000: 68).

A third alternative for email surveys is to send the questionnaire as a file attached to an email message, for instance, as a document prepared by a word processor or spreadsheet software. Although this provides control over the appearance of the survey, its completion and return require a number of technical tasks by respondents – opening the file, saving it after completion and attaching the saved file to the return mail. In order to reduce the risk that some respondents may not have the software necessary to open the attached file, a program file as additional attachment to the respondents can be sent.<sup>5</sup>

A general assumption, based on studies of postal surveys, is that use of a (personalized) cover letter increases survey response rate. Sheehan and McMillan (1999) conducted a meta-analysis on several methodological issues in three email surveys. They found evidence that the speed of response to email surveys is faster from individuals who received pre-notification of the survey than from respondents approached without such notification. According to Crawford et al. (2001) the e-mail invitation plays a disproportionately important role in eliciting responses to a Web survey. In postal surveys potential respondents can peruse the entire questionnaire and make informed judgements about matters such as content, length and design. In Web surveys the e-mail letter is the only piece of information on the nature of the questionnaire. In spite of this, “many of the standard tools available to the mail survey researcher (e.g., type of postage, type of envelope, personalized signature, use of legitimizing letterheads, and inclusion of incentives), do not have direct electronic equivalents” (Crawford et al., 2001: 160).

In the email surveys reviewed by Kay and Johnson (1999) participation of respondents was solicited in several ways. In some studies respondents were emailed cover letters and received the survey via email only upon

agreement to participate. In other studies the electronic questionnaire was sent unsolicited and without advance warning, leading to critical and hostile responses which forced the researchers to revise their strategy.

## 5.2. WEB BASED QUESTIONNAIRES

The difficulties of email surveys described earlier in this section might be overcome by Web based questionnaires, although these require respondents to be able to access Web browsers. Kay and Johnson (1999) emphasize here the rule of thumb that the more burdensome the task, the lower the response rate. Participation in a Web survey is sometimes solicited by sending an email message in which a link to the Web based questionnaire is included. Most email software offers transformation of URLs into direct links to the Web site, hence reducing additional actions on behalf of the respondents.

Whereas Kay and Johnson (1999) note a number of objections to using Web based surveys from the perspective of respondents, Pitkow and Johnson (1995) emphasize the advantages of Web based surveys to email surveys as seen from the perspective of a researcher. These authors suggest six main advantages of the use of Web technology:

- possibility of point-and-click responses;
- provision of structured responses;
- use of an electronic medium for data transfer and collation;
- provision of visual presentation of the questions permitting review;
- flexible time constraints for respondents;
- employment of adaptive questions to reduce the number and complexity of questions presented to users.

Vehovar and Manfreda (1999) provide a number of examples of surveys among general populations placed on the Web open to anyone willing to complete them, such as general self-selection surveys (e.g., Gvu surveys conducted by the Georgia Institute of Technology, Graphic, Visualization and Usability Center), Web survey panels (e.g., Harris Poll Online), or online voting by means of short questionnaires on political or public issues.<sup>6</sup> This kind of distribution of Web surveys can, in fact, be considered a form of indirect mailing. Indirect mailing requires reliance on the chance that users of the WWW might come across the questionnaire and feel motivated to complete it. Such surveys are prone to various forms of bias and consequently unlikely to provide representative results.

Researchers employing Web based surveys for investigating purposive samples often require online respondents to enter a survey identification, ensuring that only the selected sample completes the questionnaire (e.g., Pitkow and Johnson, 1995; Bertot and McClure, 1996) because

“[once] a survey is posted to the Web, anyone who stumbles across it can fill it out.” (Kay and Johnson, 1999: 326). Smith (1997) argues however against the use of survey identification procedures because this can be considered a breach of respondent anonymity, and thus negatively influencing response rate. Heerwegh and Loosveldt (2003) approached this issue empirically and conducted a study on the effect of employing different access control measures (manual, semiautomatic and automatic) in order to prevent uninvited respondents from taking part in the study, and multiple submissions by the same respondent. They found that a semiautomatic login procedure (one access code to be typed in) does not decrease response rates, whereas it does increase the overall degree of data quality.

### 5.3. WEB BASED QUESTIONNAIRES: PROBLEMS ENCOUNTERED

In one of our own studies employing a Web based questionnaire we received feedback from respondents which illustrates some of the problems and challenges online surveys pose to respondents. The objective of this study was to explore the opinions of elderly Internet users regarding an electronic political debate in which they had been involved on a Web site in the Netherlands (Jankowski and Van Selm, 2000a). The electronic debate itself formed our sampling frame, as every posting in the debate was furnished with the email address of the contributor. Potential participants were addressed by an introductory email in which a link to the Web based questionnaire was included. The email was sent to 63 potential respondents of whom 29 returned the completed questionnaire, resulting in a response rate of 46%. Some potential respondents, however, did not feel comfortable with the manner in which they were addressed. One respondent was also skeptical about the confidentiality of his remarks in as much as the header of the email contained all of the addresses of the other respondents. We had not, in other words, employed the ‘batch’ or ‘blind carbon copy’ option in sending email messages. His reaction, shared per email, vividly illustrates this skepticism:

“Well, it could be good (for you) to fill in this form, but I better not. Do you want to know why? ‘All responses will be treated confidentially’, but what do I see in the address column? I see all the email addresses of those you’ve sent this message to. Do you folks call that confidentiality!? I’ve decided not to participate in this ‘carefully composed’ study, although I do have an opinion on the subject matter.”

Sheehan and McMillan (1999) consider assurance of respondent anonymity a key issue in the debate about the potential of online surveys. It is difficult to achieve full anonymity. Such a degree of anonymity may be

attained when respondents employ anonymous remailers to respond, but this is probably beyond the technical competence of most potential respondents. Confidentiality, however, can be considered a satisfactory alternative. Confidentiality can be assured by informing respondents that their email addresses will not be recorded with their survey responses, in addition to the fact that the survey data will only be analyzed at the aggregate level.

Another experience from this study made us aware of an unforeseen factor that may have influenced the response rate in a negative way. One respondent expressed irritation about an English language signature appended to the Dutch language email message. Although many European academics consider presentation of contact information in English commonplace, this is not necessarily the case among other groups in the population. It would have been better to have prepared a Dutch language signature for correspondence with these survey respondents.

In addition, another unanticipated objection to participate in the survey was the costs incurred on behalf of the respondents during the time used for filling out of the questionnaire. Many Internet users (in the Netherlands and in other European countries) still access the Internet via telephone modems. Time online costs money and the following message from a respondent illustrates this problem:

“Indeed, I received your request to fill out your questionnaire, but I don’t want to do this online. I am sorry but I don’t feel like spending time on the telephone line for this purpose. In case you could email the questionnaire, I would be willing to participate.”

This comment reflects, in fact, the general principle followed in postal surveys of including a stamped envelope in which the questionnaire is to be returned. Researchers have the tendency to forget that respondents often must pay for the time they are online. Such people may feel reluctant to participate in an online survey, and also expect some form of reimbursement for such expenses. One possible solution to this problem would be to allow respondents to print out the email and return the completed form by a conventional postal service free of charge. Such a mixed mode design, in which both email and paper-and-pencil questionnaires are distributed, is one of the recommendations made by Schaefer and Dillman (1997).

These examples suggest that a different kind of relation may exist between a researcher and respondents to an online survey, compared to this relation when conventional survey methods are employed. Many email messages we received from respondents expressed a kind of ‘eagerness’ to discuss the value of the questionnaire and to negotiate the conditions for completing the instrument. To a degree, the roles of researcher and respondent are reversed: respondents start posing the questions and researchers are expected to provide answers. This is not only potentially valuable because of the additional

information such a situation provides, but also because such interactions may help to increase the survey response rate. In another online survey, conducted among listeners to Web radio, we were able to ‘persuade’ some members of the sample to complete the questionnaire after they had posted email messages to us (Van Selm, Jankowski and Kleijn, 2004).

## 6. Response

### 6.1. RESPONSE RATE

The main problem with Web surveys is the impossibility of calculating the response rate as “There is no way in which to know how many individuals might have seen the survey or its links but declined to participate. Only the number of completed surveys is known and not the number of refusals” (Kay and Johnson, 1999: 326). One way to deal with this uncertainty is by placing a counter on the survey site that keeps track of the number of times that a site has been accessed. The response rate can then be estimated by dividing the number of returned surveys by the number of times that a site was accessed. Kay and Johnson (1999: 326) note, however, that

“...counters typically do not keep track of the number of unduplicated visitors. For example, 10 ‘hits’ could mean that each of 10 different individuals was exposed to the survey once, or it could mean that each of 2 visitors accessed the survey 5 times.”

In addition, counters do not reflect whether a page is allowed to fully load in the browser (Smith, 1997). Response rate estimates derived from counters, in other words, are questionable. Despite this uncertainty, some studies report and discuss response rates in Web based and email surveys. Crawford et al. (2001) review a number of studies in which lower response rates for Web surveys are reported compared to traditional mail methods. Schaefer and Dillman (1998) conclude, in a review of the literature on email surveys, that online surveys have generally failed to meet the response standard set by comparable mail techniques. They found only one study that reported an acceptably higher response by email (63%), as compared to regular mail (38%). They do suggest, however, ways in which the response to email surveys can be improved. One way is to try successively contacting potential respondents – a technique common in postal, telephone, and face-to-face interviewing. Reminders to potential respondents of an online survey can be automated, an aspect much more difficult with other survey methods. In postal surveys researchers must allow a number of weeks time before sending a reminder but with online surveys these can be sent a few days after the original questionnaire was announced or distributed. This reduces loss of respondents who may tend to postpone and

then forget to complete the questionnaire. Another suggestion Schaefer and Dillman make is to send a personalized letter to respondents prior to distributing the questionnaires. Personalized letters can stress the importance of every respondent for a study. A third way is to develop a mixed-mode survey strategy where email is complemented by other methods for reaching respondents, particularly those without access to the Internet.

Schaefer and Dillman (1998) conducted an experiment among the faculty of an American university in which they examined the value of such a multi-mode strategy in order to obtain responses from individuals unreachable through email. The strategy involved using a combination of paper and email for sending a preliminary announcement of the study, a letter accompanying the questionnaire, a thank you or reminder letter, and a replacement questionnaire. The design involved four treatment groups for which the channel was modified (paper versus email) regarding four items of the survey (pre-notice, letter and survey, thank you/reminder, and replacement survey). Results of this experiment suggest that, when a mixed-mode strategy is used in order to obtain responses by mail for sampled individuals who do not have email addresses, comparable rates can be obtained from regular mail surveys and online surveys. Through inclusion of a paper element in the mixed mode design, the problem of coverage error was eliminated. In addition, other advantages of email surveys included quick responses, a slightly lower level of non-response for questionnaire items and more complete answers to the open questions included in the survey. Schaefer and Dillman (1998: 391) suggest that the lower non-response on items in email questionnaires might be explained by:

“the proposition that moving visually through an e-mail message requires more effort than filling out a paper questionnaire. Since individuals can only view a few questions at a time (fewer than in the paper version), each question may be less likely to be overlooked.”

Patrick et al. (1995) coin the term ‘random encouragement’ to describe the way they increased the response rate in their study of characteristics of the users of the National Capital FreeNet in Ottawa, Canada. Random encouragement means that a random sample, derived from persons listed in the sampling frame who had not yet volunteered to complete the questionnaire, received a personal invitation via email to complete the survey. This increased the response rate by 10%.

Sheehan and McMillan (1999), in a review of online surveys, note that many researchers seem to comment positively about the response rate and speed of email surveys, as compared to postal surveys. Few, however, have considered why these aspects vary. The authors investigated this aspect in three email surveys and found that issue salience had a positive effect on response rate; further, sending an announcement prior to the survey

increased response time. Sheehan and McMillan suggest researchers should consider making proper use of the subject line of email messages to inform potential respondents of the study.

## 6.2. RESPONSE TIME

One of the attractive features of responding electronically is the ability to monitor whether respondents have received an email message and when completed questionnaires have been returned. In our study among elderly Internet users who had participated in a political debate on a Web site (Jankowski and Van Selm, 2000a), the first responses were received within 3 hours after the email messages had been posted. This was an encouraging experience that we would not have had with a postal survey. Other researchers (Schaefer and Dillman, 1998) found in a study among the faculty of an American university that more than three-fourths of all completed email questionnaires were returned within 4 days from the time of mailing. And, in a Web based survey among an unrestricted sample of readers of the electronic version of a women's magazine conducted by one of our students, nearly 2500 respondents had responded to the survey within 2 weeks after the questionnaire had been posted behind a banner on the magazine Web site. The study by Crawford et al. (2001) provides illustrations as to how researchers monitored in real time respondents' reactions to reminders and abandonment behavior.

## 6.3. INCENTIVES

Use of respondent incentives is another way to possibly increase the response rate to surveys. Respondent incentives may, however, be more complicated for online surveys than for postal surveys. For most of such initiatives, it is uncertain whether they work and, if so, where the point of maximum benefit is – with a nominal lottery prize or with a financial remuneration commensurate with the amount of time required to complete the questionnaire. GVU has taken more extreme action and offered cash stipends between \$100 and \$250 to participants in their user surveys (Kay and Johnson, 1999).

Bosnjak and Tuten (2003) examined whether prepaid monetary incentives have the same positive effects on response rate in Web surveys as shown in traditional mail surveys. The authors found no positive effect of prepaid incentives on willingness to participate, actual completion rates, and share of incomplete response patterns when compared to postpaid incentives. Instead they found that prize draws compared to no incentives significantly increased the willingness to participate and tend to decrease the number of incomplete responses.

Brennan et al. (1999) advise caution in using incentives because those that are too attractive may well prompt multiple submissions from, for example, people wanting to increase their chances in winning a lottery drawing. 'Immaterial' incentives, however, may be an alternative. Smith (1997) offered two such incentives in a survey of Web providers. The first was opportunity to be listed on a special Web page thanking participants in the survey that was linked to the executive summary of the research report, and the second was an increased awareness of the survey topic through self-reflection during completion of the questionnaire.

## **7. Processing Data**

One of the major advantages of online surveys is that data often can be used more or less directly for analysis. In as much as respondents, when filling in a questionnaire, enter data directly into an electronic file, there is no need for a separate phase of data entry. This reduces costs and the time necessary to complete the research project. Another characteristic of online surveys is the automatic coding of close-ended questions by the computer, leaving only open-ended questions to be manually coded.

Data cleaning remains necessary in online surveys, but can be efficiently performed with Internet facilities. Employment of 'cookies', for example, may provide some value during this part of the survey. Cookies are files installed on a respondent's computer system when a Web site is visited. When a respondent completes a survey, a cookie is placed on the computer, and in the event the person tries to respond to the same survey again, the system recognizes the cookie and restricts access. Although seemingly an ideal solution to double responses, Internet browsers can be programmed in such a manner to prevent cookies from being placed on a computer. Respondents who restrict placement of such cookies on their computers are thus excluded from the survey, which in turn introduces a new source of non-response.

Another method sometimes proposed for detecting multiple questionnaire completion is by checking IP numbers. IP numbers are unique identification characteristics by which a computer connected to the Internet can be identified. Multiple responses from one IP number (i.e., from a single computer) may be a source of error. However, some computers may have a multi-user function – as is often the case in libraries, Internet cafes and in the generally accessible rooms with computers at universities. Hence, different Internet users may legitimately have the same IP address. Another problem with this strategy is that many Internet users are assigned IP addresses dynamically by Internet service providers, which means their unique identifiers differs for each dial-up session.



In the Web based survey among the readers of the electronic version of a women's magazine conducted by one of our students one of the measures that was taken to prevent generation of invalid data involved the employment of a filtering program that used the time of submission of the completed questionnaire as a unique identifying characteristic. Questionnaires returned within the interval of 2 min were subjected to an additional manual check. A total of 400 multiple responses were identified in this manner. An explanation for this large number might be the fact that the server on which the Web survey was hosted was sometimes slow, resulting in situations where the questionnaire would be transmitted after pressing the 'submit' button, but without simultaneously indicating this action to the respondent by means of an automatic 'Thank you' message. The natural reaction of respondents in such a situation would be to press the submit button again, resulting in double responses.

### **8. Supportive Technology and Software**

Supportive technology can be employed in all phases of Internet surveys. Specific survey-generating software can support researchers in constructing email or Web based questionnaires.<sup>7</sup> These software products provide a number of options with respect to decisions about survey layout, order of questions and types of responses for each question (e.g., ranking, single choice, Likert Scale, open-ended responses). Web software programs prepare the questionnaire file, store it on the computer's hard drive, and are also able to generate computer screens for completion by respondents. Medlin et al. (1999) note a number of additional functions offered by some of these programs, including the ability to:

- check for non-completion of questions;
- require completion of all questions before allowing respondents to proceed;
- automatically control for branching according to respondent answers;
- vary the order of questions during instrument testing;
- monitor response time for sections or for the whole instrument.

Mann and Stewart (2000) provide examples of commercial computer programs that not only assist in producing the text of a survey, but also interpret the emailed responses and read the answers directly into the database. Pitkow and Recker (1995) suggest that the major added value of Web based surveys lies in the use of adaptive questions, and supportive methods available for tracking users' responses across different surveys, thereby allowing more in-depth analyses of survey responses. The allocation of respondents to different versions of questionnaires (both in respect to the technology and the content) can be accomplished for WWW sur-

veys by means of Common Gateway Interface script (CGI-scripts).<sup>8</sup> This means that rather than having a single questionnaire for all combinations of questions, CGI scripts are used to determine which sections of a questionnaire are appropriate for a particular respondent. Brennan et al. (1999: 85) describe how they applied this feature in one of their studies:

“... the questionnaire was broken into separate sections. Submitting one section invoked a CGI script, which determined which section (treatment) the respondent was allocated to next. Because the first page of the Web survey was common to all respondents, it was used to collect Webographic and demographic data. This Webographic information was used; for example, to ensure that a respondent was not allocated to a treatment involving sound if their computer was not equipped to handle it.”

Another example of supportive software is the use of parsing programs that can be applied in various ways, for instance in order to scan newsgroup messages for email addresses and to store them automatically in a file (Swoboda et al., 1997). Also for preprocessing data, enabling the coding or writing to a output file of different types of survey questions (i.e., free text and multiple-choice questions), parsing programs can be useful.

## 9. Summary and Discussion

In this research methods review we have discussed challenges that researchers encounter while conducting online surveys. We presented literature on online survey and started by noting various reasons for using the Internet as a tool for conducting surveys, such as:

- the object of study may relate to Internet use;
- it may be expedient to consult key persons in well-defined Internet user groups;
- respondents with deviant or covert behaviors may be more easily recruited via the Internet;
- school-age respondents with an affinity for computers may find online surveys attractive.

The main benefits of online surveys discussed include: reduction of costs and time, ease in reaching large numbers of potential respondents, and the possibility of providing anonymity to participants. Each of these benefits, however, has a ‘price’, and we have mentioned a number of qualifications regarding the use of online surveys. But once a decision has been taken to conduct an online survey, several issues immediately follow:

- choosing between an email or a Web based questionnaire;
- selecting, when possible, a sampling frame;
- avoiding sampling and coverage errors.

The online survey questionnaire must then be constructed. In general, the same principles used for the design of traditional surveys are applicable to electronic instruments, but additional attention should be given to questionnaire length. A typical self-completion questionnaire, administered by conventional mail, may seem excessively long when placed on a Web site. It is also important to estimate the computer equipment respondents may be using as well as their competence with particular programs so as not to expect more from them than is realistic.

Depending on the type of online survey, the questionnaire is distributed via email or the Web. In both cases a number of principles should be applied during the collection of data:

- sending a pre-notification to prospective respondents tends to increase response rate;
- measures should be taken to ensure anonymity or at least confidentiality;
- email correspondence from prospective respondents may form a valuable source of information and serve to further reduce the dropout rate.

So far, online surveys have seldom achieved response rates comparable to those of postal or telephone surveys. A number of measures have been suggested, however, that may increase response rates:

- undertake multiple attempts to contact potential respondents (through, e.g., pre-notifications, reminders and replacement surveys, and thank you notes);
- utilize a mixed-mode strategy, including both electronic and pen-and-paper questionnaires, in order to reach respondents without access to the Internet;
- try to ensure that the survey topic is relevant to the target group;
- use respondent incentives to stimulate questionnaire completion.

The electronic collection of data provides several advantages over traditional surveys during data processing, such as elimination of a separate phase for data entry and automatic coding of close-ended questions, checks for multiple responses through use of computer cookies, through checking IP numbers of survey participants, and by filtering out questionnaires returned within a time-interval of a few. Finally we briefly discussed that in various phases of the research process supportive technology and software can be employed. Software products are available that support the

construction of questionnaires, storage of the document on the Web, and allocation of respondents to different versions of the questionnaire. Parsing programs form another set of tools that can be employed when selecting potential respondents from sampling frames. Various other processes can be automated, such as construction of the data file, cleaning of data and sending reminders to sample members who have not yet responded to the online survey.

After summarizing the main challenges and possible remedies, at least two issues stand out. We suggest, first of all, that the problems researchers and respondents encounter while conducting or responding to an email or Web based survey are often remedied when principles employed in conventional surveys are taken into account. Secondly, as the arena of online research is developing at a rather rapid pace, both researchers and developers of survey software are increasingly gaining expertise and continuously improving solutions to (potential) problems. We urge scholars that are interested in utilizing online surveys in their work to monitor these developments closely.

## Notes

1. Other names are also in vogue; sometimes Web page survey, or Web questionnaire are used to describe what we mean here.
2. Examples include the journals *Social Science Computer Review*, *Sociological Research Online*, *Journal of Computer-Mediated Communication*, and the mailing list: [virtual-methods@mailbase.ac.uk](mailto:virtual-methods@mailbase.ac.uk). See also the mailing list of the Association of Internet Researchers: [air-l-request@aoir.org](mailto:air-l-request@aoir.org).
3. A MUD is an acronym for 'multi user domains' and stems from the early adventure game 'Dungeons and Dragons'. MUDs are both *synchronous* and *hypertext* forms of communication, that combine visual and textual ingredients in order to create a sense of place and community among players (Marvin, 1996).
4. HTML, Hypertext Markup Language, is the standard computer language used for preparing pages for the Internet via the World Wide Web.
5. A limitation to this alternative is that programs created for the Windows operating system are not always compatible with the Apple Computer operating system, and vice versa. Other limitations are that large files may not be accepted by some service providers, and some organizations prohibit attachments to email messages altogether because of concern for computer viruses.
6. For Harris Poll Online, see <http://www.harrispollonline.com>. Sites of newspapers or television stations often offer the online polling function; see Jankowski and Van Selm (2000b).
7. A large variety of software supporting email and Web based survey research is presently available, provided by companies specialized in online survey research and readily found through a Google Web search. A small number of such products includes: MaCATI (<http://www.senecio.com/>), Survey Internet (<http://www.aufrance.com/>), Survey Connect (<http://www.surveyconnect.com/>), Survey Solutions ([www.perseus.com](http://www.perseus.com)), SurveyTracker ([www.surveytracker.com](http://www.surveytracker.com)). The major statistical software producers, such as SPSS, also have specialized products: <http://www.spss.com/spssmr/>. See also <http://www.websm.org/>. For an overview of online survey software, see Wright (2005).

8. GGI is a specification for transferring information between a World Wide Web server and a CGI program. A CGI program is any program designed to accept and return data that conforms to the CGI specification. CGI programs are the most common way for Web servers to interact dynamically with users. Many HTML pages that contain forms, for example, use a CGI program to process the form's data once it is submitted. Another increasingly common way to provide dynamic feedback for Web users is to include scripts or programs that run on the user's machine rather than the Web server. One problem with CGI is that each time a CGI script is executed, a new process is started. For busy Web sites, this can slow down the server noticeably (<http://webopedia.internet.com>).

## References

- Babbie, E. (1990). *Survey Research Methods* 2nd ed. Belmont, CA: Wadsworth.
- Beebe, T. J., Mika, T., Harrison, P. A., Anderson, R. E. & Fulkerson, J. A. (1997). Computerized school surveys. Design and development issues. *Social Science Computer Review* 15(2): 159–169.
- Bertot, J. C. & McClure, C. R. (1996). Electronic surveys: methodological implications for using the World Wide Web to collect survey data. In: S. Hardin (ed.), *Global Complexity: Information, Chaos and Control*. Medford, MA: Information Today, pp. 173–192.
- Bosnjak, M. & Tuten, T. L. (2003). Prepaid and promised incentives in Web surveys. *Social Science Computer Review* 21(2): 208–217.
- Brennan, M., Rae, N. & Parackal, M. (1999). Survey-based experimental research via the web: Some observations. *Marketing Bulletin* 10: 83–92. <http://marketing-bulletin.massey.ac.nz/article10/article9b.asp>
- Coomber, R. (1997). Using the Internet for survey research. *Sociological Research Online*, 2 (2). <http://www.socresonline.org.uk/socresonline/2/2/2.html>
- Couper, M. P., Traugott, M., & Lamias, M. (2001). Web survey design and administration. *Public Opinion Quarterly* 65: 230–253.
- Crawford, S. D., Couper, M. P. & Lamias, M. J. (2001). Web surveys. Perceptions of burden. *Social Science Computer Review* 19(2): 146–162.
- Dillman, D. A., Tortora, R. D. & Bowker, D. (1998a). *Principles for Constructing Web Surveys*. SESRC Technical Report 98-50, Pullman, Washington. <http://survey.sesrc.wsu.edu/dillman/papers.htm>
- Dillman, D. A., Tortora, R. D., Conradt, J. & Bowker, D. (1998b). *Influence of plain vs. fancy design on response rates for Web surveys*. Paper presented at Joint Statistical Meetings, Dallas, Texas. August 1998. <http://survey.sesrc.wsu.edu/dillman/papers.htm>
- Heerwegh, D. & Loosveldt, G. (2003). An evaluation of the semiautomatic login procedure to control Web survey access. *Social Science Computer Review* 21(2): 223–234.
- James, M. L., Wotring, C. E. & Forrest, E. J. (1995). An exploratory study of perceived benefits of electronic bulletin board use and their impact on other communication activities. *Journal of Broadcasting and Electronic Media* 39(1): 30–50.
- Jankowski, N. W. & Van Selm, M. (2000a). The promise and practice of public debate in cyberspace. In K. L. Hacker & J. A. G. M. van Dijk (eds.), *Digital Democracy: Issues of Theory and Practice*. London: Sage, pp. 149–165.
- Jankowski, N. W. & Van Selm, M. (2000b). Traditional news media online: An examination of added values. *Communications. The European Journal of Communication Research* 25(1): 85–101.
- Kay, B. & Johnson, T. J. (1999). Research methodology: taming the cyber frontier. Techniques for improving online surveys. *Social Science Computer Review* 17(3): 323–337.
- Kehoe, C. M. & Pitkow, J. E. (1996). Surveying the territory: GVU's five WWW user surveys. *World Wide Web Journal* 1(3). <http://www.w3.org/pub/WWW/Journal/3/s3.kehoe.html>

- Manfreda, K., Lozar, Batagelj, Z. & Vehovar, V. (2002). Design of Web survey questionnaires: Three basic experiments. *Journal of Computer-Mediated Communication*, 7(3), <http://jcmc.indiana.edu/vol7/issue3/vehovar.html>
- Mann, C. & Stewart, F. (2000). *Internet Communication and Qualitative Research. A Handbook for Researching Online*. London: Sage.
- Marvin, L. (1996). Spoof, spam, lurk and lag: the aesthetics of text-based virtual realities. *JCMC* 1(1). <http://www.ascusc.org/jcmc/vol1/issue2/marvin.html>
- Medlin, C., Roy, S. & Ham Chai, T. (1999). *World Wide Web versus mail surveys: A comparison and report*. Paper presented at ANZMAC99 Conference: Marketing in the Third Millennium. November 28 – December 1, Sydney, Australia. [www.singstat.gov.sg/conferences/ec/f112.pdf](http://www.singstat.gov.sg/conferences/ec/f112.pdf)
- Mehta, R. & Sivadas, E. (1995). Comparing response rates and response content in mail versus electronic mail surveys. *Journal of the Market Research Society* 37(4): 429–439.
- Mickelson, K. D. (1997). Seeking social support: parents in electronic support groups. In: S. Kiesler (ed.), *Culture of the Internet*. Mahwah, New Jersey: Lawrence Erlbaum, pp. 157–178.
- Miller, T. I., Kobayashi, M., Caldwell, E., Thurson, S. & Collett, B. (2002). Citizen Surveys on the Web: General Population Surveys of Community Opinion. *Social Science Computer Review* 20(2): 124–136.
- Patrick, A. S., Black, A. & Whalen, T. E. (1995). Rich, young, male, dissatisfied computer geeks? Demographics and satisfaction from the National Capital FreeNet. In: D. Godfrey & M. Levy (eds.), *Proceedings of Telecommunities 95: The International Community Networking Conference*. Victoria, British Columbia: Telecommunities Canada, pp. 83–107.
- Perseus Development Corporation (2000). *Survey 101- A Complete guide to a successful survey*. [http://www.perseus.com/surveytips/Survey\\_101.htm](http://www.perseus.com/surveytips/Survey_101.htm)
- Pitkow, J. E. & Recker, M. M. (1995). Using the web as a survey tool: Results from the second WWW user survey. GVU's WWW User Surveys. [ftp://ftp.gvu.gatech.edu/pub/gvu/www/survey/papers/survey\\_2\\_paper.pdf](ftp://ftp.gvu.gatech.edu/pub/gvu/www/survey/papers/survey_2_paper.pdf)
- Schaefer, D. R. & Dillman, D. A. (1998). Development of a standard e-mail methodology. *Public Opinion Quarterly* 62: 378–397.
- Sheehan, K. B. & McMillan, S. J. (1999). Response variation in e-mail surveys: An exploration. *Journal of Advertising Research* 39(4): 45–54.
- Sills, S. J. & Song, C. (2002) Innovations in Survey Research: An application of Web-based surveys. *Social Science Computer Review* 20(1): 22–30.
- Smith, C. B. (1997). Casting the Net: surveying an Internet population. *Journal of Computer-Mediated Communication* 3(1). <http://www.ascusc.org/jcmc/vol3/issue1/smith.html>
- Swoboda, W. J., Mühlberger, N., Weikunat, R. & Schneeweiss, S. (1997). Internet surveys by direct mailing. An innovative way of collecting data. *Social Science Computer Review* 15(3): 242–255.
- Van Dijk, J. A. G. M. (2000). Widening information gaps and policies of prevention. In K. Hacker & J. van Dijk (eds.), *Digital Democracy: Issues of Theory and Practice*. London: Sage, pp. 166–183.
- Van Selm, M., Jankowski, N. W. & Kleijn, B. (2004). Dutch web radio as a medium for audience interaction. In A. Crisell (ed.), *More than a Music Box: Radio Cultures and Communities in a Multi-media World*. New York, Oxford: Berghahn, pp. 265–282.
- Vehovar, V. & Manfreda, K. L. (1999). *Web surveys: Can the weighting solve the problem?* Paper presented at the 1999 AAPOR Conference, St. Petersburg, Florida, May 13–16. [http://www.amstat.org/sections/srms/Proceedings/papers/1999\\_168.pdf](http://www.amstat.org/sections/srms/Proceedings/papers/1999_168.pdf)
- Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication* 10(3), <http://jcmc.indiana.edu/vol10/issue3/wright.html>