Research Design:

Internet-Based Data Collection: Promises and Realities

Jacob A. Benfield
Department of Psychology, Colorado State University, Fort Collins, CO 80521-1876, USA
jake.benfield@colostate.edu

William J. Szlemko
Department of Psychology, Colorado State University, Fort Collins, CO 80521-1876, USA
william.szlemko@colostate.edu

Abstract

The use of Internet to aid research practice has become more popular in the recent years. In fact, some believe that Internet surveying and electronic data collection may revolutionize many disciplines by allowing for easier data collection, larger samples, and therefore more representative data. However, others are skeptical of its usability as well as its practical value. The paper highlights both positive and negative outcomes experienced in a number of e-research projects, focusing on several common mistakes and difficulties experienced by the authors. The discussion focuses on ethics and review board issues, recruitment and sampling techniques, technological issues and errors, and data collection, cleaning, and analysis.

Keywords: Internet; data collection; research ethics; sampling

1. Internet as a Research Tool

With the advancement of information and communication technology, researchers have found new methods of data collection and analysis. This has evolved from telephone surveys, computerized data analysis, and use of cell phones and pagers, to collecting information at random intervals, use of Personal Digital Assistants (or "PalmPilots"), and use of the Internet in research. Although the Internet is fast becoming a common fixture in contemporary life in many parts of the world, it remains relatively unused for primary data collection in many research fields. For example, social science research is yet to respond to the emergence of the Internet, as shown by only 494 peer reviewed articles with keywords "Internet research" published within major social science journals over the decade 1996-2006 (as per our search in the *CSA Illumina*® bibliographic database). Increasingly, however, the Internet is being treated as a rich source for literature and secondary data in social science research.

Until relatively recently, use of the Internet for primary data collection required the researcher to either know HTML or have someone else create a new program. Fortunately, within the past few years a number of new technological solutions and services have emerged that allow the researcher to create studies (i.e., surveys, experiments, etc.) online without needing the knowledge of computer programming. This has coincided with a large increase in studies using the Internet to collect primary data. A search in the *Web of Science*® bibliographic database indicates that the number of publications during the six-year period 2000-2005, using "Internet research" as keywords, is 128, which is 312 per cent higher than the corresponding figure during the six-year period prior to 2000, i.e.,1994-1999. Similar results are seen for "Internet data collection" (325 per cent), "web based research" (333 per cent), and "electronic data collection" (327 per cent). Of course, these impressive percentages are based on low base figures; Internet use in research still remains rather limited.

By its very nature, the Internet appears to be a very promising medium for researchers. As a vehicle for data collection, it promises increased sample size, greater sample diversity, easier access and convenience, lower costs and time investment, and many other appealing features. It is even possible to use the Internet for pilot testing media messages and advertisement campaigns. But without careful attention, the researcher may get into difficulties. It is the purpose of this article to expose some of the potential pitfalls awaiting the unwary researcher. Along with the potential pitfalls, solutions utilized by the authors are also discussed.

2. Manual vs. Internet-Based Data Collection

We have encountered a number of issues in our various attempts at using the Internet for primary data collection. A list of such issues must include those associated with research ethics guidelines, technical snags arising from power failures, data cleaning requirements, and low response rate. Sometimes, the experience has been so frustrating as to make manual data collection through paper-and-pencil research packets appear more attractive.

However, with experience, we have learnt to be judicious in selecting the appropriate data collection method for a given research project and taking the necessary precautions if we choose to use the Internet.

Researchers, especially psychologists, have often looked at the method of data collection with regard to the impact it can have on results. The issues of questionnaire design, for example the implications of using forced choice, Likert scales, open response, or multiple response formats, are all issues much older than the Internet (Orlich, 1978; Schuman & Presser, 1981; Sudman & Bradburn, 1982). These will always be important when designing data collection instruments. The design of the instrument should be informed by the research question being addressed. Any advantages or disadvantages offered by a specific question format will not be altered by technology, but technology may introduce additional issues (Manfreda, Batagelj, & Vehovar, 2002). Each of these response types is easily available in an electronic format. Some researchers have compared manual and electronic formats, examining the issues of *validity* and *reliability* of research instruments (Berrens, Bohara, Jenkins-Smith, Silva, & Weimer, 2003; Schilewaert & Meulemeester, 2005; Sethuraman, Kerin, & Cron, 2005). They have found test-retest reliabilities for both formats to be nearly equal, indicating that both formats can generate equally reliable data assuming that the participants are cooperative and truthful, and the questions are valid. They have also found internal consistency, predictive validity, and recruitment trends within socio-demographic categories to be comparable between the two formats. In essence, the mode of data collection (i.e., manual or electronic) does not, in itself, seem to significantly alter the type of respondent recruited or the quality of data given by the respondent.

Collecting data from people with poor reading comprehension or those not accustomed to taking paper-and-pencil tests is already known to be difficult. Similarly, while using electronic data collection methods, the respondents' lack of familiarity with computers could be an issue. In some of our survey research projects, we have compared the paper-and-pencil method with the computer based method. In our pilot tests, we have found that the computer based method was usually faster (because of the respondents' familiarity and ease with the computer keyboard and the mouse). However, during the actual data collection, the mobile laboratory had a touchpad instead of a mouse, which slowed down the respondents using the electronic version, in comparison with those who used the paper version. In short, computer skills and familiarity with the input devices affect a respondent's ability to complete an electronic survey. This is in addition to problems experienced by respondents who have poor reading comprehension or who are not comfortable with filling out questionnaires.

Another relevant difference between paper-and-pencil and electronic formats is the level of rapport possible with the respondent. The impact of such rapport may be unpredictable. For some respondents, the signed letter accompanying a paper-and-pencil format may be more persuasive than an e-mail from a stranger, commonly sent with the electronic format. It is uncertain whether face-to-face interaction with a person or the relative anonymity of the Internet would produce more authentic responses.

With *identity theft* (i.e., the deliberate assumption of another person's identity without the latter's knowledge) being a major issue of current concern, Internet data collection may not seem as legitimate as data collected in a community center or a university laboratory. Internet data collection could indeed be problematic from the point of view of *source credibility*--an important issue in persuasive communication, as research in the area of persuasion indicates (Hong, 2006; Hovland & Weiss, 1951; Olson & Cal, 1984). Additionally, as the psychologist Stanley Milgram (1974) argues, people are more likely to obey an authority that is present in the room compared to one that is in the next room or on the phone. Accordingly, the manual paper-and-pencil method can be expected to produce higher-quality data compared to the Internet-based method, the former being more tangible, more personal, and in short, more credible to the respondents, especially if the research staff is in the room with them (Nosek, Banaji, & Greenwald, 2002).

On the more positive side, Internet-based data collection, if utilized properly, can reduce costs and make unfunded projects feasible, yield larger and more representative samples, and obviate hundreds of hours of data entry. Table 1 compares the advantages and disadvantages of manual and online modes of data collection.

Table 1. *Comparison between Manual and Internet-Based Data Collection**

	Manual	Internet
Cost (in USD)	• $0.03-0.05 per sheet • Initial and return postage (Mail surveys) • 1,000 5-page surveys = $150-250 • 10,000 5-page surveys = $1,500-2,500	• Initial start-up fee ($300-2,000), for one year • No return postage even if using paper recruitment letter • 1,000 5-page surveys = $300-2,000 • 10,000 5-page surveys = $300-2,000
Sample Size	• Both large and small samples • Increased sample size increases cost	• Both large and small samples • Increased sample size does not change cost
Recruitment	• Any traditional form including phone, mail, posters, flyers, in-class, etc.	• Any traditional form as well as the use of e-mail and Web links
Data Entry	• Data are entered manually • Data are in *Scantron* format and entered electronically (increased cost for sheets and scoring service)	• Data are entered digitally as participant completes the project • Data often need reformatting to match statistical program or scale scoring system
Time Loss	• Researchers' time is taken up by data entry (back-end)	• Researchers' time is taken up by setting up the project electronically (front-end)

*Assumes electronic data collection is being facilitated by a third party Internet survey company and not by computer savvy researchers

Internet is a tool that is out there, for better or for worse. Its usefulness in research is largely dependent on its judicious use. As depicted in Figure 1, a series of questions pertaining to different stages of the research project need to be answered before making a

final choice regarding the data collection format. In this figure, the solid lines represent the progression of the decision making process concerning the use of electronic data collection. The broken lines lead to the likely decision, with the lines on the right representing a negative answer to the question posed at each stage (thus favoring manual data collection) and the lines on the left representing a positive answer (thus favoring electronic data collection).
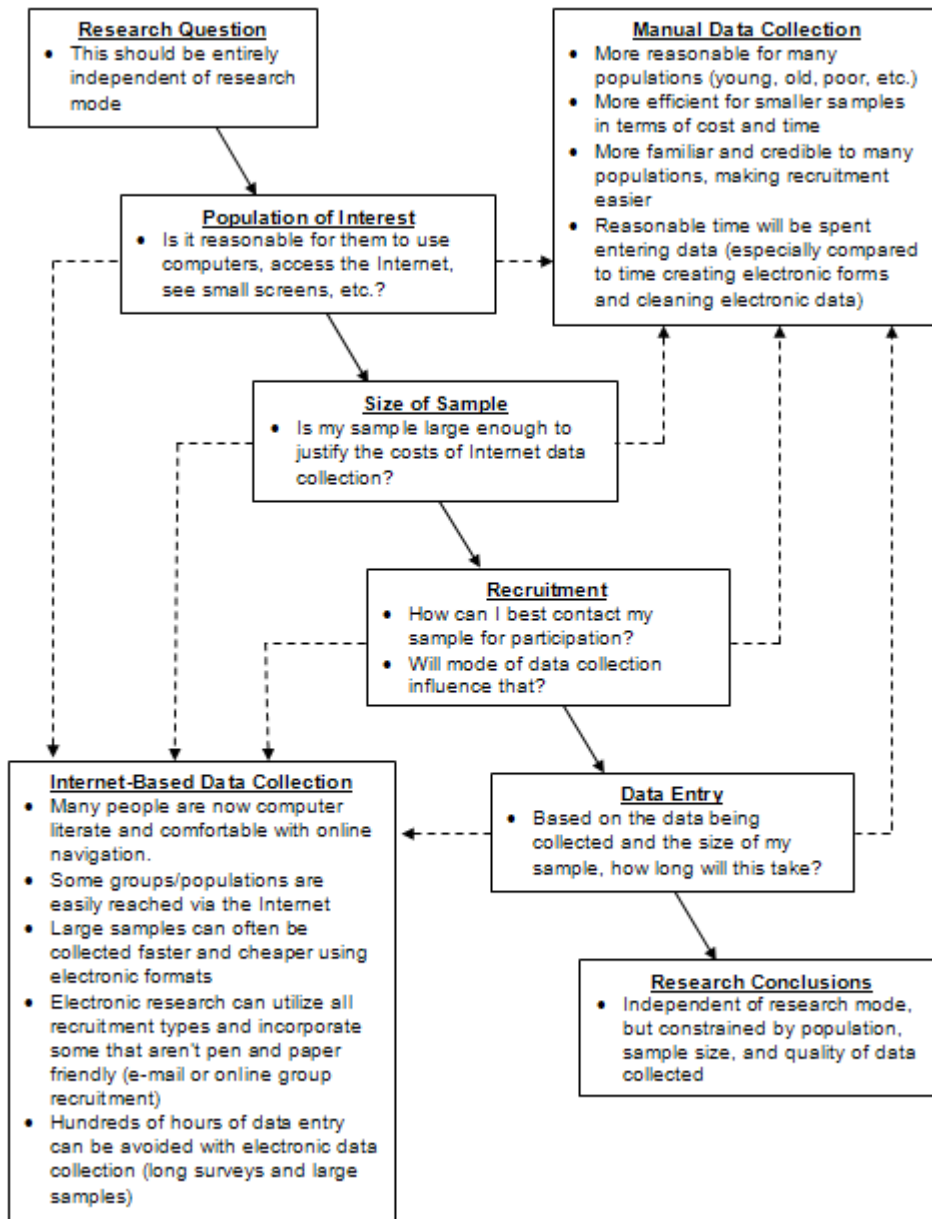

Figure 1. Considerations for Incorporating Internet-Based Data Collection in a Research Project

3. Research Ethics

The Institutional Review Board (IRB) is the US version of the research ethics committees created in many universities and other research institutions in response to the rising concerns about both human and animal use in research. The IRB's role is to oversee research being conducted within an institution in an attempt to ensure that participants' rights and privilege are being upheld. In the United States, IRBs generally focus on the principles laid out in the *Belmont Report* (1978). When considering whether or not a specific research project should be allowed to be completed, IRB reviewers focus on three key principles: (a) *beneficence* (i.e., lack of harm and/or received benefit), (b) *respect for persons* (i.e., confidentiality and ability to withdraw from research), and (c) *justice* (i.e., opportunity for all participants to benefit from outcome). In essence, the IRB serves as the research participants' informed and trained advocate.

Some IRB members may have some special concerns when dealing with proposals involving primary data collection via the Internet (Naglieri et al., 2004; Nosek, Banaji, & Greenwald, 2002). Anonymity and confidentiality are always concerns in data collection, but the potential for recording the IP (Internet Protocol) addresses, thereby the identity of the remote computers, makes Internet-based proposals more complicated (Berry, 2004). Other issues, such as data security during transmission, are unique to Internet-based data collection. Some common IRB issues the authors have encountered are discussed in the following paragraphs.

Primary data collection via the Internet presents a unique issue during data transmission (Hewson, Laurent, & Vogel, 1996). The data are most susceptible to hacking, corruption, etc., while these are being transferred from the respondents' computers to the researchers' computer. One relatively easy method of limiting these possibilities is the *encryption* of data during transmission. Data encryption may be accomplished through various methods, but from the IRB viewpoint, the method of encryption appears to be of less importance than the fact that encryption is being done. Of course, providing for data encryption can add to the cost of the project.

Irrespective of the mode of data collection, physical security of data is a major issue once data have been collected. With Internet-based data collection, physical security includes much more than a locked file cabinet in a secure room. Consideration must be given to both physical and electronic security of the server where data are stored. Physical security of the server should minimally include a room with restricted access. Internet data collection can be facilitated by numerous agencies that specialize in allowing researchers to create their own study. These agencies often provide adequate physical security. One physical security measure that may be overlooked is environmental controls that regulate temperature, humidity, and air flow. Environmental controls are particularly relevant for electronic data. Papers locked in a file cabinet will not be affected by a 105 degree Fahrenheit temperature, but this may cause problems with computer hard-drives. These extensive safeguards may not be necessary depending on the IRB, but having them will provide peace of mind for researchers and IRB members alike. Electronic security begins with the encryption process described above; it does not, however, end there. It would be

necessary for the server to have *firewalls*. Firewalls protect the server from unauthorized electronic entry (i.e., hacking). Other electronic security commonly includes the use of passwords, PIN codes, and access codes.

When conducting Internet surveys, there is a potential threat to anonymity of the respondent that needs to be considered (Pittenger, 2003; Waern, 2001). It is possible for a computer program to record the IP address of the computer being used by the respondent. The IP address is a numerical code that is unique to each computer connected to the Internet. It is also possible to record the time when the data were entered. These capabilities mean that the actual respondents can be traced out in many cases. We have dealt with this issue by either deleting the IP addresses from the dataset early in the cleaning process or electing to not record the IP addresses, wherever possible. As an interesting aside, IP addresses collected from personal computers may be useful for matching sets of longitudinal data without collecting specific identifiers or using matched lists of identities and participant codes. In this case, recording IP address is an advantage--not an ethical liability. However, IRBs should be made aware that this is the intent behind recording IP addresses in such a case.

Research involving persons requires some form of *informed consent*, wherein the persons agree to participate and acknowledge the risks, benefits, and their rights. This can take the form of a verbal consent or a written one. In both verbal and written consents it is ascertainable whether the person providing the consent is indeed the person participating in the research. With Internet-based data collection this is not possible, as there is no visual reference (Pittenger, 2003). Additionally, it is not possible to determine that the person providing the consent meets the inclusion or exclusion criteria, as may be specified by the researcher. Thus, the issue of consent for Internet-based data collection includes issues of the respondent's personal integrity. Commonly, the consent to participate in Internet surveys takes the form of either choosing a box on the screen and pressing a button or choosing the "agree" option. Some IRBs may not consider this to be true informed consent, viewing it simply as the respondent's acknowledgement of reading the page. Since verifying this is next to impossible, some version of a "waiver of consent" becomes appropriate before conducting Internet-based data collection. This is especially relevant considering the possibility of respondents being minors without parental consent. Seeking and securing waivers from IRB for both parental and individual consent has been our approach to avoid subsequent disputes regarding consent, acknowledgement, and participation by minors.

One of the usual conditions of informed consent is that withdrawal from participation or refusal to participate cannot invalidate incentives. In Internet surveys with incentives provided this means that in the event of refusal to participate or early exit from the survey, the participant must be routed to the page meant for debriefing and incentive enrollment. Clearly, this is not perfect as the participant may simply close their Web browser to exit, rather than choose a button marked "exit survey." There is no simple and effective way to ensure that this does not happen and participants always have access to the incentives they are entitled to.

Other considerations that must be weighed are issues of *burden* and *beneficence*. Does using the Internet constitute an undue burden on a specific population, for example, computer illiterate individuals? There is no easy answer to this and it may in part depend on the subject matter being researched. Similarly, if participants receive benefits from being involved in the research, are these benefits available to non-computer users? These are difficult questions that each IRB would view differently; however, the best answer is that it depends on the research being conducted and the population being targeted for data collection. Our practice has been to anticipate these issues and, when applicable, justify the decisions in the design of the survey. Open communication with the IRB representative has helped us avoid unforeseen issues, thus leading to faster, more efficient approval processes.

4. Recruitment of Respondents

The Internet appears to be a mechanism to access the most representative participant pool in the world. Because of this, consumer researchers and marketing firms have created dedicated websites and electronic mailing lists designed to send out surveys to the willing public (e.g., NPD Online Research). However, it may not be correct to assume that recruitment of respondents in a virtual setting must be easy. We have utilized a variety of recruitment techniques and learned that, (a) different recruitment procedures can have different effects on the resulting sample and (b) the right recruitment procedure, with some luck, can yield interestingly large samples for the study.

Issues of recruitment have been widely discussed in the context of survey research (Cochran, 1977; McCready, 1996; Rosnow & Rosenthal, 2005; Sudman, 1983). Some of the recruitment methods are discipline-specific while others are more general. Most of these methods can be applied in an Internet-based project with simple alterations (Andrews, Nonnecke, & Preece, 2003; Hewson, Laurent, & Vogel, 1996; Koo & Skinner, 2005; Schillewaert & Meulemeester, 2005). For example, psychologists often utilize student pools from psychology classes--a convenience sample, while sociologists are usually more purposive in trying to sample groups meeting certain criteria (e.g., low-income minorities). Researchers using the Internet can recruit these same groups by either mass e-mailing the survey to the target group or sending out the survey Web site link to community leaders or organizations that interact with the target group.

If an electronic survey is being used simply to speed up data entry and analysis, the common method involving a group of participants meeting at a specified location and time can be used, with the provision of computers at the desired location. In this case, the recruitment procedure would be based on the accessibility of the population being sampled. Of course, the benefit of speedy data entry needs to be weighed against the risks associated with technology and those involved in data preparation processes (see Sections 5 and 6).

Despite the potential participant pool of hundreds of millions, the actual number of respondents in an Internet survey can be quite low (Zhang, 2000). In fact, response rates can be dismal enough to make the time-honored mail-in surveys seem more attractive. Using four of our Internet surveys as a basis, we have presented a discussion of the

recruitment techniques which worked for us and those which did not. Our experience indicates the prudence in following multiple recruitment strategies in any project. Moreover, strategies that worked before the Internet generally also work with the Internet.

In a project concerning health behaviors and activity, designed to survey college students, all 25,000 students on a college campus were e-mailed the Web link to the 15-page survey containing several validated and time-tested scales along with an explanation of the study and the opportunity to win prizes. A second e-mail was sent out two weeks later with a reminder and the link. One month after the original recruitment e-mail we had only 509 respondents (i.e., 2 per cent response rate). The inclusion of paper reminders placed in dormitory mailboxes increased participation within freshmen to 5 per cent, which was about 2 per cent prior to this.

A second project of ours with severe recruitment woes involved an attempt to get a community sample of driving behaviors within six cities in three states. The original recruitment procedure involved placing 600 paper leaflets or flyers per community (N=3,600) on vehicles parked in public parking lots during business hours. The flyers contained information about the study employing several persuasion tactics, the link to the Internet survey, and the contact information for the researchers--should a potential respondent have any questions or need help accessing the survey. Because of an inability to give reminders and the need for the respondent to manually enter the Web address, we planned on a 90 per cent non-response rate in order to get 60 participants per community (n=360). After 1,200 flyers distributed in two cities and one month of waiting, five respondents had attempted the online survey with only two finishing it in entirety. Interestingly enough, two respondents had accessed the survey the day before recruitment leaflets were sent out, which indicates that perhaps the IRB members were checking on the link and survey materials.

Not discouraged by a 0.5 per cent response rate, we adopted a *snowball sampling* technique in which we sent the survey out to friends, family, and colleagues. This recruitment e-mail contained study information, the link to the survey, and instructions to forward the e-mail to friends, family, and colleagues. Using this approach, the 60 initial e-mails yielded three times as many responses (189 to be precise) within the first month. Follow-up information seems to indicate that the snowballing process stopped at the third or fourth iteration. While this technique did yield higher response rates, it did not allow for community-specific analyses to be conducted because the e-mail contacts were distributed in other cities. It did however provide a broad sample with several professions and ages being represented. As an interesting aside, the e-mail sent out by one of the authors reached the other author, at the fourth iteration of snowballing, through routes that neither could have foreseen.

A team of researchers (including one of us) interested in individual attitudes related to the loss of local wildlife also utilized the electronic method to collect data. These researchers focused on college students for their sample and recruited participants by going into a diverse range of courses and verbally recruiting students by providing them with the

Internet link on an overhead projector. Interestingly, some course instructors offered extra credit for participation while others did not. For courses providing extra credit, more than 90 per cent of the students responded. The response rate was only 10 per cent where this incentive did not exist. This drastic difference based on extra credit was found to hold irrespective of the class size.

In another study we utilized a participant pool from an Introductory Psychology course. The participants received research credit for their participation that counted towards a course requirement. As it is to be expected, recruitment turned out to be a virtual non-issue. We just posted the study on the sign-up page and then e-mailed the link to those who signed up. Sections 5 and 6 below, focusing on technological and data preparation issues discuss this project and other similar projects which use the electronic method to reduce data entry time and labor.

Recruitment methods such as community sampling, telephone surveys, and mail-in surveys, widely used in different fields of research (Dillman, 1978), have also proved their merit in our Internet surveys. An incentive to participate is not essential but definitely helps and that has been known for some time (Brennan, 1992). In our projects, offering guaranteed benefits yielded greater than 90 per cent response rates. Surveys offering the possibility of some benefit, but no guarantee, had much lower response rates but were better than those without the possibility of such benefit. Reminders have also been shown to improve response rates in manual surveys (Nederhof, 1988; Sheehan & McMillan, 1999). Reminders doubled responses among college freshmen in our health survey even though the resulting response rate was not sufficiently high. The driving behaviors project had no reminders or incentive for the first round of data collection and was a total failure. However, we overcame the lack of incentive and our inability to offer reminders by utilizing snowball sampling that originated with people motivated to help-- our friends, family, and colleagues.

5. Technical Snags

Using the Internet to collect data is convenient and can greatly extend sample representativeness; however, the use of Internet is not without some risk. During the doctoral research of one of the authors, data were being collected using a mobile computer laboratory with an array of laptop computers, so as to avoid the time-consuming data entry process. Participants arrived every hour, completed the questionnaire online and left. Shortly into one of the sessions, the electricity supply to the building went out. Fortunately, the laptop batteries were fully charged and so no data were lost, and data collection continued. With desktop computers without uninterrupted power supply (UPS), the data entered till power-failure would have been lost and data collection would have to discontinue until power gets restored. Even with laptops this could have resulted in major inconvenience had the batteries not been charged or had the server been located in the building where the power supply was disrupted. After this experience the researcher printed out research packets to have on hand for future emergencies.

During the same project, the wireless Internet connection was lost for a period of time. This resulted in incomplete data from 18 respondents and created delays for the next session of data collection. A solution that was used in another Internet project conducted by the authors was to have a disc with the survey materials on it and have the respondents record their answers directly onto a Word document, which could later be transferred. To use this option, it is necessary to save each respondent's responses into a separate file for later retrieval, which requires enough disk space and the required level of access to save files.

In another research study conducted by one of the authors in a computer laboratory, all the computers contracted a virus. This was rather unfortunate, resulting in incomplete data from 14 respondents and lost data from 35 respondents. Considering that the sample size was 150, this resulted in approximately one-third of the sample being lost. Amendments for more participants had to be sent to the IRB since one of the experimental conditions was severely compromised by sheer luck of random assignment. Additionally, those 14 participants who were completing the study at the time had their university Internet accounts temporarily deactivated for using an infected computer. Prior to starting data collection each computer had been scanned for viruses and had antivirus updates installed. The virus came from another computer laboratory using the same server and infected the entire university network. Apart from keeping current on antivirus updates and timely virus scans, backing up the data more frequently during data collection could minimize virus-induced losses of already collected data. The paper-and-pencil back-ups will prevent losing participants who are present during the computer infection.

Another technology issue, especially in a laboratory setting, relates to the hardware devices used. In one of the studies mentioned above (i.e., the one with power-failure), the respondents were required to navigate the survey Web site using a touchpad. This resulted in delays and some confusion because the respondents were more used to a mouse, rather than a touchpad. Similarly, the type of screen and keyboard used can also make a difference. Specific screen sizes may be more appropriate for specific groups. Small screen size might be a disadvantage for groups with vision impairment. Similarly, perhaps a touch sensitive screen would be better than a keyboard while working with younger children.

In situations where multiple users may use the same computer to complete the study, it is necessary to determine if the survey software enters the data as new data or if, recognizing the same IP address, records over the previous data. This is not only a concern in laboratory settings; some hostel or dormitory rooms may have a single computer for multiple users. Even in the private home different family members may respond from the same computer. Another software issue is how it handles a respondent who exits the survey or closes the Web browser without completing the survey, whether accidentally or otherwise. Are they allowed to pick up at the point they exited, or do they need to start over? One study the authors were involved in did not allow the respondents to start where they left off. This resulted in numerous partially duplicated data points. For example, one would answer the first third of the survey and then accidentally exit, only to

discover that one needed to start at the beginning to take the survey. This would result in the first third of the survey being duplicated, requiring increased time in data cleaning later. Perhaps, this also resulted in frustration and withdrawal from the study, indicated by the fact that after data cleaning to eliminate duplicate entries, approximately 7 per cent of the data sets were incomplete.

When using flyers to recruit respondents, the Web address of the survey can cause a practical difficulty. Since IRBs tend to require data encryption, this necessitates the use of secure Web sites. Secure Web sites are designated with "https" in their address (rather than the usual "http"). This can lead to the respondents not typing the address correctly and consequently being unable to locate the survey. In a laboratory setting, one of the authors discovered that about 13 per cent of the respondents typed the Web address incorrectly. Specifically, they were all making the same error mentioned above. Even when told to be sure to type "https" and emphasizing the letter 's' the error rate was approximately 4 per cent. This tendency may be even more pronounced when using paper flyers or windshield leaflets for recruitment and possibly contributed to the dismal 0.5 per cent response rate encountered in the driving behaviors study.

The authors, jointly or individually, have been involved in over ten Internet-based surveys. Not a single one of those surveys has avoided technical or recruitment problems. Keeping back-up plans ready seems to be the major lesson from these experiences.

6. Data Preparation Issues

The single most appealing advantage of the electronic method of data collection is the elimination of the tedious data entry process. With the electronic method the data are entered into a database at the same time as the respondent completes the survey. If a researcher plans on collecting large amounts of data or having a large sample size, electronic data collection can be invaluable. It is a solution in itself when facing mountains of data and weeks worth of data entry. An additional advantage is that typing errors by the researcher are avoided. The data file is an exact replica of the responses received. However, electronic data files can easily lead to other types of error.

Electronic data files almost always need to be transformed, merged, and/or reformatted before use. Most available electronic formats separate the survey into sections and the data are provided in separate files for each section. These must be merged together so that analyses can be performed. Additionally, some programs that help facilitate creating e-surveys use their own coding schemes, which are not what the researcher might use. For example, 1-7 Likert scales may be recorded as 0-6 scales by the computer. Also, many established subscales have specific scoring criteria. Because of this, simple transformations are usually performed on the data. Also, when the data are downloaded into a database program, some programs default everything to string format, even if the data were meant to be numeric. As a result, another reformatting of the data becomes necessary. None of these issues is hard to correct. However, the more steps we add to the process, the more likely are we to make a mistake.

7. Conclusion

Data collection over the Internet has many potential benefits. Unfortunately, it also has many potential problems. Properly used, Internet-based data collection can generate large samples, be a solution to funding problems, ease logistics, and eliminate data entry. However, problems can arise during any phase of the research. With careful planning, many issues can be avoided altogether. While not all inclusive, this paper presents many of the issues the authors have encountered while conducting Internet-based data collection.

Advantages of Internet-based research have allowed us to dream a little bigger and pursue projects and research questions we would never have considered. Who would want to collect data in six cities in three states without formal funding? The Internet and some "creative budgeting" allowed the two of us to put the finishing touch on a project that had been two years in the making but confined to the available student pool for data collection. However, we will not discard the paper-and-pencil format either. For some projects, the inclusion of electronic data collection is not only unnecessary but also impractical. It can add unnecessary costs, time commitments, and headaches when used for smaller samples that are easily available. Conducting Internet-based research remains a decision that the researcher must weigh carefully.

Acknowledgements

References

Andrews, D., Nonnecke, B., & Preece, J. (2003). Electronic survey methodology: A case study in reaching hard-to-involve Internet users. *International Journal of Human-Computer Interaction, 16*(2), 185-210.

Belmont report: Ethical principles and guidelines for the protection of human subjects of research (Vol. 2, DHEW No. 5, 78-0014). (1978). Washington, DC: US Government Printing Office.

Berrens, R. P., Bohara, A. K., Jenkins-Smith, H., Silva, C., & Weimer, D. L. (2003). The advent of Internet surveys for political research: A comparison of telephone and Internet samples. *Political Analysis, 11*(1), 1-22.

Berry, D. M. (2004). Internet research: Privacy, ethics, and alienation--An open source approach. *Internet Research, 14*(4), 323-332.

Brennan, M. (1992). The effect of monetary incentive on mail survey response rates: New data. *Journal of the Market Research Society, 34,* 173-177.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: Wiley.

Dillman, D. A. (1978). *Mail and telephone surveys.* Newbury Park, CA: Sage.

Hewson, C. M., Laurent, D., & Vogel, C. M. (1996). Proper methodologies for psychological and sociological studies conducted via the Internet. *Behavior Research Methods, Instruments, & Computers, 28*(2), 186-191.

Hong, T. (2006). The influence of structural and message features on web site credibility. *Journal of the American Society for Information Science and Technology, 57*(1), 114-127.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15,* 635-650.

Koo, M., & Skinner, H. (2005). Challenges of Internet recruitment: A case study with disappointing results, *Journal of Medical Internet Research, 7*(1), Article e6. Retrieved April 25, 2006, from http://www.jmir.org/2005/1/e6/

Manfreda, K. L., Batagelj, Z., & Vehovar, V. (2002). Design of web survey questionnaires: Three basic experiments. *Journal of Computer-Mediated Communication, 7*(3). Retrieved April 25, 2006, from http://jcmc.indiana.edu/vol7/issue3/vehovar.html

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York: Harper and Row.

McCready, W. C. (1996). Applying sampling procedures. In F. T. L. Leong & J.T. Austin (Eds), *The psychology research handbook: A guide for graduate students and research assistants* (pp. 98-110). Thousand Oaks, CA: Sage.

Naglieri, J. A., Drasgow, F., Schmidt, M., Handler, L., Prifitera, A., Margolis, A., et al. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist, 59*(3), 150-162.

Nederhof, A. J. (1988). Effects of a final telephone reminder and questionnaire cover design in mail surveys. *Social Science Research, 17,* 353-361.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues, 58*(1), 161-176.

Olson, J. M., & Cal, A. V. (1984). Source credibility, attitudes, and the recall of past behaviours. *European Journal of Social Psychology, 14,* 203-210.

Orlich, D. (1978). *Designing sensible surveys*. Pleasntville, NY: Redgrave.

Pittenger, D.J. (2003). Internet research: An opportunity to revisit classic ethical problems in behavioral research. *Ethics and Behavior, 13*(1), 45-60.

Rosnow, R. L., & Rosenthal, R. (2005). *Beginning behavioral research: A conceptual primer* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Sheehan, K. B., & McMillan, S. J. (1999). Response variation in e-mail surveys: An exploration. *Journal of Advertising Research, 39*(4)*,* 45-54.

Schillewaert, N., & Meulemeester, P. (2005). Comparing response distributions of offline and online data collection methods. *International Journal of Market Research, 47*(2), 163-178.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys.* New York: Academic Press.

Sethuraman, R., Kerin, R. A., & Cron, W. L. (2005). A field study comparing online and offline data collection methods for identifying product attribute preferences using conjoint analysis. *Journal of Business Research, 58,* 602-610.

Sudman, S. (1983). Applied sampling. In P. Rossi, J. Wright, & A. Anderson (Eds), *Handbook of survey research* (pp. 145-194). New York: Academic Press.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Waern, Y. (2001). Ethics in global Internet research (Report from the Department of Communication Studies, Linkoping University). Retrieved April 25, 2006, from http://www.cddc.vt.edu/aoir/ethics/public/YWaern-globalirethics.pdf

Zhang, Y. (2000). Using the Internet for survey research: A case study. *Journal of the American Society for Information Science, 51*(1), 57-68.