

## THEORY AND METHODS

## Collecting behavioural data using the world wide web: considerations for researchers

S D Rhodes, D A Bowie, K C Hergenrather

*J Epidemiol Community Health* 2003;57:68–73

See end of article for authors' affiliations

Correspondence to:  
Dr S D Rhodes, University  
of North Carolina School  
of Public Health,  
Department of Health  
Behavior and Health  
Education, Campus Box  
7440, Chapel Hill, NC  
27599-7440, USA;  
Scott\_Rhodes@unc.edu

Accepted for publication  
15 April 2002

**Objective:** To identify and describe advantages, challenges, and ethical considerations of web based behavioural data collection.

**Methods:** This discussion is based on the authors' experiences in survey development and study design, respondent recruitment, and internet research, and on the experiences of others as found in the literature.

**Results:** The advantages of using the world wide web to collect behavioural data include rapid access to numerous potential respondents and previously hidden populations, respondent openness and full participation, opportunities for student research, and reduced research costs. Challenges identified include issues related to sampling and sample representativeness, competition for the attention of respondents, and potential limitations resulting from the much cited "digital divide", literacy, and disability. Ethical considerations include anonymity and privacy, providing and substantiating informed consent, and potential risks of malfeasance.

**Conclusions:** Computer mediated communications, including electronic mail, the world wide web, and interactive programs will play an ever increasing part in the future of behavioural science research. Justifiable concerns regarding the use of the world wide web in research exist, but as access to, and use of, the internet becomes more widely and representatively distributed globally, the world wide web will become more applicable. In fact, the world wide web may be the only research tool able to reach some previously hidden population subgroups. Furthermore, many of the criticisms of online data collection are common to other survey research methodologies.

"Nothing is permanent but change."

Heraclitus, circa BC 500.

With up to 15 million people world wide accessing the internet each day at a rate that is increasing by an estimated 25% every three months,<sup>1</sup> and each person averaging nearly 10 hours per week on line,<sup>2</sup> new opportunities exist for researchers to conduct online research. To date, health researchers have harnessed this new technology using the world wide web to provide health information,<sup>3–8</sup> conduct health appraisals,<sup>9–12</sup> deliver interactive health related interventions,<sup>11–16</sup> and collect epidemiological data.<sup>7, 17–19</sup> Most recently, a growing number of researchers have begun to collect behavioural data via the world wide web.<sup>20–24</sup> This preliminary experimentation with behavioural data collection using the world wide web seems promising as nearly all data collection that once relied on paper and pencil can be completed electronically.

While advantages exist for online collection of behavioural data, researchers must be aware of the challenges and ethical considerations that online data collection generate. We outline the advantages and challenges associated with online data collection and discuss ethical considerations that must be tackled as online data collection is further developed in research applications.

## ADVANTAGES OF WEB BASED DATA COLLECTION

### Electronic dexterity

Nearly any survey instrument, including mail, self administered, and interviewer administered questionnaires, can be posted on the world wide web for pilot testing and administration. The process for web based data collection is comparatively simple. A questionnaire is translated into

HTML (hypertext markup language), the de facto language of the internet. With a point and click interface, respondents can complete a web based survey that is visually and functionally similar to traditional, written surveys. Form elements of web based surveys are similar to many self administered surveys including check boxes and numeric entry boxes as well as radio buttons, selection lists, and pull down menus that facilitate data entry and minimise error.

The electronic nature of the medium allows researchers to make adjustments to a survey as unforeseen problems to comprehension are discovered. Just as questions can be revised or removed, new questions or follow up questions can be added as new issues arise based on new information or preliminary findings. Online data collection can even document the length of time that a respondent took to complete a survey.

### The provision of numerous potential respondents

The internet provides nearly limitless numbers of potential study respondents across geographical and cultural boundaries.<sup>25–26</sup> In a study headed by the first author, the internet server that hosted the survey and the researchers were based in Birmingham, Alabama, USA, while collecting data from throughout North American, Eastern and Western Europe, Africa, Asia, and Australia.<sup>27</sup> Although the distance between the researcher and respondent can be inconsequential with mailed questionnaires, for example, extended time allotted for postal service, the costs associated with mailed surveys, and the typically low response rates can be prohibitive.<sup>25–30</sup>

### Access to previously hidden populations

Internet communities and chat rooms, newsgroups, electronic mailing lists, search engines, hypertext links from related web sites, and web rings, offer unique recruitment opportunities to

develop and test strategies to capture the attention of potential respondents and motivate their participation in a web based survey. Often these respondents are difficult to access using traditional methodologies. For example, one online survey used hypertext links from other web sites to collect quantitative psychological data on gay men's participation in "bareback sex,"<sup>31</sup> the term used to describe the psycho-social phenomenon among gay men of consciously choosing to have unprotected anal intercourse. Many of these men are HIV positive or may desire to become HIV positive. The researcher established links from a multitude of web sites that were designed to facilitate this type of sexual networking.

Using newsgroups for recruitment, another study collected data to understand the demographic, symptomatic, and predisposing characteristics of chronic prostatitis.<sup>32</sup> Newsgroups are topic specific and allow members to communicate in an electronic mail-like manner; members can share concerns and experiences with others who share a health status, a political stance, a profession, or some other characteristic or interest. Other behavioural survey studies have used a combination of recruitment strategies, including hypertext links and snowball approaches in which respondents distribute the uniform resource locator (URL) of an online questionnaire to their electronic mail correspondents and distribution lists.<sup>23 27 33</sup> Non-electronic techniques highlighting the URL and encouraging participation can be used as well to recruit respondents such as radio announcements, television stories, the distribution of business cards, and advertisements within print media.

These types of recruitment methodologies tend to be "active," with the researcher calling on potential respondents to participate in a survey. Keyword META-tags, a more "passive" recruitment approach, are another recruitment option. Briefly, META-tags are embedded keywords that index web sites on the world wide web.<sup>34</sup> As a person "surfing" the web searches a topic on various search engines (for example, Ask Jeeves, Lycos, Magellan, Microsoft Explorer, Netscape, etc), he or she uses key search words to locate information pertaining to a topic of interest. These words result in listing web sites that are META tagged or described with these invisible keywords. At one time, one of the most popular META-tags was "Pamela Anderson Lee."<sup>35</sup> By searching for "Pamela Anderson Lee," a person searching the web would get access to a variety of web sites that may or may not pertain to Pamela Anderson Lee. Web site operators, who had sites that did not pertain to Pamela Anderson Lee but wanted to capture the attention of the types of people who might search for such a web site, chose to META-tag their sites with the name "Pamela Anderson Lee."

Currently, the many search engines offer a day by day ranking of popular key search words that can be incorporated into a web site as META-tags depending on the type of respondents a researcher wants to attract. For example, researchers hoping to attract the attention of and collect data from people who order pharmaceuticals on line may META-tag their survey with specific drug brand names or descriptive words such as "drug," "prescriptions," and "vitamin."

Further research is needed to determine how different recruitment strategies affect sample composition. Although special recruitment efforts may be made to solicit particular types of respondents, researchers may have limited control over contamination to a recruitment protocol. For example, as respondents locate an online survey through the researcher's intended recruitment protocol, respondents may then distribute the web address through another strategy. For example, we initially contacted web masters of gay and bisexually oriented web sites to establish hypertext links to our data collection web sites. However, web masters quickly distributed the web sites to their friends, distribution lists, and newsgroups.<sup>23 27</sup> While this snowball approach did not seem to harm our efforts, the composition of a sample can be changed substan-

tially and change a study's findings. An example of this type of sampling contamination may occur as a researcher establishes a web site to collect sexual behaviour data. A well intentioned respondent could announce the survey and distribute the URL of the site via a newsgroup created for survivors of sexual abuse thus resulting in the recruitment of a high percentage of respondents who reported a history of sexual abuse. The researcher may have no indication that this has happened. Of course, surveys requiring passwords solve this problem, but recruitment is further complicated and participation rates are jeopardised.<sup>36</sup>

### Speed: from research questions to answers

The web serves as an expedient method to collect data. The time between planning a study and reporting findings is reduced because hundreds of respondents can access a data collection web site and submit their data at any given time. There is no testing site or appointment/outreach scheduling, and data entry is eliminated. Of course, creating a survey is required; however, inexpensive and user friendly software exist that can facilitate the process.

### Better data through reduced error

Online data collection is likely to yield more useable data than other data collection methodologies<sup>39</sup> by reducing error two ways. Firstly, error is reduced by the inclusion of explanatory material, prompts, and menus on the online survey. Respondents interact with a questionnaire in a structured format, minimising entry of erroneous or unacceptable data. With a paper and pencil survey, respondents can skip items or enter a word when a number is requested, for example. Multiple responses to questions requiring single answers can be prohibited, and completion of all questions can be required before a questionnaire can be successfully submitted and accepted. Complicated branches and skip patterns can be programmed into the survey, requiring less respondent attention to the survey format. Also, responses can be reviewed before acceptance of the data to eliminate answer inconsistencies within a respondent's survey.

Secondly, error resulting from variation of survey administration, interviewer interpretation, and data entry is prevented using online data collection. The respondents' interactions with the survey are standardised; thus interviewer bias is eliminated. Furthermore, the automatic data entry that occurs as a survey is completed ensures that errors from data entry are non-existent. A common gateway interface (CGI) script can be used to automatically compile data and export them into statistical software package such as SAS (Cary, NC) or SPSS<sup>37</sup> for data analysis.

### Sensitive topics and the reduction of bias

Web based data collection seems to be an effective format to ask respondents about sensitive and difficult to discuss topics.<sup>33 38</sup> Preliminary research suggests that people share information and experiences electronically that they might not disclose using traditional survey methodologies.<sup>39-43</sup> Thus, online surveying may reduce social desirability and yea-saying biases<sup>44</sup> and allow the researcher to collect accurate sexual behaviour data that tend to be highly sensitive. The reasons for increased self disclosure and uninhibited responses are unclear; however, respondents to online data collection may believe (correctly or wrongly) that their responses are more secure and anonymous.<sup>38 45-47</sup>

### Full participation

Online respondents, in fact, may participate more fully than in other survey methodologies.<sup>45</sup> They are likely to give feedback, offer help and support with survey distribution, and ask for a summary of the study findings. In our experiences with online data collection, respondent altruism seemed evident. Web

masters, who were solicited to establish links from their web sites to an online survey, frequently distributed the URL address of the survey by electronic mail to their friends. Furthermore, 65 of 628 respondents in one of our studies requested notification of the study's results by electronic mail, and another 28 electronically communicated qualitative context to their own personal risk behaviour and disease progression. Perhaps because taking part in such a study requires the participant to seek it out, or because respondents are more in control of the experience and their participation, there is a shift from being a subject of experimentation to being an active participant.

To further the respondents' participation and benefit, researchers using a web based methodology also can provide feedback to respondents in the form of summary statistics about aggregate results up to that particular point in time or after some temporal delay. Furthermore, online survey respondents can participate in a survey at their leisure and individual paces. Respondents are not limited to participating during specific hours or in specific settings and can participate privately at their convenience. Moreover, online respondents may feel more able to terminate participation before completion without the perceived social pressure that they may feel in face to face encounters.<sup>48</sup>

### An application for students

Web based research is a simple yet effective method for both experienced researchers as well as students to collect and analyse data as a learning exercise.<sup>49</sup> The first and third authors collected hepatitis B and C virus infection behavioural data from MSM using web based questionnaires while they were doctoral students. No costs were associated with the research as the university provided the server that hosted the questionnaires and collected respondent data. This research has led to two publications<sup>23 27</sup> and funding for subsequent research.

### Cost: the bottom line

The decreased need for data collection staff and the automatic electronic data entry that occurs as respondents complete an online questionnaire may save between 20% to 80% of total data collection costs associated with collecting and entering the paper and pencil questionnaire into a database.<sup>26 41 44 50-52</sup> With web based data collection there is no need to train interviewers in survey administration, scoring, and data entry. No costs accrue from space required from questionnaire administration, paper and printing, postage, or paper storage.

## THE CHALLENGES OF WEB BASED DATA COLLECTION

### Sampling issues

When collecting data on line, response rates are incalculable because of the unknown number of potential respondents who received recruitment materials or examined a data collection web site but chose not to participate.<sup>53</sup> Web counters allow researchers to document the number of visits that a web site received but this technique does not count non-respondents who received the URL via active or passive recruitment strategies. As has been noted, online data collection does not ensure true random sampling.<sup>22 54</sup> However, the degree of fit between a sample and the target population about which generalisations can be made is a common challenge in many studies<sup>6 33 33 55</sup>; in fact, nearly all studies of sexual behaviour among MSM, for example, are based on self selected samples<sup>56 57</sup> or clinical populations.<sup>22 58</sup>

Furthermore, although the internet uses a self administered format that may minimise response bias, these results remain based on self reported data with their potential limitations.<sup>59</sup> Techniques found to increase validity of self

reported behaviour when applied to the paper and pencil questionnaire<sup>60 61</sup> can be adapted and applied easily to web based data collection.

Although multiple submissions have not proved to be a serious problem in online data collection to date, the potential for multiple submissions is clearly present. A respondent can complete a survey, click the submit button, read the appreciation and debriefing page, and press the "back" key to complete the survey multiple times, resulting in another set of data from the same respondent. There are several confirmed techniques that can be used to limit this from occurring initially or affecting the analyses subsequently. Firstly, the first page of the survey that explains the purpose can specifically ask that each person complete the survey only once.<sup>36</sup> Secondly, a question embedded within the survey can ask whether the respondent has completed the survey previously. Thirdly, identifying information can be examined to identify possible multiple or duplicate responses. For instance, in our confidential surveys of sexual risk behaviour, we have used respondent zip code, sex, and date of birth (month/day/year) to identify possibly duplicate responses.<sup>23 27</sup>

### Competition

With the rapid growth rate of internet web sites, reaching the target population may be increasingly difficult as competition for the attention of potential respondents increases. Current estimates suggest that between two and three billion web pages exist.<sup>62 63</sup> Relying on internet communities, newsgroups, electronic mailing lists, search engines, hypertext links from related sites, and web rings to disseminate a questionnaire or web site address (URL) pose challenges to data collection methodology, analysis, and interpretation that must be explored repeatedly.

### The digital divide

Research has uncovered educational, economic, racial, and gender disparities among those who have access to, or use, the web.<sup>25 54 64-68</sup> In the US, this "digital divide," which suggests that younger, more educated, higher income, white men have greater access to the web, seems to be changing as the numbers of people on line increase.<sup>69</sup> In 1996, in the US specifically, 30% of adults had computers at home and 27% logged on line; today over 50% of adults have computers in their home and over 80% log onto the internet.<sup>69</sup> In fact, over 700 new households join the internet every hour.<sup>50</sup> These figures do not include access to the internet through schools, libraries, and other public institutions. Furthermore, data suggest that population subgroups within the US and European countries may not follow aggregate trends of internet access and use. Studies of gay and bisexual populations, for example, suggest that web users from these populations include a higher proportion of lower educated, lower income, unemployed, and disabled persons.<sup>22 58 70</sup> This difference may reflect the adoption of the world wide web by some marginalised groups as a safe place to interact without fear of negative social consequences.<sup>70</sup>

Recent studies also have found that women are more likely to participate in online research despite the conventional wisdom that the web is male dominated.<sup>48 49 69</sup> Thus, the "digital divide" assumption inaccurately may characterise web access and use among subgroups within the US population, and as has been suggested, the "digital divide" may not be as relevant as once assumed.<sup>69</sup> Of course, further research will be necessary to identify and understand internet access and use patterns within cultures and populations and within subgroups.

Moreover, targeted recruitment may reach respondents from populations that were previously inaccessible through traditional survey methods, including, but not limited to, non-deviant adult recreational drug users,<sup>71 72</sup> non-dependent



### Key points

- As access to and use of the internet increase, new opportunities exist for behavioural researchers to collect data on line.
- Advantages to online research include: reaching large numbers of potential respondents and respondents from hidden populations; expediency; the reduction of error and bias; full participation; and lower costs.
- Challenges include: sample representativeness; competition for attention; assumptions about the "digital divide;" literacy, and disability; and, the internet's limited international reach.
- Ethical considerations include: anonymity and privacy; ensuring informed consent; and, malfeasance.
- The use of the world wide web in research will continue to pose a variety complex challenges for researchers, the general public, especially potential respondents, and ethics committees and institutional review boards that warrant discussion.

problem drinkers,<sup>73</sup> men who are or have been abused physically by their heterosexual female sexual partners,<sup>74</sup> paedophiles,<sup>75</sup> and sellers of illegal drugs.<sup>76</sup> The potential use of the internet to reach hidden populations is vast but the internet still may not reach some hidden groups that traditional methodologies cannot access.

### Literacy and disability

Currently, over 32% of all web pages are in languages other than English, including Catalan, Czech, Dutch, Greek, Polish, Romanian, Slovenian, Hebrew, Malay, and Thai among others.<sup>77</sup> Although the percentage of web pages in languages other than English continues to increase rapidly, local literacy rates limit access to the world wide web.<sup>78</sup> In the US alone, nearly 100 million adults lack the basic reading skills to function successfully in society,<sup>79</sup> and thus it is assumed that the internet, which is primarily text based, is out of their reach. However, rapid advances in computer technology and telecommunications continue to provide new mechanisms for access to the internet and world wide web without regard to literacy, vision, or physical disabilities. Examples of these technologies include onscreen keyboards with head pointers, voice command systems, and software providing graphic representations of concepts and multi-sensory interactions. In fact, the US government has mandated by that federal agencies move towards making programs offered on web sites accessible to all people with disabilities.<sup>80</sup> The US Intragency Committee on Disabilities Research is exploring the development and transfer of assistive technology and universal design to ensure equal access to resources on the internet.<sup>69</sup> Thus, these developments will help ensure that some of the world wide web is accessible to people without regard to disabling conditions. To date, steps taken within the US at the federal level in terms of assistive technology have not been undertaken in other countries.

### Limited international scope

Although the physical locations of a researcher and a respondent are unimportant, as previously noted, access to the internet is limited currently for many people within the US, Western Europe, and developing countries. However, this limitation may be changing; of the roughly 100 million new internet users, who logged onto the internet in 2000, three quarters were located outside the US with the majority located in the United Kingdom and Germany. Examples of countries with growing internet use include China, Italy, and South Korea; each of these countries experienced growth rates of numbers of internet users of over 140% between 1999 and 2000, and Brazil, France, and Germany doubled the US growth rate.<sup>69</sup> In

more than 50 countries world wide, internet use has increased over 100% in one year.<sup>81</sup> Thus, with time, online global data collection may become a more realistic possibility.

Obviously the necessity for an adequate telecommunications infrastructure is paramount. Researchers collecting international data must recognise that the sometimes limited access to a reliable telecommunications infrastructure may affect response and data collection beyond limited access to computer technology, especially within developing countries.

### ETHICAL CONSIDERATIONS

While the potentials for online research seem nearly limitless, the potentials for intentional and unintentional misuse are also broad. Because research using the world wide web is a new domain for research, ethical issues have not been well resolved within the research community and among oversight boards including ethics committees and institutional review boards.<sup>82-83</sup> Current concerns include whether researchers truly can promise anonymity and confidentiality, and what constitutes informed consent.

### Anonymity and privacy

Encryption, used to protect the transmission of confidential information, such as credit card numbers from unauthorised third parties, can protect responses to an online questionnaire but cannot protect the respondent's IP address. A computer "hacker" may be able to determine a respondent's IP address and the site visited, perhaps a site associated with socially undesirable or illegal behaviour for example, and use this information maliciously.<sup>18-53</sup> This information can be more or less problematic depending on the type of survey but in all cases is a breach of privacy. Researchers must educate potential respondents about their inability to ensure true, 100% anonymity or confidentiality and provide alternate mechanisms. A link to a free service that guarantees anonymous access to the internet, such as <http://anonymizer.com>, may solve this problem. URL addresses of surveys that are distributed via electronic mail or some other recruitment strategy may provide this information without an initial web site visit but the use of hyperlinks would require an initial visit to a data collection web site. Understanding the changing technology and the adaptations necessary to ensure respondent anonymity, confidentiality, and security are challenges for researchers who choose to collect data via the world wide web and for ethics committees and institutional review boards who must approve and oversee these studies.

### Informed consent

When using the world wide web to collect data, a standard of what constitutes informed consent has not been well established within the research community. Is it necessary to display a consent form to potential respondents and what respondent designation of consent is sufficient? Respondents can be asked to type their names, create a code of their birth date and other identifying information, or simply click a check box indicating their consent. Knowing whether the respondent truly understands the research and providing the respondent with opportunities for clarification are limited, if not impossible on line. Furthermore, with such distance between the researcher or data collection staff and respondent, researchers cannot verify basic demographics of a respondent, perhaps most importantly, the respondent's age. Potential respondents could be minors, which in other research settings would require consent of parent or legal guardian. For example, US federal regulations require that an operator of a commercial web site or online service directed towards children, or any operator who has actual knowledge that he or she is collecting or maintaining personal information from a child must, in most cases, obtain verifiable parental consent before any collection and use of information

from children.<sup>64</sup> Permissible methods of obtaining verifiable parental consent include: providing a consent form that must be signed by a parent or legal guardian and returned by either postal mail or facsimile; requiring a parent to use a credit card in connection with an online transaction; having a parent call a toll free telephone number staffed by trained personnel to verify consent; using a digital certificate that uses public key technology; and using electronic mail accompanied by a personal identification number (PID) or password. The issue of consent is especially important as youth access to the world wide web increases dramatically every year in schools, libraries, community centres, and private homes.<sup>62</sup>

To complicate matters further, new US federal health privacy regulations, promulgated by the US Department of Health and Human Services (HHS), establish the conditions under which health information may be used or disclosed for research purposes.<sup>65</sup> These regulations are in addition to the informed consent procedures currently required by HHS and the US Food and Drug Administration for research involving human subjects<sup>66, 67</sup> and may affect online data collection methodologies and the reporting of findings. Most parties covered by the regulations must be in compliance by April 2003.<sup>65</sup>

### Malfeasance or do no harm

In traditional survey methodologies, specifically face to face settings, the researcher or the study staff monitors respondent reactions and rectifies or provides resources during negative reactions. With mail or telephone assisted surveying, the researcher can know the locations of the respondents and can provide targeted, local resources should a respondent need support as a result of the research. Counselling service or other types of referrals can be provided as needs are identified. With online surveying, however, the opportunities to identify needs and provide support are limited. Lists of resources can be provided after the survey is submitted but these resources do not reach those respondents who discontinue participation in mid-survey as a result of emotional distress. Moreover, these resources must be national or international given the global catchment area of the world wide web, and depending on the nature of a survey, national or international resources may be non-existent.

### CONCLUSION

With an estimated 450 million people world wide having access to the internet by the end of 2001,<sup>68</sup> unique and exciting opportunities currently exist and will be developed for researchers to collect valuable data using the internet. Numerous respondents, many of whom previously were hidden from researchers, are available electronically via the world wide web. Understanding the changing technology and the subsequent necessary adaptations, and the diffusion of access and use of the world wide web will continue to pose a variety of complex challenges for researchers, the general public, especially potential respondents, and ethics committees and institutional review boards.

### ACKNOWLEDGEMENTS

Manuscript preparation was supported in part by the Community Health Scholars. Program funded by the W K Kellogg Foundation (Scott D Rhodes).

### Disclaimer

The views expressed in this article are those of the authors and do not necessarily represent the views of the Federal Trade Commission or any individual Commissioner.

### Authors' affiliations

**S D Rhodes**, The Department of Health Behavior and Health Education, University of North Carolina School of Public Health, Chapel Hill, NC, USA

**D A Bowie**, United States Federal Trade Commission, Washington, DC, USA

**K C Hergenrather**, Department of Counseling/Human and Organizational Studies, George Washington University, Washington, DC, USA

### REFERENCES

- 1 **Science Panel on Interactive Communication and Health**. *Wired for health and well-being: The emergence of interactive health communications*. In: Eng TR, Gustafson DH, eds. Washington, DC: US Department of Health and Human Services, US Government Printing Office, 1999.
- 2 **Cooper A**, Delmonico DL, Burg R. Cybersex users, abusers, and compulsives: New findings and implication. *Sexual Addiction Compulsivity* 2000;**7**:5-29.
- 3 **Baujard O**, Baujard V, Aurel S, et al. Trends in medical information retrieval on Internet. *Comput Biol Med* 1998;**28**:589-601.
- 4 **Kim HS**, Kim E, Kim JW. Development of a breast self-examination program for the Internet: Health information for Korean women. *Cancer Nurs* 2001;**24**:156-61.
- 5 **Hardin PC**, Reis J. Interactive multimedia software design: concepts, process, and evaluation. *Health Educ Behav* 1997;**24**:35-53.
- 6 **Pealer LN**, Weiler RM. Research notes. Web-based health survey research: a primer. *Am J Health Behav* 2000;**24**:69-72.
- 7 **Soetikno RM**, Provenzale D, Lenert LA. Studying ulcerative colitis over the world wide web. *Am J Gastroenterol* 1997;**92**:457-60.
- 8 **Soetikno RM**, Lenert LA, Provenzale D. Research and the internet. *Am J Gastroenterol* 1998;**93**:485.
- 9 **Coiera E**. Information economics and the Internet. *J Am Med Inform Assoc* 2000;**7**:215-21.
- 10 **Fotheringham MJ**, Wonnacott RL, Owen N. Computer use and physical inactivity in young adults: public health perils and potentials of new information technologies. *Ann Behav Med* 2000;**22**:269-75.
- 11 **Paperny DM**. Computerized health assessment and education for adolescent HIV and STD prevention in health care settings and schools. *Health Educ Behav* 1997;**24**:54-70.
- 12 **Rhodes F**, Fishbein M, Reis J. Using behavioral theory in computer-based health promotion and appraisal. *Health Educ Behav* 1997;**24**:20-34.
- 13 **Bennett SJ**, Hays LM, Embree JL, et al. Heart messages: a tailored message intervention for improving heart failure outcomes. *J Cardiovascular Nurs* 2000;**14**:94-105.
- 14 **Johnson KB**, Ravert RD, Everton A. Hopkins Teen Central: assessment of an internet-based support system for children with cystic fibrosis. *Pediatrics* 2001;**107**:E24.
- 15 **Marcus BH**, Nigg CR, Riebe D, et al. Interactive communication strategies: implications for population-based physical-activity promotion. *Am J Prev Med* 2000;**19**:121-6.
- 16 **Tate DF**, Wing RR, Winett RA. Using internet technology to deliver a behavioral weight loss program. *JAMA* 2001;**285**:1172-7.
- 17 **Klausner JD**, Wolf W, Fischer-Ponce L, et al. Tracing a syphilis outbreak through cyberspace. *JAMA* 2000;**284**:447-9.
- 18 **Kushi LH**, Finnegan J, Martinson B, et al. Epidemiology and the internet. *Epidemiology* 1997;**8**:689-90.
- 19 **Houston JD**, Fiore DC. Online medical surveys: Using the internet as a research tool. *MD Comput* 1998;**15**:116-20.
- 20 **Baehring TU**, Schulze H, Bornstein SR, et al. Using the world wide web: a new approach to risk identification of diabetes mellitus. *Int J Med Inform* 1997;**46**:31-9.
- 21 **Musch J**, Reips UD. A brief history of web experimenting. In: Birnbaum MH, ed. *Psychological experiments on the internet*. San Diego: Academic Press, 2000:61-87.
- 22 **Ross MW**, Tikkanen R, Mansson SA. Differences between internet samples and conventional samples of men who have sex with men: implications for research and HIV interventions. *Soc Sci Med* 2000;**51**:749-58.
- 23 **Rhodes SD**, DiClemente RJ, Yee UJ, et al. Correlates of hepatitis B vaccination in a high-risk population: an internet sample. *Am J Med* 2001;**110**:628-32.
- 24 **Lamb M**. Cybersex: research notes on the characteristics of the visitors to online chat rooms. *Deviant Behavior* 1998;**19**:121-35.
- 25 **Eter JF**, Perneger TV. A comparison of cigarette smokers recruited through the Internet or by mail. *Int J Epidemiol* 2001;**30**:521-5.
- 26 **Houston TK**, Cooper LA, Vu HT, et al. Screening the public for depression through the Internet. *Psychiatr Serv* 2001;**52**:362-7.
- 27 **Rhodes SD**, DiClemente RJ, Yee UJ, et al. Factors associated with testing for hepatitis C in an Internet-recruited sample of men who have sex with men. *Sex Transm Dis* 2001;**28**:515-20.
- 28 **Cassell MM**, Jackson C, Cheuvront B. Health communication on the internet: an effective channel for health behavior change? *J Health Commun* 1998;**3**:71-9.
- 29 **Clark TJ**, Khan KS, Gupta JK. Provision of pen along with questionnaire does not increase the response rate to a postal survey: a randomised controlled trial. *J Epidemiol Community Health* 2001;**55**:595-6.
- 30 **McAvoy BR**, Kaner EF. General practice postal surveys: a questionnaire too far? *BMJ* 1996;**313**:732-3.
- 31 **Cason G**. Barebacking: politics or recklessness? *WebMDHealth* 31 May 2000. Available at: <http://my.webmd.com/content/article/1700.50606>. (Accessed 5 September 2001.)
- 32 **Alexander RB**, Trissel D. Chronic prostatitis: results of an internet survey. *Urology* 1996;**48**:568-74.

- 33 **Bailey RD**, Foote WE, Throckmorton B. Human sexual behavior: a comparison of college and internet surveys. In: Birnbaum MH, ed. *Psychological experiments on the internet*. San Diego, CA: Academic Press, 2000:141–68.
- 34 **Gaffney BJ**. Metatags and keying: Is their use trademark infringement? *Topics in Cyberspace Law* Fall 2000.
- 35 **Weber T**. Net of fame: Who rules the web? Pamela Anderson Lee, The b-movie actress - Linked to naughty video, she's a massive draw regrettably, at times that 'GNP' of \$77 million. *Wall Street Journal* 14 Apr 1999:A1.
- 36 **Birnbaum MH**. *Introduction to behavioral research on the internet*. Upper Saddle River, NJ: Prentice Hall, 2000.
- 37 **SPSS**. *SPSS user's guide, version 10.0 for Windows*. New York: McGraw Hill, 2000.
- 38 **Davis RN**. Web-based administration of a personality questionnaire: comparison with traditional methods. *Behav Res Methods Instrum Comput* 1999;**31**:572–7.
- 39 **Joinson A**. Social desirability, anonymity, and internet-based questionnaires. *Behav Res Methods Instrum Comput* 1999;**31**:433–8.
- 40 **Joinson AN**. Self-disclosure in computer-mediated communication: the role of self-awareness and visual anonymity. *Eur J Soc Psychol* 2001;**31**:177–92.
- 41 **Kiesler S**, Sproull LS. Response effects in the electronic survey. *Public Opinion Quarterly* 1986;**50**:402–13.
- 42 **McKenna KY**, Bargh JA. Coming out in the age of the internet: identity "demarginalization" through virtual group participation. *J Pers Soc Psychol* 1998;**75**:681–94.
- 43 **Babbie E**. *The practice of social research*. 8th edn. Belmont, CA: Wadsworth Publishing, 1998.
- 44 **Streiner DL**, Norman GR. *Health measurement scales: a practical guide to their development and use*. 2nd edn. Oxford: Oxford University Press, 1995.
- 45 **Buchanan T**. Potential of the internet for personality research. In: Birnbaum MH, ed. *Psychological experiments on the internet*. San Diego, CA: Academic Press, 2000:121–40.
- 46 **Malone RE**. Research, the internet, and the way things are. *Health Educ Behav* 2000;**27**:695–7.
- 47 **McKenna KY**, Bargh JA. Plan 9 from cyberspace: the implications of the internet for personality and social psychology. *Journal of Personality and Social Psychology* 2000;**4**:57–75.
- 48 **Reips UD**. The web experiment method: advantages, disadvantages, and solutions. In: Birnbaum MH, ed. *Psychological experiments on the internet*. San Diego, CA: Academic Press, 2000:89–117.
- 49 **Baron J**, Siepmann M. Techniques for creating and using web questionnaires for research and teaching. In: Birnbaum MH, ed. *Psychological experiments on the internet*. San Diego, CA: Academic Press, 2000:235–5.
- 50 **Budman SH**. Behavioral health care dot-com and beyond: computer-mediated communications in mental health and substance abuse treatment. *Am Psychol* 2000;**55**:1290–300.
- 51 **Fawcett J**, Buhle EL. Using the internet for data collection: an innovative electronic strategy. *Comput Nurs* 1995;**13**:273–9.
- 52 **Yokley JM**, Coleman DJ, Yates BT. Cost effectiveness of three child mental health assessment methods: Computer-assisted assessment is effective and inexpensive. *Journal of Mental Health Administration* 1990;**17**:99–107.
- 53 **Hilsden RJ**, Scott CM, Verhoef MJ. Research and the Internet. *Am J Gastroenterol* 1998;**93**:484–5.
- 54 **Binik YM**, Mah K, Kiesler S. Ethical issues in conducting sex research on the internet. *J Sex Res* 1999;**36**:82–90.
- 55 **Schmidt WC**. World-wide web survey research: benefits, potential problems, and solutions. *Behav Res Methods Instrum Comput* 1997;**29**:274–79.
- 56 **Rhodes SD**, DiClemente RJ, Yee LJ, et al. Hepatitis B vaccination in a high-risk MSM population: the need for vaccine education. *Sex Transm Infect* 2000;**76**:408–9.
- 57 **Sy FS**, Rhodes SD, Choi ST, et al. The acceptability of oral fluid testing for HIV antibodies. A pilot study in gay bars in a predominantly rural state. *Sex Transm Dis* 1998;**25**:211–15.
- 58 **Koch S**, Schockman HE. Democratizing internet access in the lesbian, gay and bisexual communities. In: Ebo BL, ed. *Cybergetto or cybertopia? Race, class and gender on the internet*. Westport, CT: Praeger, 1998:170–84.
- 59 **Serraino D**, Franceschi S, Dal Maso L, et al. The classification of AIDS cases: Concordance between two AIDS surveillance systems in Italy. *Am J Public Health* 1995;**85**:1112–14.
- 60 **Fishbein M**, Pequegnat W. Evaluating AIDS prevention interventions using behavioral and biological outcome measures. *Sex Transm Dis* 2000;**27**:101–10.
- 61 **Pequegnat W**, Fishbein M, Celentano D, et al. NIMH/APPC workgroup on behavioral and biological outcomes in HIV/STD prevention studies: a position statement. *Sex Transm Dis* 2000;**27**:127–32.
- 62 **Pastore M**. The world's online populations. *CyberAtlas com* 24 Aug 2001. Available at: [http://cyberatlas.internet.com/big\\_picture/geographics/article/0,1323,5911\\_151151,00.html](http://cyberatlas.internet.com/big_picture/geographics/article/0,1323,5911_151151,00.html) [Accessed 17 Sep 2001].
- 63 **Pastore M**. The web: more than 2 billion pages strong. *CyberAtlas com* 12 Jul 2000. Available at: [http://cyberatlas.internet.com/big\\_picture/traffic\\_patterns/article/0,,5931\\_413691,00.html](http://cyberatlas.internet.com/big_picture/traffic_patterns/article/0,,5931_413691,00.html) [Accessed 17 Sep 2001].
- 64 **Bernhardt JM**. Health education and the digital divide: building bridges and filling chasms. *Health Educ Res* 2000;**15**:527–31.
- 65 **Sorensen AA**. Promoting public health through electronic media: a challenge for schools of public health. *Am J Public Health* 2001;**91**:1183–5.
- 66 **Wolf A**. Exposing the great equalizer: demythologizing. In: Ebo B, ed. *Cyberghetto or cybertopia: race, class, and gender on the internet*. Westport, CT: Praeger, 1998:15–32.
- 67 **National Telecommunications and Information Administration**. *Falling through the net: defining the digital divide. A report on the telecommunications and information technology gap in America*. Washington, DC: US Department of Commerce, 1999.
- 68 **Milio N**. The internet, public health, and the globalisation of just about everything. *J Epidemiol Community Health* 2001;**55**:74–6.
- 69 **US Government Working Group on Electronic Commerce**. *Leadership for the new millennium. Delivering on digital progress and prosperity*. Washington, DC: US Department of Commerce, 2000.
- 70 **Rhodes SD**, DiClemente RJ, Cecil H, et al. Risk among men who have sex with men in the United States: a comparison of an Internet sample and a conventional outreach sample. *AIDS Ed Prev* 2002;**14**:41–50.
- 71 **Nicholson T**, White J, Cline R, et al. Parents who report using illicit drugs: findings and implications from the DRUGNET study. *Psychol Rep* 2001;**88**:245–51.
- 72 **Reneau J**, Nicholson T, White JB, et al. The general well-being of recreational drug users: a survey on the WWW. *International Journal of Drug Policy* 2000;**11**:315–23.
- 73 **Humphreys K**, Klaw E. Can targeting nondependent problem drinkers and providing Internet-based services expand access to assistance for alcohol problems? A study of the moderation management self-help/mutual aid organization. *J Stud Alcohol* 2001;**62**:528–32.
- 74 **Migliaccio TA**. Marginalizing the battered male. *J Men's Stud* 2001;**9**:205–26.
- 75 **Durkin KF**, Bryant CD. Propagandizing pederasty: a thematic analysis of the on-line exculpatory accounts of unrepentant pedophiles. *Deviant Behavior* 1999;**20**:103–27.
- 76 **Coomber R**. Using the Internet for survey research. *Sociological Research Online* 1997;**2**. Available at: <http://www.socresonline.org.uk/socresonline/2/2/2.html> [Accessed 10 Sep 2001].
- 77 **Global Research**. *Global Internet statistics (by language)*. Available at: <http://global-reach.biz/globstats/index.php3>. [Accessed 17 Mar 2002].
- 78 **Lakeman R**. Using the Internet for data collection in nursing research. *Comput Nurs* 1997;**15**:269–75.
- 79 **National Institute for Literacy**. 1992 National Adult Literacy Survey [http://www.nifl.gov/nifl/faqs.html#literacy\\_rates](http://www.nifl.gov/nifl/faqs.html#literacy_rates). [Accessed 4 Jul 2001].
- 80 **The Rehabilitation Act Amendment**, 29 USC §794d (1998).
- 81 **Bush NE**, Wooldridge J, Foster V, et al. Web site design and development issues: The Washington State Breast and Cervical Health Program Web Site Demonstration Project. *Oncol Nurs Forum* 1999;**26**:857–65.
- 82 **Crocker SD**. Operational security. In: Lynch DC, Rose MT, eds. *Internet system handbook*. Greenwich, CT: Manning; 1993:677–704.
- 83 **Bier MC**, Sherblom SA, Gallo MA. Ethical issues in the study of internet use: uncertainty, responsibility, and the spirit of research relationships. *Ethics Behav* 1996;**6**:141–51.
- 84 **Children's Online Privacy Protection Rule**, 16 CFR §312.5a (2000).
- 85 **Standards for Privacy of Individually Identifiable Health Information (Privacy Rule)**, 45 CFR §160, 164 (2001). Technical guidance on the Privacy Rule from the US Department of Health and Human Services is available at: <http://www.hhs.gov/ocr/hipaa/research.html>
- 86 **Department of Health and Human Services Protection of Human Subjects Rule**, 45 CFR Part 46 (2000).
- 87 **Food and Drug Administration Protection of Human Subjects Rule**, 21 CFR Part 50 (2001).