The Growth of Internet Research Methods and the Reluctant Sociologist*

Dan Farrell, Western Michigan University
James C. Petersen, University of North Carolina at Greensboro

Between 1999 and 2004 only one article appeared in the American Sociological Review, the American Journal of Sociology, or Social Forces using primary data collected with Web-based research techniques. Since then there have been only a handful of studies published in these core sociology journals drawing on Web-based surveys or other forms of Web-based data. The use of Internet-based data has become widespread in many academic fields, especially health research and education; Web-based techniques are becoming routine in the practice and study of politics; and online commercial and market research has become a billion dollar industry. At the same time, the utility of random digit dialing surveys has eroded considerably owing to declining contact rates, increased use of technologies to screen unwanted telephone calls, and the replacement of landline telephones with cell phones. There is increasing evidence that Internet research can produce representative data. Although Web-based surveys may overrepresent some populations, Internet usage in the general population is now well over 75 per cent and is especially strong among some hard to reach populations. Internet surveys have the potential to reduce measurement error, missing data, and respondent attrition. Sociologists must overcome their fear of participation in stigmatized Internet research and actively engage in the development of techniques and refinements that will increase the utility and validity of Internet-based data collection.

The Reluctant Sociologist

Although a few sociologists have been key players in the emergence of Web-based research techniques, the discipline of sociology as a whole has lagged behind many other academic and professional groups in utilization of Sociology (AJS), and Social Forces, between 1999 and 2004. An article (Griswold and Wright 2004) based on data from Survey 2000, an online survey conducted in 1998 by the National Geographic Society, was published in the American Journal of Sociology at the end of this period. In addition, an examination of volumes of the Annual Review of Sociology between 1999 and 2008 showed that no methods in chapters had focused on Internet research methods.

Since 2004 there have been just a handful of articles published in the ASR, AJS, and Social Forces that either used Web-based survey data to sup-
plement other data sources or made limited use of other forms of Web-based
data. For example, a study of social status in an Internet open-source software
community explored the use of a peer certification system by software devel-
opers using data drawn from individual members’ Web sites (Stewart 2005).
A study of storytelling related to the September 11, 2001 attacks drew on data
from a questionnaire provided to participants in a face-to-face forum, storytell-
ing at the forum, and postings on an online forum about the future of Lower
Manhattan (Polletta and Lee 2006). In a study of factors related to academic
compensation, Leahey (2007) obtained data from a variety of secondary
sources and a Web-based survey of faculty. Finally, a study of social relation-
ships among capitalists combined fieldwork on industry peer networks (IPN),
a survey of U.S. industries, telephone interviews, and data from 2 years of a
Web-based survey of one IPN—Business Networks (Zuckerman and Sgourev
2006).

This very limited use of Internet research is especially surprising given
the extent to which sociologists have traditionally been innovators in develop-
ing and refining data collection techniques that were later adopted by other
fields, especially other social sciences. Sociologists have, however, begun to
recognize the rapidly growing importance of e-mail communication, virtual
communities, and widespread access to the Internet as factors influencing
human behavior. DiMaggio et al. (2001:329) observed that “sociology has
been slow to take advantage of the unique opportunity to study the emergence
of a potentially transformative technology in situ.” They called for socio-
lologists to conduct more research on the Internet and to link research findings on
individual Internet users with macro-level political, economic, and institutional
factors (p. 307). Zhao (2006) examined the impact of the Internet on the social
construction of reality and observed that the Internet has created a new spatio-
temporal zone, a new form of communication—electronic text chat, and a new
form of social gathering—the online public domain. The emergence of Face-
bok, MySpace, blogs, Twitter, and other forms of electronic networking have
created new sources of potential data for sociological research. Although the
advantages and limitations of Web-based surveys were certainly recognized
(Sills and Song 2002), it was not until 2009 that a special issue of a sociology
journal, Sociological Methods and Research, was devoted to Web surveys.

Much of the reluctance of sociologists to collect data for scholarly
research using Web-based techniques seems to be based on concerns about the
quality of data (Shropshire, Hawdon, and Witte 2009) obtained through such
approaches, a belief that it is not possible to obtain broad population coverage
or representative samples with Web-based approaches, and perceptions that
threats to both internal and external validity abound in the Web environment
(Schonlau et al. 2009). Issues of such importance properly demand our full
attention. The problem, however, is that there is a growing body of evidence that some of these widely held assumptions are either false or at least less and less accurate as the technology associated with the Internet develops, data collection techniques designed for the Web environment are refined, and social acceptance of Web-based technology increases.

**Increasing Use of Internet Research**

Nowhere has the growth in the use of Web-based surveys and other Internet approaches been stronger than in commercial survey and market research. Dan Coates, an executive for SPSS, Inc., reported that a research industry survey found that expenditures for online research grew by 24 percent in the United States in 2003 and were approaching a billion dollars (Coates 2004). Extensive use of Internet research is, however, certainly not confined to market research and other forms of commercial survey research. Government agencies in the United States including the Census Bureau and the U.S. General Accounting Office have utilized various forms of Internet research. In addition, both the practice of politics and the scholarly study of politics are rapidly making use of Web-based techniques. For example, Knowledge Networks, founded by sociologist Norman Nie, was commissioned to use Internet polling to provide data for a study of voter perceptions and misperceptions about the war in Iraq (Kull, Ramsey, and Lewis 2003).

Scholarly journals in many fields other than sociology have shown a willingness to publish research based on data collected through e-mail surveys or through various Web-based approaches. Such articles have been especially frequent in health research and studies in education. The ability of Internet research methods to provide study participants with both convenience and privacy is an asset in studies of special populations such as alcoholics (Brodey et al. 2004), surgical patients (Naylor et al. 2002), persons participating in a smoking cessation intervention (Brendryen and Kraft 2008), college students reporting a traumatic loss (Schnider, Elhai, and Gray 2007), and college students enrolled in a university dining plan (Kolodinsky et al. 2007).

**Declining Effectiveness of Telephone Surveys**

The emergence of innovative Web-based research, especially approaches suited for survey research, could not be timelier. Random Digit Dialing (RDD) telephone surveys have been considered the gold standard for accurately representing the general population since the mid-1970s (Dillman 2000). However, beginning in the 1990s, survey researchers have reported erosion in the efficacy of RDD surveys. Studies conducted by the Gallup Organization (Tortora 2004) show 24 consecutive quarters of declining contact rates. This trend has also been reported in U.S. government-conducted surveys during the
same period (Atrostic et al. 2001). Interestingly the explanation lies not solely in refusals, which may be declining (Tortora 2004) and is probably not because of completion burden or complexity (Atrostic et al. 2001), but rather owing to the difficulty of reaching individuals.

Social and technological changes apparently account for most of decreasing response rates. Answering machines, caller ID, and privacy management technologies that give individuals the ability to block unfamiliar callers and unlisted telephone numbers are all barriers to response (O’Rourke et al. 1998). The proliferation of telephone numbers for fax and Internet lines also contribute to declining response rates (Tuckel and O’Neill 2001). Finally, there are now many more cell phones in the United States than landline phones. CIA’s (2008) *The 2008 World Factbook* reported that there were 233 million mobile cellular telephones in the United States in 2006 as compared with 172 million main lines in use. In recent years wireless telephones have been increasingly replacing landline telephones. By 2008, 17.1 percent of U.S. households had replaced landlines with cell phones (Nielsen Company 2008). The growth of cell phones and their increasing replacement of landlines as the only phone in a household raise additional concerns about measuring each individual’s probability of being included in a sample.

Social changes in the attitudes toward phone surveys reinforce the changing technology and make reaching respondents difficult. The language of researchers regarding the impact of telemarketing portrays the frustration. “The proliferation of telemarketing” (Couper 2000:465) is threatening the use of telephone surveys as a mode of data collection for representative samples of the general population by creating an “over surveying effect.” Further explaining the effect, researchers (O’Rourke et al. 1998: 1) note that these “unsolicited, much hated calls from telemarketers ... often attempt to mimic surveys.” In a combination of social and technological change, it appears that the growing use of unlisted phone numbers is associated with a propensity to be “hostile toward participation” (Tortora 2004:222) when reached in RDD telephone surveys.

**Internet Research as a Possible Replacement Technology**

*Representativeness*

Empirical studies increasingly suggest that Internet polling may be a replacement technology (Couper 2000). There is increasing evidence that Internet research can produce representative data. The key issue in representativeness is coverage error, that is, the degree to which there is a match between the target population and the sampling frame population (Dillman 2000). Several years ago, Dillman (2000: 356) noted that “e-mail and web
surveys have only minor coverage problems for higher computer use popula-
tions.” The Pew Internet and American Life Project reported data from a
late 2007 survey of Americans adults (18 years and older) showing that 75
percent use the Internet at least occasionally. Especially high Internet utiliz-
ation was found among respondents 18–29 years old (92 percent), those with
household incomes of $75,000 and above (93 percent), and those with at
least a college education (93 percent; Pew Internet 2008).

Other early coverage issues such as the predominance of men on the
Internet have disappeared (Bruzzone 1999). Internet surveys resulted in “a
final sample that more closely matched the target sample in gender mix than
did the U.S. mail survey mode” (McCabe et al. 2002:755). Comparisons of
RDD recruited, but Internet-administered surveys and Internet panel surveys
(Krosnick and Chang 2001:3) revealed, “in general average deviations (from
U.S. Census) are not large and sample representativeness is never dramatically
poor in terms of percentage point deviation of any survey estimate from the
population.” Pew research (Couper 2000) comparing RDD, volunteer Web
samples and selected Web panels concluded,

there were no predictable patterns of success or failure of Internet surveys. Respondents
who took the survey online were not consistently more conservative or liberal than nation-
wide telephone surveys, nor were they more optimistic or pessimistic. (p. 472)

Empirically there is evidence that Internet surveys can replace RDD tele-
phone surveys as a representative tool. A valuable study comparing the repre-
sentativeness of a RDD telephone survey with an online survey (Bethell et al.
2004) demonstrated that each approach has representational bias in portraying
the underlying U.S. population. Comparing online and telephone surveys, even
after corrections are applied, both underestimate the 18- to 24-year-old popula-
tion and the less educated population and overestimate the college educated
population. Online surveys underestimate the over 65 population slightly and
more substantially overestimate the white population. The researchers note,
however, that despite these imperfections “conclusions about the level and
variations in health care quality in the United States are similar whether based
on data collected online or data collected using more elaborate and costly
survey methods” (Bethell et al. 2004:e2). Lee and Valliant (2009) also
demonstrated that weighting adjustments can improve Web survey estimates
impacted by sample selection bias.

Sampling Frames

Affirmative consent e-panels are emerging as an alternative sampling
frame. To obtain addresses, list compilers typically engage in a cooperative
process with other organizations that obtain affirmative consent of potential
respondents. Large retailers such as video rental companies, banks, and credit card providers collect e-mail addresses of their customers. These retailers then offer to refer customers to list companies, noting that incentives are usually offered to participants. This is known as the enrollment process. Use of an enrollment process avoids legal problems and usually overcomes screening software and firewalls. The use of e-panels is an alternative to so-called Internet river sampling where a continuing flow of potential respondents is recruited through pop-ups and promotions on various Web sites. Sample frame compilers (e.g., e-rewards, FGI Research, and Affordable Samples) now advertise panels exceeding 1 million names. Partnering with AOL, Digital Marketing expects to offer panels of 5 million (Bruzzone 1999:5).

As one researcher noted (Bruzzone 1999), all commonly accepted sampling techniques have some self-selection component, commonly called "opting in." The difference with affirmative panels is timing. Internet panels are built following consent and RDD surveys seek consent following contact. Internet panelists, of course, still retain the right to decline participation. Further, the much smaller size of the e-panel list clearly suggests that there may be a threat to validity based on repeated exposure. An Internet research firm, e-Rewards, considers this by both tracking the number of surveys a panelist completes each year and asking at the enrollment point how frequently a respondent wishes to participate. Some corrections for conditioning are being tested (Couper 2000). Finally, sampling issues may be addressed by hybrid approaches. Knowledge Networks has used e-panels recruited through RDD and then surveyed through the Internet.

**Internal Validity**

Opportunities for improving internal validity may further accelerate the trend to online surveys. Reduced measurement error is often cited as an advantage of Internet data collection over other surveys. A careful comparison study (Krosnick and Chang 2001) demonstrated that Internet responses were more accurate than those provided through telephone interviews. The list of explanations for measurement errors with self-administered questionnaires includes (Couper 2000) lack of motivation, comprehension problems, poor wording, and deliberate distortion. The most common explanation for reduced measurement error with Internet surveys is the opportunity for visible response options (Bruzzone 1999; Couper et al. 2004). When well-designed procedures are used, response rates with Internet surveys are equivalent to mail surveys (Crawford, Couper, and Lamias 2001). Pictorial presentations, such as candidate images, provide a clear advantage over telephone surveys. Recall, varied question streams, and other custom features can be controlled better in self-administered Web polls than with other technologies.
Although not yet well established, Internet surveys may reduce problems posed by missing data and respondent attrition. Comparing Web survey respondents with mail survey responses (McCabe et al. 2002:759) revealed “web respondents had marginally lower missing data rates.” Earlier reported surveys of surgical patients indicated improved effectiveness in avoiding respondent attrition. Internet surveys avoid fatigue problems as respondents can usually pause at will, self-pace, and/or self-schedule completion at their convenience. Shropshire, Hawdon, and Witte (2009) also note the ability with Web surveys of pushing information to engage respondents and to make use of partial data from terminated surveys. Whether Internet surveys improve perceptions of privacy is debatable. On one side is the perception that interviewer reassurances of confidentiality improve perceptions of privacy (Couper 2000). The alternative view holds that the impersonal contact provided by a video display makes responding to personal questions seem less intimate and less embarrassing. Overall, there is a potential for empowerment in Internet surveys that should reduce respondent attrition.

Technology Transition

The extent to which Web-based surveys replace telephone surveys as the dominant survey methodology will be decided in the near future. The 2004 and 2008 election coverage included CNN’s Poll of Polls. This informal meta-analysis of election polls attempted to ensure that the results were based on methodologies generally accepted among professionals. Included in these analyses were surveys from the Zogby organization. The Zogby methodology relied on 25 percent Web-based interviews (Zogby International 2008). The Zogby surveys were especially accurate in predicting caucus-state results in the 2008 election.

Overcoming Reluctance

Sociologists and other social researchers must overcome their fear to engage in stigmatized Internet research and actively engage in the development of techniques and refinements that will increase the utility and validity of Internet-based data collection approaches. The telephone-based survey techniques that have served sociology so well in recent decades are being rendered ineffective by social and technological changes. It is understandable that sociologists wish to continue utilizing techniques where major investments have been made in developing RDD approaches, specialized sampling approaches such as Mitofsky–Waksberg techniques, and computer-assisted telephone interviewing facilities and software. If sociologists continue to cling to these familiar methods while ignoring the increasing utility of Internet-based approaches, the discipline risks falling far behind those fields more open to new technologies.
Internet research methods vary widely in their potential for significant social research. A few approaches may have little to contribute to scholarly work. River sampling via pop-ups, for example, may be the virtual analog to mall intercepts. Other approaches, however, have a great deal of potential and many of their limitations are receding rapidly. Göritz and Wolff (2007) recently noted, for example, that online panels are increasingly popular mechanisms for data collection because of a wide array of benefits including “easy identification of key sample segments, usually high response rates, short field times, validation of respondents’ answers on the basis of previously collected data, limitation of questionnaires to novel items, and ethical advantages” (Göritz and Wolff 2007:99). In his discussion of technology trends in data collection, Couper (2005) notes the radically divergent perceptions of those who see technological innovations as opening up new possibilities to enhance and extend survey approaches and those who see such change as threatening an end to survey research.

Recommendations

If Internet surveys are to be a major tool in the research arsenal of sociologists, a number of challenges must be addressed. Motivation to respond to Internet surveys has multiple dimensions. Beyond the “opt-in” issues discussed earlier with respect to sampling frames, e-mail invitations do not have the demand characteristics of either a ringing telephone or a sealed, first-class letter. Survey researchers need to learn the optimal combination of e-mail or surface mail pre-notice (Porter and Whitcomb 2003; Trouteaud 2004), which reminder protocols are optimal, and what small adjustments and personalizations increase responses (Kaplowitz, Hadlock, and Levine 2004).

A major issue in Internet response is overall motivation. Two major approaches have been explored—financial incentives and emotional appeals. Small guaranteed payments ($10) apparently increase response rate (Tourkin et al. 2005) whereas entry into lotteries have no significant effect (Marcus et al. 2007) or a first-wave effect only (Göritz and Wolff 2007). Another approach is the emotional appeal: “Pleading for the help of the respondent throughout the survey can have constructive effect” (Trouteaud 2004:390). Multiple motivational approaches may well be needed as Internet surveys do lack the interpersonal warmth and rapport that can be subtly communicated in the telephone interviews.

An additional challenge to Internet surveys apparently not yet represented in the scientific literature is the impact of spam defense. The consequences of spam filters will likely vary by target population. Academic and business environments and most of Europe have already demonstrated a high level of con-
cern regarding spam. With targeted population research, surveyors may request white listing to avoid e-mail blocking programs as an initial step in the data collection process.

The immediate future appears to favor mixed mode approaches. Combinations of Internet and RDD surveys (Lee and Valliant 2009) may allow one to be used as a reference sample and weightings to be used to adjust for possible sampling bias. In any case, the distinction between methods is blurring (Couper 2005). Both telephone and Internet surveys offer branching, randomization, skips, and jumps. Experiments are becoming larger, approaching the size of small surveys. Images online and interactive blogs make ready substitutes for focus groups.

Research design may need to be specifically tailored to demographic groups (Diment and Garrett-Jones 2007). Groups about which the “basic parameters such as population size and member location” are well known (Trouteaud 2004:385) are already seen as an excellent fit for Web-based research PR TOOLBOX (2006). The reasons to move beyond this to studies of the general population are numerous. First, there are sociological issues that are issues of the general population. Voting behavior undoubtedly tops the list of these issues. Given that e-voting is a rapidly developing technology (Kenski 2005; Oostveen and Van Den Besselaar 2005), it follows that Internet polling can offer valued predictions. Broad social changes also suggest that Web-based surveys will have increasingly relevant application to the general population. The landline or fixed-line household sampling unit that may have mirrored reality in the 1970s appears quaint in light of today’s focus on mobile telephone numbers and e-mail addresses, ubiquitous computing, and increasing diversity in the definitions of family and household. An undeniable force for Internet surveys of the general population is reduced cost compared with other technologies and/or respondent convenience. Researchers must recognize that the communication media of an e-powered society is not fixed-line telephony.

ENDNOTE

*The authors thank the editor, an anonymous reviewer, and Eric J. Petersen for their comments and many valuable suggestions on an earlier draft of the manuscript. They also thank Cathy (Cat) Collins for her careful copyediting of the manuscript. Please direct correspondence to James C. Petersen, The Graduate School, University of North Carolina at Greensboro, 241 Mossman Building, Greensboro, NC 27402, USA; e-mail: jcpeters@uncg.edu.
REFERENCES


