

Strategies to address participant misrepresentation for eligibility in Web-based research

JESSICA KRAMER,¹ AMY RUBIN,² WENDY COSTER,¹ ERIC HELMUTH,³ JOHN HERMOS,²
DAVID ROSENBLOOM,³ RICH MOED,⁴ MEGHAN DOOLEY,¹ YING-CHIA KAO,¹
KENDRA LILJENQUIST,¹ DEBORAH BRIEF,⁵ JUSTIN ENGGASSER,⁵ TERENCE KEANE,⁵
MONICA ROY⁵ & MARK LACHOWICZ⁵

1 Department of Occupational Therapy, Boston University, Boston, MA, USA

2 Department of Medicine, Boston University School of Medicine, Boston, MA, USA

3 Department of Health Policy and Management, Boston University, Boston, MA, USA

4 CREcare LLC

5 VA Boston Healthcare System, Boston, MA, USA

Key words

Internet data collection, World Wide Web data collection, sampling methods, recruitment and enrollment methods, participant misrepresentation

Correspondence

Jessica Kramer, Department of Occupational Therapy, Boston University, 635 Commonwealth Ave. (SAR 503), Boston, MA 02215, USA.

Email: kramerj@bu.edu

Received 19 April 2012;
revised 8 January 2013;
accepted 16 January 2013

Abstract

Emerging methodological research suggests that the World Wide Web (“Web”) is an appropriate venue for survey data collection, and a promising area for delivering behavioral intervention. However, the use of the Web for research raises concerns regarding sample validity, particularly when the Web is used for recruitment and enrollment. The purpose of this paper is to describe the challenges experienced in two different Web-based studies in which participant misrepresentation threatened sample validity: a survey study and an online intervention study. The lessons learned from these experiences generated three types of strategies researchers can use to reduce the likelihood of participant misrepresentation for eligibility in Web-based research. Examples of procedural/design strategies, technical/software strategies and data analytic strategies are provided along with the methodological strengths and limitations of specific strategies. The discussion includes a series of considerations to guide researchers in the selection of strategies that may be most appropriate given the aims, resources and target population of their studies. Copyright © 2014 John Wiley & Sons, Ltd.

Introduction

Research conducted on the World Wide Web (or “Web”) provides expanded opportunities for reaching target populations (Wright, 2005). Emerging methodological research suggests that the Web¹ is an appropriate venue for survey data collection (Bethell *et al.*, 2004), and a promising area for delivering behavioral interventions

(Eysenbach and Wyatt, 2002; Riper *et al.*, 2011). However, enrollment methods commonly accepted as best practices in familiar research contexts such as mailed surveys or face-to-face intervention studies may not readily translate to Web-based research. Web-based studies may need to utilize new methodological strategies that take into consideration how recruitment and enrollment on the Internet

may present unique challenges to sample validity (i.e. legitimacy) and representativeness (Murray *et al.*, 2009). This concern can be particularly problematic when participants receive monetary incentives for participation (Bowen *et al.*, 2008; Konstan *et al.*, 2005; Murray *et al.*, 2009).

A review of procedures used in Web-based studies suggests that researchers are attempting to use some methodological strategies to address this threat, such as not paying participants for their time and effort (Wilson *et al.*, 2010), recruiting participants in person for a Web-based study (Sinadinovic *et al.*, 2010) or recruiting on the Web but assessing eligibility either in person or on the phone and then mailing consent forms and/or questionnaires to be returned via email or mail (Bromberg *et al.*, 2011; Eysenbach and Wyatt, 2002; Riper *et al.*, 2011). Relying on these more traditional strategies for recruitment and enrollment does not capitalize on the full benefits and appeal of Web-based research.

There are a number of reasons to develop new strategies that enable researchers to conduct rigorous, secure, and valid research on the Web. First, given that one typical goal of Web-based research is to efficiently acquire a large, broadly based sample, an equally efficient means for excluding inappropriate participants is necessary. A second potential advantage of Web-based, automated approaches to screening is that the fidelity of the screening, recruitment and eligibility processes precludes some of the subjective judgments that may occur in traditional, face-to-face clinical trials. Third, studies conducted completely on the Web provide a relatively anonymous environment for high risk or sensitive populations and for those concerned about the stigma of being labeled with certain conditions. Finally, recruitment on the Web may be the most appropriate way to reach the target population. If the eventual goal of a study is to establish effectiveness of an online intervention, then a sample recruited directly from online communities would more accurately represent that population.

This paper focuses on the validity of samples recruited and enrolled on the web, particularly in instances where monetary incentives are offered. Specifically it addresses the problem of participant misrepresentation; that is, individuals who misrepresent their characteristics and eligibility for a study in order to obtain benefits of participation such as payment, attention, or access to treatment. The purpose of this paper is to describe the challenges experienced in two Web-based studies in which participant misrepresentation threatened sample legitimacy and to suggest a broad range of potential solutions. Although the objectives, samples recruited and extent of anonymity required for the two studies were quite different, both

involved an initial design in which there was no direct contact with participants. Each study team chose different methods to rectify the vulnerabilities in their design, but their choices were based on similar considerations. Based on the lessons learned from these experiences, this paper will propose three types of strategies that researchers can use to reduce the risk of misrepresentation for enrollment in Web-based research. The two studies are described below to provide context for the discussion of these strategies.

Overview of the research studies

PEDI-CAT-ASD

This study was a one-time survey of parents of children with Autism Spectrum Disorders (ASDs). The purpose of this study was to determine initial validity for a commonly used assessment of adaptive behavior, the Pediatric Evaluation of Disability Inventory – Computer Adaptive Test (PEDI-CAT) (Haley *et al.*, 2011; Kramer *et al.*, 2012) for children and youth with ASDs ages 3–21. The PEDI-CAT is a parent self-report of children's ability to perform everyday activities in four domains: Daily Activities, Mobility, Social/Cognitive and Responsibility. The PEDI-CAT is widely used as a standard outcome measure in pediatric rehabilitation and also as a clinical assessment tool.

The initial recruitment and data collection procedures for the PEDI-CAT-ASD survey utilized a universal user name and password posted on the research website to provide potential respondents with immediate, open access to the survey portal. Eligibility was confirmed online; respondents answered a series of questions to determine if they met inclusion criteria. Respondents could only proceed with the full survey if the answers to the questions confirmed eligibility. There was no direct contact with the research team. This approach was selected in order to reduce the burden on parents of children with ASDs, who may be very busy and have limited time and resources to participate in research. Nationwide recruitment was undertaken via advocacy and parent support groups associated with autism and disability, both virtually through LISTSERVs and website postings and through mailed flyers that were distributed at conferences and service agencies. Recruitment materials stated that, upon completion of the full online survey, participants would receive payment (an Amazon electronic gift card) for their time. In the first three weeks of data collection using this protocol, a pattern of suspicious responses was identified. For example, overnight between 15:00 and 07:29, 53 new respondents began the survey only minutes apart. Data collection was shut down while the research team modified the enrollment procedures.

VetChange

This study was a Web-based behavioral intervention program for veterans returning from the Iraq and Afghanistan wars suffering from high rates of problem drinking and traumatic stress symptoms (Brief *et al.*, 2012; Hoge *et al.*, 2004; Jacobson *et al.*, 2008). The Web-based program was conceived and developed to provide a readily available and essentially anonymous self-help intervention as a way to mitigate the adverse consequences of heavy drinking while being sensitive to the needs of returning veterans, who were concerned about both stigma and access to care (Hoge *et al.*, 2004). The program consisted of eight modules based on Motivational Interviewing (Miller and Rollnick, 1991) and cognitive behavioral therapy with research assessments at baseline, after active intervention, and at three months follow-up. A wait list control design was used, in which participants randomized to the delayed treatment group were invited back at eight weeks to complete the program.

All aspects of the research were conducted on the Web: recruitment, informed consent, randomization, assessment, treatment, and payment via Amazon electronic gift cards. Participants could be identified only via their valid email address; no other identifying information was collected. Starting in the sixth week of recruitment, a surge of 119 new participants enrolled during the early morning hours on a weekend. The website was immediately shut down and measures were taken to address this seemingly fraudulent enrollment.

This paper describes the strategies identified retrospectively by the two research teams that may help reduce the risk of participant misrepresentation during enrollment in web-based studies.

Strategies to address participant misrepresentation in web-based studies

Researchers can use three types of strategies to identify and minimize misrepresentation by participants seeking enrollment in Web-based studies: procedural/design strategies, technical/software strategies and data analytic strategies (see Table 1). Although we describe these strategies separately, both the PEDI-CAT-ASD and VetChange studies used a combination of these strategies to ensure eligibility within their samples.

Procedural/design strategies

Rigorous research studies typically implement procedural safeguards to ensure the validity and eligibility of samples. However, Web-based studies may require unique or

additional procedures during the recruitment and enrollment process.

One strategy is to *limit access to Web-based data collection*. Enabling any potential visitor to a research website to access and enroll in a study (“open enrollment”) can make it easy for individuals misrepresenting themselves to participate multiple times. This risk may not be sufficiently reduced by requiring individuals to create a unique username and password or provide a valid email address. Individuals seeking to misrepresent themselves can easily create multiple, unique email addresses (Bowen *et al.*, 2008; Murray *et al.*, 2009). In contrast, there may be some benefits to allowing open enrollment for Web-based studies. Open enrollment may enable a research team to recruit a larger sample, and this convenience sample approach may be appropriate for some research aims. Researchers may also use a modified open enrollment procedure in which all visitors have the ability to access the study website but must answer qualifying, or inclusion questions in order to proceed with study activities. The PEDI-CAT-ASD original procedures offered open enrollment as researchers hoped this procedure would reduce the burden associated with participation in research for those respondents. If open enrollment is the most effective way to build a Web-based sample, additional safeguards described later may reduce the risk of participant misrepresentation.

A second procedural strategy is to *require potential respondents to answer inclusion criteria questions or demographic questions that demonstrate insider knowledge*. For example researchers may ask questions or solicit answers that use terms or slang unique to specific populations or sub-cultures. Responses can then be monitored and reviewed for credibility. The revised enrollment procedures for the PEDI-CAT-ASD study asked potential participants to name the types of services their child received via email before receiving a confidential password to access the study site. The research team looked for respondents to name specific professions commonly providing intervention services to children and youth with autism, or list the acronyms used to refer to those professions when talking with another individual “in the know”. Respondents who gave vague or brief answers, such as “all of them” or “school” were asked to clarify their responses before access was granted. In the VetChange study, after baseline assessment was completed, a review of answers to questions on deployment revealed suspect answers such as “shock troops” being listed as a branch of the US military service. The patterns of deployment reported by respondents were also monitored by a research team member knowledgeable

Table 1. Strategies to ensure sample validity in Web research

Strategy category	Specific techniques
Procedural/design	<ul style="list-style-type: none"> • Limit access to Web-based data collection sites. (PEDI) • Require potential respondents to answer questions that demonstrate “insider knowledge”. (PEDI, VC) • Ask potential participants to report how and where they found out about the research study and provide actual links to these Web pages. (PEDI) • Do not advertise the amount or type of compensation that will be provided to study participants. (VC) • Collect the same information at multiple points and then examine for consistency. (PEDI) • Develop a plan to re-contact potentially suspicious respondents as part of the research protocol prior to data collection. (PEDI, VC)
Technical/software strategies	<ul style="list-style-type: none"> • Track the Internet Protocol (IP) address of individual respondents to identify potential multiple enrollees. (VC) • Gather date and time stamps for Web-based responses to identify respondents completing surveys more quickly than expected. (PEDI, VC) • Use software created to reduce risk of attacks from programs written to automatically populate surveys. (e.g., CAPTCHA) • Restrict enrollment to respondents who enter a Web data collection site through approved Web links (URLs). (VC)
Data analytic strategies	<ul style="list-style-type: none"> • Identify pairs of items that can be examined together to evaluate the logic of an individual participant’s responses. (PEDI, VC) • Analyze the extent to which responses are in keeping with previous research with the target population. (PEDI, VC) • Analyze the extent to which responses from participants suspected of misrepresenting eligibility differ from the remaining study sample (sensitivity analysis).

Note: PEDI, PEDI-CAT-ASD web-based study; VC, VetChange web-based study.

about typical patterns, and unusual reports were flagged for further scrutiny.

Another procedural strategy is to *ask potential participants to report how and where they found out about the research study, and when feasible, to provide actual links to Webpages, posts, or other announcements about the study.* This strategy enables researchers to track the usefulness of various recruitment efforts as well as to confirm that potential participants learned about the study through legitimate recruitment efforts. For example, several potential participants in the PEDI-CAT-ASD project reported they learned about the study through the “lab website”, which they did not name and which appeared to be, arguably, a generic response. It was unlikely that parents heard about the study through this source and the researchers did not grant access to individuals who reported this as their source of information. This strategy may be less effective if researchers are relying on a snowball sampling or a word-of-mouth recruitment strategy.

The risk of misrepresentation may also be reduced by *not advertising the amount or type of compensation that*

will be provided to study participants. Some research with Web-based samples suggests that the amount of compensation may not impact response rate (Wilson *et al.*, 2010). However, in the Wilson study the target population was other researchers who may not have been motivated by financial incentives. Researchers recruiting from broader Web-based populations may find that financial incentives encourage respondents to misrepresent themselves and engage multiple times in the same research study. Even small amounts, such as \$10, can be financially lucrative if a respondent is able to participate multiple times in the same study. After the fraudulent enrollments in VetChange, the researchers found that a Web search for “paid online surveys” identified a number of websites that provide information about Web-based studies offering financial compensation, making it easier for individuals or groups of people to identify studies that allow them to earn money.

An alternative is to *avoid offering financial incentives such as electronic gift certificates to all participants,* and instead to hold a drawing for a gift from the final pool of eligible participants after all data collection is complete.

The time delay and the reduced certainty of receiving compensation may reduce the perceived benefits of participation for fraudulent responders. If anonymity is not a strong consideration, another alternative is to obtain physical addresses for the purpose of mailing incentives to participants and then confirming the address via phone directories or other means before providing the incentive. Using a similar strategy, the PEDI-CAT-ASD team requested a telephone number from participants who were eligible to receive research compensation, and completed brief validation phone calls with suspicious respondents before subjects were authorized for study payment.

Another procedural strategy is to *collect information that respondents would likely know at multiple points of data collection and then examine that information for consistency*. Some examples of information most people have memorized are email addresses, phone numbers, or names (such as the first name of one's oldest child, or the name of the city in which one works). For example, the PEDI-CAT-ASD study asked participants to provide their first name during the consent process at the beginning of the survey and again at the end of the survey to receive compensation. Numerous respondents provided different names at these two time points and, as a result, were eliminated from the participant pool. A related strategy is to ask respondents to answer the same question twice within one period of data collection and to examine responses for inconsistencies. However some participants may have limitations such as cognitive impairments associated with traumatic brain injury that make them more likely to make errors. It is not uncommon for individuals completing information on the Web to make entry errors such as spelling, letter reversal, or keyboarding mistakes. Therefore it may be difficult to distinguish invalid responses from response errors. This strategy is probably best used with caution in selected populations, and preferably used in combination with other strategies.

As a final procedural strategy, *research teams can plan to re-contact potentially suspicious respondents as part of their research protocol*. The protocol can include a set of circumstances or criteria that would necessitate the need to re-contact potentially suspicious respondents. The research team can then use a pre-approved script to contact respondents and ask them to confirm eligibility by contacting the research team directly for a screening, answering additional inclusion questions, completing new eligibility assessments or retaking sections of the original data collection protocol. Both the PEDI-CAT-ASD and VetChange studies employed this strategy.

Technical/software strategies

These strategies harness the power of computer software and other technology to track and monitor participant behaviors during data collection. This information can then be used to identify potentially invalid respondents. The successful implementation of technical strategies typically requires that the research team include an information technology professional with expertise in Web-based surveys, marketing via the Web, Web-application management, and internet security.

The first technical strategy that can help monitor the validity of research participants is to *track the computer Internet Protocol (IP) address of individual respondents*. An IP address is an individual identifier associated with each computer (in the United States, an IP address is considered Protected Health Information (Gunn *et al.*, 2004)). Data collection tools can be programmed to automatically capture and store the unique IP address of each respondent. Examining IP addresses can enable the research team to identify multiple accounts created on the same computer to help identify respondents who are attempting to participate multiple times in Web-based studies. Researchers can also use an IP geographic lookup service to look up a user's approximate likely location (country, state, and sometimes city or town). If the location associated with the IP address is inconsistent with the research team's recruitment strategy, it may be an early warning sign that an individual is misrepresenting him or herself. For example, the IP addresses of the surge of participant enrollments in VetChange were all from a single city in China, which would be a highly unusual location for returning American veterans.

However, there are limitations to tracking IP addresses. The same IP address may appear multiple times if individuals access the Web at public locations such as libraries, community centers, or educational sites. Reliance on IP address alone could lead to over-identification of invalid respondents. Another limitation is that some Internet service providers do not always assign the individual computer one IP address, but instead use a transient IP address (Bowen *et al.*, 2008; Koo and Skinner, 2005). Finally, researchers should be aware that sophisticated users may be able to forge or hide their IP address by using proxy servers and directing their data through multiple IP addresses so that webpage requests from their computer appear to be coming from multiple machines and/or from another geographic location.

Another technical strategy involves *programming Web data collection software to gather date and time stamps for Web-based responses*. When respondents submit data, a

date and time stamp can be associated with each response field or designated response fields; this data can be analyzed to identify unusual or unexpected completion patterns. For example, the researchers on the PEDI-CAT-ASD team found that some respondents were completing the total survey of over 250 items in less than 20 minutes, whereas the majority of respondents required 35–45 minutes. The team also identified “response surges” in which numerous respondents began the survey in a condensed period of time. These time surges were not associated with new Web-postings or specific recruitment events, raising the suspicion that one individual was completing multiple surveys simultaneously. Similarly, the VetChange team administered the baseline assessment to several research assistants naïve to the study to determine the average amount of time to completion (12 minutes). That time was used as an initial guide to flag participants for further examination; those taking less than three minutes to complete the assessment were told that they were ineligible.

However, simply monitoring the dates and times of survey completion may provide a limited picture of misrepresentation without considering other factors. It is possible that respondents familiar with surveys and assessments frequently used with their client population (such as surveys determining stress symptoms or levels of care) may complete data collection quickly. Therefore, research teams are encouraged to obtain median, inter-quartile, and extended time estimates from their full sample in order to identify response times substantially shorter than those of the entire sample.

Researchers can also *use software created especially to reduce the risk of attacks from computer programs written to automatically populate surveys*. This type of software is difficult to circumvent by “phishing” programs, that is, programs written with the specific intent to automatically populate Web fields in order to inundate the system. An example of software that aims to minimize the influence of phishing programs is CAPTCHA, which produces words in distorted, swirly text that respondents must re-type correctly to proceed through a Web-based system. The official CAPTCHA site (<http://www.captcha.net/>) offers a free plugin that is also accessible for Web-users who are blind or visually impaired.

A final technical strategy is to *limit enrollment to respondents who enter a Web data collection site through approved Web links (URLs)*. Web links can be programmed to contain a code that acts like a password to permit enrollments on the site. The coded portion of such URLs can be changed as often as desired to discourage people from sharing them or using them over and over. Because

this restriction need only apply to the enrollment process, enrolled participants may return to use the study website using the public URL without difficulty. The VetChange team used this strategy after revising their online recruitment and only allowed participants to enroll if they came directly from a coded link in a Facebook ad placed by the investigators. The link included an identifier that was changed at least once a day, so if potential respondents distributed or saved the link it could not be used the next day. The code embedded in the URL can be recorded with individual data to be used later to group participants according to the code used to access the website. This strategy may deter individuals from enrolling multiple times. However, the downside to this strategy is that recruitment materials used within or outside of the Web, such as posters or print ads, cannot advertise a direct link to the study website. Therefore, this strategy can limit recruitment options.

Data analytic strategies

Finally, investigators can use data analytic techniques *after data collection* to identify highly unexpected or unlikely response patterns that could indicate that research participants were misrepresenting themselves and their eligibility for a research study, and exclude them from data analysis.

One analytic strategy is to *identify pairs of items that can be examined together to evaluate the logic of an individual participant's responses*. Candidate item pairs should elicit information that is expected to have a particular relationship. For example, in the PEDI-CAT-ASD survey, parents were asked if their child had a current diagnosis of mental retardation and, in a separate item, were also asked to report the results of their child's most recent IQ test. The responses to these two items were compared for congruency. This analytical strategy may be most rigorous when paired items do not appear consecutively and when questions are not grouped or ordered by expected level of difficulty; this reduces the likelihood that respondents who are misrepresenting themselves will remember their responses to previous items and will respond in ways that are consistent with answers to related items. However, this strategy can also increase the amount of time and cognitive effort valid respondents exert during data collection. Therefore, researchers should consider the familiarity of the question or content for the target client group when ordering and spacing paired items throughout their survey instruments.

Researchers can also *analyze the extent to which responses to their study are in keeping with or deviate from*

previous related research with the target client population. For example, Lieberman (2008) compared responses from a Web-based sample with data generated from a mail sample and found congruencies in the findings across the samples. Researchers can also compare their data with known demographic, incidence, and/or prevalence data. For example, the PEDI-CAT-ASD research team was surprised to see a large number of respondents who identified as minority fathers of children and youth with ASDs. The project did not target recruitment to this specific subgroup of parents, and it is well-established in the literature that Caucasian mothers are typically over-represented in research pertaining to children with ASDs. Therefore, respondents who identified as both minority and fathers had their responses further evaluated. In this regard, VetChange researchers compared the age, gender, ethnic makeup and deployment history of their sample with data published by the Department of Defense, finding these parameters to be quite similar.

Researchers may also conduct *sensitivity analyses*, that is, determine if data from a sample of respondents who

are suspected of misrepresenting their eligibility differs significantly from the rest of the study sample or if results of the study are substantially different when either including or excluding their data from analyses. Of course, this analytical strategy may not always be feasible or possible, as there may be no previous research with the target population, the Web-based sample may represent a unique population of respondents, or a subgroup of respondents may represent a unique population.

The effectiveness of analytical strategies is enhanced when the initial data collection plan builds in targeted questions that can be used in a variety of analyses to triangulate the authenticity of responses.

Discussion

Just as in face-to-face enrollment, misrepresentation to gain entry into a study represents a serious threat to validity. Three types of strategies were described to mitigate the risk in Web-based research: procedural, technical and data analytic strategies. When implementing

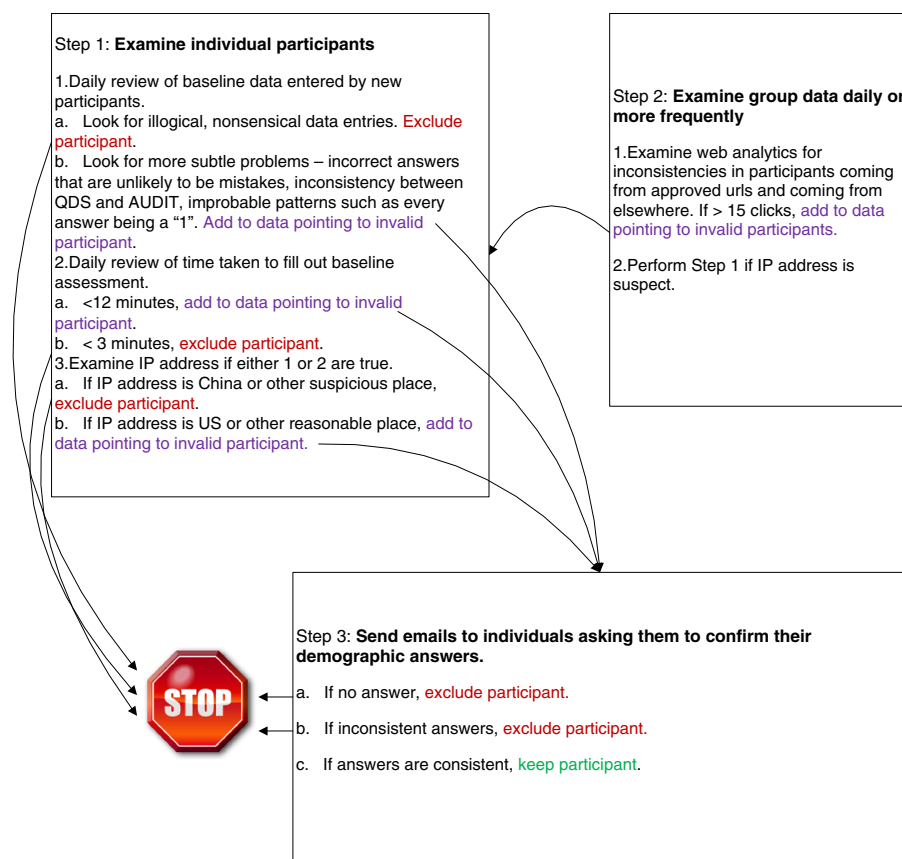


Figure 1. VetChange web security steps flowchart.

these strategies, it is important to recognize that the advantages of conducting research on the Web, such as increased reach and more efficient data collection, come at the expense of more intensive daily monitoring of potential participants. Close collaboration between the research team and skilled information technology (IT) staff is imperative if these strategies are to achieve their purpose.

The specific measures used in each study are indicated in Table 1. The VetChange study encompassed a combination of the three types of strategies conceptualized in this paper, including restricting enrollment to respondents entering the website through approved Web links (URLs), tracking the IP addresses to identify multiple or

inappropriate enrollees and, through re-contacting some subjects, validating questionable responses in areas such as military service history. In addition to a surge of 119 enrollees initially excluded from the study, 10 additional applicants (the total study population was 600) were subsequently excluded from participating after initial data collection. The number of individuals who tried to access the site and were prevented by measures put in place after this incident, such as only accepting enrollment from approved URLs, is not known. It should be emphasized that rarely was one criterion used in isolation; as shown in Figure 1, a combination of indicators was used to decide to exclude individuals. The additional costs for revising the

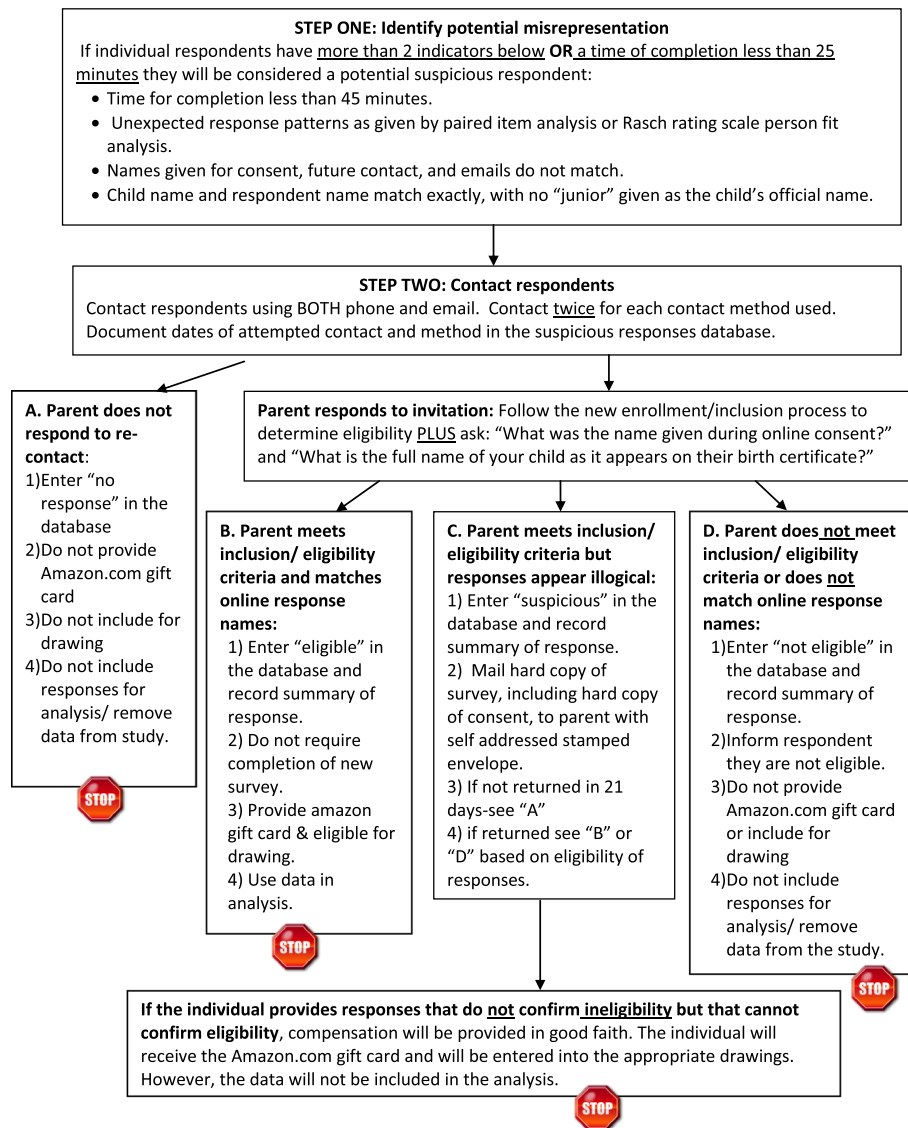


Figure 2. PEDI-CAT-ASD flowchart for identifying participant misrepresentation.

VetChange protocol included approximately 80 hours of programming time and 40 hours of debugging and testing these protocols. Ongoing monitoring by the technical director, research assistant and project director of VetChange added about 35 hours a week. The study was suspended for almost three months while the plan was developed and the approval of two Institutional Review boards were sought, then the new programming and protocol changes were put in place, and previous enrollees reviewed for eligibility using the new protocol. These estimates are based on custom software and protocol, and could vary greatly depending on the circumstance.

The PEDI-CAT-ASD study originally identified 293 responses as suspicious using a combination of strategies illustrated in Figure 2 including: illogical responses to related paired items, shorter than expected time to complete the survey, and inconsistently reported personal information such as names. However, the use of an additional strategy to re-contact all suspicious respondents via phone or email led the team to confirm the eligibility of 30 of those respondents. The revised enrollment strategy adopted by the PEDI-CAT-ASD team that required participants to demonstrate insider knowledge appeared to be effective; some participants seeking enrollment provided obviously illogical or unlikely information in response to emailed screening questions and therefore were not provided access to the web survey. Thus, experience suggests that prospective strategies may be more effective than retrospective strategies. The greatest cost associated with this approach was personnel time; data collection was suspended for over one month while new enrollment procedures were put into place and a team of three graduate research assistants needed three months to determine the eligibility of suspicious respondents.

Confirming eligibility and the validity of respondents is enhanced when procedural, technical and data analytic strategies are used in concert. Investigators need to carefully weigh the pros and cons of the different strategies proposed in this paper based on their target population and the purpose of their research. There is a dearth of research in this area, and the researchers were forced to use common sense and an understanding of their target populations to come up with

the protocols for mitigating this risk. A key consideration is the extent to which gathering identifiable information could reduce individuals' willingness to participate in a study, particularly for sensitive, hard-to-reach populations. For example, the VetChange study did not screen potential participants by telephone because the web provided (1) relative anonymity to veterans who may be concerned about stigma, and (2) the capacity to reach veterans all across the country, or indeed, around the world. The PEDI-CAT-ASD team used this strategy because the study was low risk and completion of the survey already required participants to provide identifiable information.

Researchers planning Web-based studies should also consider the time and resources needed to ensure the validity of the sample at enrollment, as well as the costs of these actions. Finally, researchers should consider local and national research ethics requirements when selecting strategies that increase the burden or risk to research participants, such as strategies that involve the collection of identifiable information. These considerations can help identify those strategies that are most appropriate for a particular project.

While the strategies proposed in this paper may reduce the risk of attack from automated programs and reduce the number of times a human user may feasibly complete Web-based data collection procedures, they do not provide an absolute safeguard against invalid responses. The expanded reach of Web-based survey and intervention research and the accumulating evidence of the efficacy of these approaches means that investigations using this channel of communication will continue to grow. However, as in the past when the telephone was introduced as a medium for data collection, researchers now need to examine how the medium of the Web may change individual's behaviors and how clinical research is best conducted in this newer medium. Investigators must adapt their thinking and their strategies to fit this new, exceptionally promising, research environment.

Endnote

1. The World Wide Web is a system used to access the network of computers and other technology that form the Internet.

References

- Bethell C., Fiorillo J., Lansky D., Hendryx M., Knickman J. (2004) Online consumer surveys as a methodology for assessing the quality of the United States health care system. *Journal of Medical Internet Research*, **6**(1), e2. DOI: 10.2196/jmir.6.1.e2
- Bowen A.M., Daniel C.M., Williams M.L., Baird G.L. (2008) Identifying multiple submissions in Internet research: preserving data integrity. *AIDS and Behavior*, **12**(6), 964–973. DOI: 10.1007/s10461-007-9352-2
- Brief D.J., Rubin A., Enggasser J., Roy M., Lachowicz M., Helmuth E., Rosenbloom D., Benitez D., Hermos J., Keane T.M. (2012) Web-based intervention for returning veterans with risky alcohol use. *Alcoholism, Clinical and Experimental Research*, **36**(Suppl S1), 347A.

- Bromberg J., Wood M.E., Black R.A., Surette D.A., Zacharoff K.L., Chiauzzi E.J. (2011) A randomized trial of a web-based intervention to improve migraine self-management and coping headache. *Headache*, *52*(2), 244–261. DOI: 10.1111/j.1526-4610.2011.02031.x
- Eysenbach G., Wyatt J. (2002) Using the internet for surveys and health research. *Journal of Medical Internet Research*, *4*(2), e13. DOI: 10.2196/jmir.4.2.e13
- Gunn P.P., Fremont A.M., Bottrell M., Shugarman L.R., Galegher J., Bikson T. (2004) The Health Insurance Portability and Accountability Act Privacy Rule: a practical guide for researchers. *Medical Care*, *42*(4), 321–327. DOI: 10.1097/01.mlr.0000119578.94846.f2
- Haley S.M., Coster W.J., Dumas H., Fragala-Pinkham M., Kramer J., Ni P., Tian F., Kao Y., Moed R., Ludlow L. (2011) Accuracy and precision of the Pediatric Evaluation of Disability Inventory Computer Adapted Tests (PEDI-CAT). *Developmental Medicine and Child Neurology*, *53*(12), 1100–1106. DOI: 10.1111/j.1469-8749.2011.04107.x
- Hoge C.W., Castro C.A., Messer S.C., McGurk D., Cotting D.L., Koffman R.L. (2004) Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine*, *351*(1), 13–22. DOI: 10.1056/NEJMoa040603
- Jacobson I.G., Ryan M.A.K., Hooper T.I., Smith T.C., Amoroso P.J., Boyko E.J., Gackstetter G.D., Well T.S., Bell N. (2008) Alcohol use and alcohol-related problems before and after military combat deployment. *Journal of the American Medical Association*, *300*(6), 663–675. DOI: 10.1001/jama.300.6.663
- Konstan J.A., Simon Rosser B.R., Ross M.W., Stanton J., Edwards M.E. (2005) The story of Subject Naught: a cautionary but optimistic tale of internet survey research. *Journal of Computer-Mediated Communication*, *10*(2). DOI: 10.1111/j.1083-6101.2005.tb00248.x
- Koo M., Skinner H. (2005) Challenges of internet recruitment: a case study with disappointing results. *Journal of Medical Internet Research*, *7*(1), e6. DOI: 10.2196/jmir.7.1.e6
- Kramer J., Coster W., Kao Y.C., Snow A., Orsmond G. (2012) A new approach to the measurement of adaptive behavior: development of the PEDI-CAT for children and youth with autism spectrum disorders. *Physical and Occupational Therapy in Pediatrics*, *32*(1), 34–47. DOI: 10.3109/01942638.2011.606260
- Lieberman D.Z. (2008) Evaluation of the stability and validity of participant samples recruited over the internet. *Cyberpsychology, Behavior, & Social Networking*, *11*(6), 743–745. DOI: 10.1089/cpb.2007.0254
- Miller W.R., Rollnick S. (1991) *Motivational Interviewing: Preparing People to Change Addictive Behavior*, New York, Guilford Press.
- Murray E., Khadjesari Z., White I.R., Kalaitzaki E., Godfrey C., McCambridge J., Thompson S.G., Wallace P. (2009) Methodological challenges in online trials. *Journal of Medical and Internet Research*, *11*(2), e9. DOI: 10.2196/jmir.1052
- Riper H., Spek V., Boon B., Conijn B., Kramer J., Martin-Abello K., Smit F. (2011) Effectiveness of e-self-help interventions for curbing adult problem drinking: a meta-analysis. *Journal of Medical Internet Research*, *13*(2), e24. DOI: 10.2196/jmir.1691
- Sinadinovic K., Berman A.H., Hasson D., Wennberg P. (2010) Internet-based assessment and self-monitoring of problematic alcohol and drug use. *Addictive Behaviors*, *35*(5), 464–470. DOI: 10.1016/j.addbeh.2009.12.021
- Wilson P.M., Petticrew M., Calnan M., Nazareth I. (2010) Effects of a financial incentive on health researchers' response to an online survey: a randomized controlled trial. *Journal of Medical Internet Research*, *12*(2), e13. DOI: 10.2196/jmir.1251
- Wright K.B. (2005) Researching internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and Web survey services. *Journal of Computer-Mediated Communication*, *10*(3). DOI: 10.1111/j.1083-6101.2005.tb00259.x