HHS Public Access

Author manuscript

J Law Med Ethics. Author manuscript; available in PMC 2015 December 04.

Published in final edited form as:

J Law Med Ethics. 2015 ; 43(1): 116–133. doi:10.1111/jlme.12200.

Detecting, Preventing, and Responding to "Fraudsters" in Internet Research: Ethics and Tradeoffs

Jennifer E. F. Teitcher,
Research assistant for Dr. Robert Klitzman at Columbia University

Walter O. Bockting, Ph.D.,
Professor of Medical Psychology (in Psychiatry and Nursing) at Columbia University Medical Center and a Research Scientist with the New York State Psychiatric Institute

José A. Bauermeister, M.P.H., Ph.D.,
John G. Searle Assistant Professor of Health Behavior and Health Education (HBHE), and Director of the Center for Sexuality & Health Disparities (SexLab) at the University of Michigan School of Public Health

Chris J. Hoefer,
Research Project Coordinator at the Program in Human Sexuality within the University of Minnesota Medical School

Michael H. Miner, Ph.D., and
Professor and Director of Research at the Program in Human Sexuality, Department of Family Medicine and Community Health at the University of Minnesota Medical School

Robert L. Klitzman, M.D.
Professor of Psychiatry and the Director of the Masters of Bioethics Program at Columbia University

---

Research that recruits and surveys participants online is increasing, but is subject to fraud whereby study respondents — whether eligible or ineligible — participate multiple times. Online Internet research can provide investigators with large sample sizes and is cost efficient.[1] Internet-based research also provides distance between the researchers and participants, allowing the participant to remain confidential and/or anonymous, and thus to respond to questions freely and honestly without worrying about the stigma associated with their answers. However, increasing and recurring instances of fraudulent activity among subjects raise challenges for researchers and Institutional Review Boards (IRBs).[2] The distance from participants, and the potential anonymity and convenience of online research allow for individuals to participate easily more than once, skewing results and the overall quality of the data.

Duplicate entries not only compromise the quality of the research data, but also impact the studies' budgets if not caught before participants' payment — a growing concern with decreasing NIH funding lines. Though reports have begun to explore methods for detecting and preventing fraud,[3] the ethical issues and IRB considerations involved have received little systematic attention. Researchers and IRBs may be unfamiliar with these issues and thus be overly restrictive or lax with Internet research protocols.

In the past, researchers have identified several problematic patterns: 1) eligible individuals who take a study twice, presumably without malicious intent; 2) eligible individuals who take a study repeatedly to receive additional compensation; and 3) ineligible individuals who take a study once or repeatedly to profit from compensation.[4] Despite using methods to detect and prevent fraud, a recent study of transgender and sexual health conducted by Swinburne Romine et al. nonetheless, uncovered more serious fraudulent behavior. Specifically, these researchers found that individuals with IP addresses from China participated in the study by creating fake IP addresses and providing U.S. home addresses that, upon review, were not residential locations.[5] These "fraudsters" may not themselves have been in China, but may have routed these IP addresses through that country in order to avoid detection. Nonetheless, after Swinburne Romine et al. first encountered this problem in 2011–2012, the media has revealed widespread hacking activities from that country.[6] Given these phenomena, we decided to review the literature in light of increasing use of online surveys in academic research and potential fraud by survey participants.

Early studies regarding Internet-based research suggested that multiple submissions were a valid concern but were rare,[7] below 3% in most studies.[8] Reasons given for duplicate responses to surveys were not due to malicious intent, but rather to the respondents' curiosity of how his or her results may change if s/he gave different answers,[9] entertainment (such as a fun game or intellectual challenge), and beliefs that providing more data — even if duplicate — would aid the researchers.[10] Prevention strategies have been recommended — such as providing a link to allow respondents, if they want, to continue to participate without the responses counting toward the data, and simply requesting respondents not to participate more than once.[11] But these strategies do not deter participants with malicious intent from repeatedly entering a study. Reips mentions that high incentives may increase multiple submissions,[12] and Mustanski states that different forms of compensation (direct, lottery, or a donation to charity of choice) may lead to multiple entries, as well as that current prevention strategies are ineffective deterrences,[13] yet they both fall back on the assumption that fraudulent behavior is "extremely rare."[14] Birnbaum writes that providing compensation or a prize can lead to multiple entries for additional compensation or higher chances at winning a prize. He suggests that merely stating that participants will only be compensated once for their participation is a possible solution, but he does not take into account sophisticated and/or malicious "fraudsters."[15]

Ten years ago when these articles were written, incentives were rarely used.[16] But over the past decade, as response rates have decreased, incentives have become more frequent.[17] According to a meta-analysis by Göritz, participants receiving an incentive were 19% more likely to respond and 27% more likely to complete an online survey than those who did not receive an incentive.[18] Additionally, incentives have been shown to boost retention rates in longitudinal studies.[19] However, monetary compensation seems to be increasing both response rates and multiple submissions.[20]

We have found only five sexual health studies that have examined the frequency of multiple submissions. The percentages of entries that were multiple submissions were, respectively, 10% (of which 55% were from the same person),[21] 8% among young men who have sex with men (YMSM),[22] 16% among a sample predominantly of heterosexual young adults,[23]

and approximately 33% of the submissions (of which 51% of multiple submissions were from subjects who participated between 11–67 times).[24] In a recent study conducted by Bauermeister, of the 2,329 YMSM participants who seemed eligible and completed the study, 15% of entries were multiple submissions.[25] Bowen et al. concluded that participants eligible for reimbursement were six times more likely to engage in repeated responses than those who were not offered compensation.[26]

Discussions concerning the ethics of online research often focus on protecting participants' confidentiality to encourage them to trust the researchers.[27] But critical problems can also arise concerning researchers' abilities to trust the participants. Methods of detection and prevention of both duplicate submissions and fraudulent behavior are at times the same, while at other times they are different. Hence, we will discuss both duplicate submissions and fraud below, but highlight issues pertaining to "fraudsters" — those who are ineligible for studies and participate solely for compensation.

Methods for Detecting and Preventing Fraud

In brief, as indicated in Table 1 and described below, several possible methods exist for detecting and preventing fraud, each with pros and cons, and logistical and ethical questions and implications. Researchers can detect and prevent Internet research fraud in four broad ways: at the level of the questionnaire/instrument, the participants' non-questionnaire data and external validation, computer information, and study design. Researchers and IRBs face ethical questions of whether to report "fraudsters" to external authorities, and whether and how to include these methods in an informed consent form.

Questionnaire/Instrument Level

Questions in Survey

Researchers have suspected fraudulent behavior from the inconsistent responses participants provide.[28] For example, Romine et al. excluded participants whose ages did not match birth dates or whose answers to questions about sex, gender, and sexuality were inconsistent (e.g., I was born with a penis/I have had genital reconstructive surgery/I still have a penis; I have had insertive vaginal sex with multiple female partners/None of my partners have vaginas).[29] Researchers can also check that participants have not answered in an "All or Nothing" manner (i.e., answering all 0s or 6s in a survey, or following other patterns of responding [e.g., 1,2,3,4,1,2,3,4]),[30] or skipped large portions of the survey. However, participants may skip questions due to discomfort answering particular questions, and not necessarily due to fraudulent behavior. Nevertheless, examining the types of questions skipped, and how those questions were answered could be helpful in determining discomfort or lack of attention. Similarly, Nosek et al. suggest including choices to survey questions that are not likely to be true.[31] Participants who are not taking the survey seriously may be more likely to select an odd response, though this strategy should be used sparingly, as it may impact the experimental design.

Including questions about social desirability/sociopathy could potentially identify personality traits correlated with providing inaccurate responses.[32] However, tests of such

personality traits may have low, if any, predictability for intentional fraud behavior, as "fraudsters" may not respond to these questions honestly.

Lastly, some entries can be submitted by "bots," instead of individuals. "Bots," short for "robots," are a type of software application that can perform automated tasks over the Internet at a much quicker pace than individuals can. Thus, "bots" can fill out surveys quickly and repeatedly, allowing for the bots' programmers to complete surveys and receive additional compensation quickly. For example, in 1999, Slashdot.com created an online poll asking which was the best graduate school in computer science. Students at Carnegie Mellon and MIT wrote a voting program using "bots" to complete the ballots, resulting in over 21,000 votes for each of these schools, while every other school submitted fewer than 1,000 votes.[33] Similarly, Bauermeister has conducted studies where their own system detected "bots" after flagging rapid re-entries into the system from the same IP address and randomized answer patterns from these entries. As suggested above, researchers can review inconsistent answers (though often needing to do so by hand) to remove submissions from "bots" as well as "fraudsters."

Software for Administering Surveys

Online survey software can be engineered to help prevent Internet fraud. Disabling the back button on the web-browser can prevent "fraudsters" from going back through the survey and revising and resubmitting their responses easily. However, legitimate participants may change their mind about an answer upon greater reflection, and may legitimately want to alter a previous response but would be unable to do so. To solve this issue, the survey could be constructed to allow respondents to review answers periodically. Investigators can also construct the survey to change the order of the questions with each administration, so answers that do not match the questions would be flagged as suspicious.

"Bots" are also commonly prevented by Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA), which frequently requires the user to type in letters and numbers from a distorted image that a computer cannot copy, to ensure that the respondent is indeed a person and not a "bot." This approach, however, may decrease the participation of individuals with low computer literacy, who have visual disabilities (though some CAPTCHA programs offer an audio/sound option), and/or have outdated computer systems that do not work appropriately with CAPTCHA.[34] Additionally, not all CAPTCHA codes are secure, allowing "bots" to invade the system. Some CAPTCHA codes are also used frequently, motivating programmers to create "bots" that can bypass these common CAPTCHAs.[35]

Researchers can check other information beyond what participants provide through the survey's technology. Reviewing the administrative data, also known as paradata, on each subject's behavior can indicate if participants paid attention to the content of the questions or changed answers, potentially shedding light on whether the participant is confused or deliberately being dishonest.[36] A researcher can look at the timestamp, the length of time it took participants to take the study, the ways their mouse moved on the screen, the deletions or changes in their answers, and more. Miner, Bockting and colleagues removed submissions if participants took fewer than 30 minutes to complete the survey, or fewer than

19 minutes to complete the three most important portions of the survey.[37] These cut-offs were based on the overall distribution of respondents' completion times. In each case the cut-off was set at greater than two standard deviations from the mean completion time.

It is important to note, however, that paradata are generally included in relatively costly, private survey programs such as Sawtooth Software,[38] and not accessible through other, "free" survey systems, such as SurveyMonkey.[39] With Sawtooth Software, for example, only researchers have access to participants' information (paradata and other information) as the data may be deposited into the researchers' own data server.[40] Easily accessible online survey tools like SurveyMonkey, on the other hand, may store the information in their data servers and in their terms of agreement may include their right to review participants' data.[41] Hence, the researchers are not the only ones with access to this information — raising concerns regarding the survey's confidentiality when used in these systems. Other public surveys like Qualtrics may store the paradata for free, yet for a fee allow the researchers alone to store and access these data.[42] Consequently, researchers and IRBs must be cautious of which survey service is used to avoid breaches in data safety and security.

Tracking Participants Non-Questionnaire Data

As discussed more fully below, investigators can obtain additional information about participants outside the questionnaire in the form of general information (username, password) or through the computer (IP addresses, cookies). These methods each pose both similar and different ethical issues.

Personal Information

Similar or Same Email, Username, and/or Passwords among Participants— Investigators can check for the same or similar email addresses, usernames, or passwords among participants in the study. Effective cross-referencing may reveal that a username in one entry is similar to an email address in another entry. However, certain common usernames or passwords among participants (e.g., 123456) may not indicate suspicious activity,[43] but may, in fact, be a way for subjects to take part in the study without providing personal information. Removing all such frequent usernames and/or passwords as duplicates from the study could thus result in losing important data. Moreover, "fraudsters" may have multiple, dissimilar, valid email addresses that researchers would not be able to detect. Other means of detection would thus need to be used.

To ensure valid entries are counted while avoiding "fraudsters," researchers can also contact participants about the duplicate entries to clear up any misunderstandings that might have occurred. Bowen et al. deactivated accounts that were identified as multiple submissions and participants were sent a message to contact the project due to a problem with their account, and no subjects asked to be reactivated.[44]

However, contacting participants about "red flags" can dissuade eligible participants, and/or yield a response bias, and risk excluding valid data. Additionally, contacting participants can reveal to "fraudsters" the methods researchers use to detect fraud, thus helping the "fraudsters" to cheat the system more effectively. Researchers may find it advantageous not

to reveal explicitly what was flagged as suspicious, so that fraudulent participants will not know how researchers detected the fraudulent behavior.

Phony Address/Phone Number/Birth Date — External Validation—Checking the name, address, phone number, and age and birth date of participants to determine whether participants provided accurate and unique information can prevent both ineligible participants from taking part in the study and eligible participants from taking part multiple times.[45] Yet participants can provide friends' addresses or phone numbers or a work address or phone number to appear as two different participants, or provide fake addresses and phone numbers.[46] Similarly, in Romine et al., phone numbers were required to complete the registration process (an automated robocall to their number of record then provided a PIN that would allow the participant to continue with the registration process), yet "fraudsters" set up and used temporary Google numbers to circumvent this step.[47]

Additionally, investigators can confirm subjects' eligibility through external validation such as looking up the individual through publicly available search engines, or checking websites such as Facebook or LinkedIn. Bauermeister's study found that using Facebook and MySpace were most helpful in straightening out suspicious data. However, participants did not always have an account for verification, and sometimes privacy restrictions were activated or the profile was associated with a different email address.[48] Researchers can also use Google Earth/Google Maps, whitepages.com, Accurint (which has access to individual's driver's licenses and birthdates, among other records), and NCOA (National Change of Address, a database of changes of address that have been filed) to determine and confirm valid addresses and phone numbers. Unfortunately, eligible participants may be discouraged from taking part in the study if researchers look at information beyond what participants provide for the study. A solution to this issue could be to make providing personal information optional. Bowen et al. requested that participants include their phone numbers for follow-up and retention, yet this request was optional. Bowen and colleagues then used "reverse look-up" on the Internet to determine whether the phone number was valid.[49] Providing optional personal information may be a good way to facilitate participation since eligible subjects can remain anonymous and comfortable. But fraudulent participants may also opt-out of providing information that might identify them as ineligible.

To confirm eligibility, investigators can ask participants to provide a website where they are listed. This request can deter ineligible participants from taking the survey and deter eligible participants from taking the survey more than once, since they cannot assume another identity without proof. However, both eligible and ineligible participants can provide fake information (creating a fake Facebook account, for example) which would "confirm" eligibility yet be completely inaccurate.[50] Moreover, eligible participants may also be deterred from participating.

Publicly-available online information about subjects, if collected without interacting with an individual, would presumably not be considered human subject research, and would not require informed consent. Thus, examining outside sources might appear similar to Humphreys' tearoom trade study, where he collected individuals' license plates without informing them, obtained their names and addresses and contacted them. However,

Humphrey's study was deemed unethical in part because the researcher collected data on individuals without their consent in order to identify and later contact these individuals.[51] Collecting information on individuals separate from what is collected as part of the survey would not be used to gather identifying information that subjects wish to withhold, as was the case with Humphrey's study. But questions nevertheless arise as to whether subjects should be told that such information would be collected. Individuals who make information publicly available on the Internet presumably should not have expectations that the information is private and confidential. Nonetheless, these individuals may mistakenly think that information they provide online is private, when that is not in fact the case (e.g., companies may sell data on online customer purchasing behavior). These individuals may also scroll through and unwittingly accept legal agreements that limit their privacy, but not understand these legal statements. Researchers could also include in the consent form that they will be seeking external validation of subject information.

These strategies raise questions of what is considered personal identifiable information. As Ohm points out, providing date of birth, sex, and zip code — three seemingly vague, innocuous descriptions — can accurately identify a person 87% of the time.[52] Participants might be hesitant to provide potentially identifying information, especially if the questions are sensitive or personal; hence researchers must be careful to ensure participants of the confidentiality of information to encourage eligible subjects to participate.

Computer Information

IP Addresses

Researchers can detect multiple entries through tracking the IP address of the computer used to take the survey. Investigators can see how many times the participant took the survey and whether the participant meets geographic location eligibility (i.e., a survey may only want to study residents of the U.S.; IP addresses would reveal the participants' geographic location). If researchers see an IP address used by many participants, or an IP address from the wrong geographic location, researchers can identify those participants and block those IP addresses, thus preventing participants from taking the study again.[53]

However, problems arise when multiple eligible participants complete the survey from the same computer (e.g., roommates), or a study is being conducted on a large campus where students on the network receive the same IP addresses at different points in time.[54] Some companies that offer internet connectivity at home may also have rotating IP addresses for an area. Consequently, depending on a given day, a home may have two different IP addresses. Without fixed IP addresses, one participant may have different IP addresses, creating problems in determining whether entries are from "fraudsters" or are merely a single individual with an IP address that legitimately rotates. Additionally, eligible participants may be traveling outside a required geographic location while taking the study, in which case a foreign IP address will show up, raising unnecessary red flags. Yet respondents could potentially be asked if the computer they are using is not their usual one, and if so, why. Bauermeister and colleagues, as well as Bowen et al., used other prevention techniques to determine if entries with the same IP address were valid (time completion,

asking how many people use the computer, etc.), and concluded that some were indeed valid entries with duplicate IP addresses.[55]

In addition, IP addresses can be encrypted, scrambled or even faked; "fraudsters" can obtain a U.S. IP address in a different country, preventing researchers from knowing exactly where the participant is, and whether s/he has taken the survey multiple times. Indeed, after further examination in the Romine study of transgender individuals in the U.S., IP addresses were registered to people from China who fit the study's category of "fraudsters." While it was not certain where some of the other "fraudsters" were located, the researchers realized that these individuals were making efforts to produce multiple false records. This realization prompted the researchers to review the demographic information that was provided and determine fake addresses in order to systematically remove these participant records.[56] Similar to paradata, there are costly tracking systems that can determine if someone is re-routing an IP address.

Researchers' examination of IP addresses poses several ethical questions. Researchers may deem a participant's first entry valid, and the subsequent entries as duplicates or fraudulent. Yet, researchers should consider whether the first entry should be deemed valid, as it may not be an eligible participant submitting multiple times, but rather an ineligible "fraudster." By reviewing the results both with and without the first entry, researchers can see how the entries impacted the data.

Additionally, while the United States does not consider IP addresses to be personal information/identification (except for HIPAA purposes),[57] the European Union does.[58] European participants may not want to participate if IP addresses will be tracked, posing problems in conducting research internationally. Researchers may thus be limited in their ability to track IP addresses and face questions of whether to list such tracking in the consent form. Anecdotally, some IRBs have initially been wary of researchers collecting IP addresses, viewing this information as identifying and unnecessary for answering the research questions per se. In a study conducted by Bauermeister, the IRB first discouraged researchers from tracking IP addresses (despite the fact that the U.S. does not consider IP addresses to be personal information/identification). Upon explaining to the IRB the need for this personal data, the IRB agreed but required the researchers to include in the consent form that IP addresses would be tracked. Yet researchers and these committees should consider the possibilities that collection of this information is justified in order to ensure research integrity, and hence scientific and social benefits. A balance of what to track and how to convey this information will be discussed later.

Internet Cookies

Internet cookies are bits of data sent from a website that are stored in an individual user's web browser while the user is visiting that website. Each time the individual user accesses the site, the browser sends the cookie back to the website with information about the user's previous activity. Cookies can also detect if an individual has accessed and/or completed a survey, as well as track the URL to determine from where online participants accessed the survey. If the individual attempts to access the website from the same browser, the cookies

can detect if the individual has completed the survey and can note additional attempts to complete the survey.

Researchers can also provide a link to the survey exclusively in an email, thereby controlling the number of times participants can access the survey, as cookies can show researchers the number of times on which a link was clicked, and investigators can thus detect "fraudsters." Van Gelder et al. suggested recruiting a targeted population via email with a link to a password-protected study in the email.[59] A username would be assigned to each individual who received the email, so that multiple entries could be prevented, and recruiting a targeted population would obclude "fraudsters" from participating.

Relying on cookies presents several challenges. Participants can access the survey from different browsers or delete the cookies stored on their computers, preventing researchers from knowing whether participants have taken the study multiple times. Furthermore, if multiple usernames/emails are provided, cookies would not be able to detect multiple submissions from the same user. Cookies can also reveal and identify someone as a participant in a study; for instance, parents may check the cookies of their teen's computer and see that s/he participated in an LGBT survey. Regarding recruitment via email, Van Gelder et al. suggested that IRBs may be disinclined to recruit participants via individualized email,[60] and/or researchers may not know in advance the email addresses of all the potential participants (e.g., conducting a study on groups that are not easily identified, such as many substance abusers).

Additionally, investigators can enable cookies to be stored on subjects' hard disk on their computers without the subjects' knowledge or consent. Alternatively, some websites issue a pop-up before the user accesses any of the website's contents, noting that by continuing to use the website, the individual agrees to accept cookies on the website. While enabling cookies may assist in detecting "fraudsters" and multiple submissions, informing participants of cookies may discourage eligible subjects from participating.

Similar to IP addresses, enabling cookies may prevent eligible participants who live together or share a computer from participating, if the researcher's software detects that the study has already been conducted from the shared computer. If multiple individuals use the same computer, researchers should decide if cookies should be enabled. If so, the researchers will in effect only be able to include one participant from each shared computer, losing eligible participants.

Tracking Survey URL

Tracking the referring URL and/or searching for the URL online can show researchers if the enrollment site has been posted elsewhere. There are websites that post links to studies for users intending to earn easy money (such as paidsurveysonline.com, onlinejunkie.com, ranksurveys.com and swagbucks.com),[61] so knowing where the URL has been posted allows researchers to see where participants are hearing about the study and researchers can then act accordingly to have the re-posting taken down. This situation in fact occurred in the Romine study: participants notified the researchers, sending screen captures of a chat room where users were mocking and planning to fraud the study.[62] While this method does not

prevent eligible participants from taking the study multiple times, it controls where the study is advertised and can help avoid ineligible participants.

Study Design Level

Elements of the study's design, such as breaking up the consent form, controlling how participants are compensated, and including a face-to-face, online chat or Skype interview as part of the study, can help prevent Internet research fraud.

Informed Consent

Investigators can provide the informed consent form online not as one long document, but instead as separate sections and webpages requiring the participants' consent for each section of the form as it appeared on the screen. The compensation component of the informed consent would be listed at the end. Researchers can have the order of consent options (YES, I agree vs. NO, I don't agree) randomized at each page. This process requires participants to pay more attention to what they are clicking, and creates a longer process to receive the compensation, as opposed to scrolling down quickly through the consent form and "consenting" to the study. These mechanisms can also help reduce "bots" from entering the system. Additionally, not knowing the compensation initially may discourage some "fraudsters" from participating, as they may find that the time is not worth it, given that the amount of compensation is not clear initially, though eligible participants may also be discouraged if the survey is too long and compensation is unknown. While this new structure of the consent form does not detect "fraudsters" or multiple submissions, it can help prevent these situations from initially occurring.

Compensation

Altering the amount, description or type and timing of compensation can also help prevent fraudulent activity. Studies have suggested that lowering incentives would lower fraudulent behavior.[63] Researchers may also be able to de-emphasize the incentive by paying participants less money, or emphasizing the social and community benefits of the study and the costs of fraud. By focusing on the importance of the research and the costs of fraud, some participants may feel less inclined to submit duplicates or falsify results. Bauermeister et al. sent out an email post-survey about the harmful effects of fraudulent behavior in studies to participants suspected of fraudulent behavior and two of the participants apologized.[64] The note stated:

> Dear Participant,
>
> We appreciate your interest and willingness to complete our survey. Unfortunately, we noticed irregularities during data collection. Specifically, a few individuals chose to provide false data, refer ineligible individuals, and/or create multiple entries so that they may receive one or more incentives. We cannot underscore how disappointing this has been for us. Legally, this behavior constitutes fraud.
>
> As public health practitioners, we strive to collect quality and robust data through research that will inform smoking prevention and sex education programs for

> young women. False data diminishes our ability and actually harms the population that we seek to help through science and social services.
>
> We hope that similar events will not occur in future efforts. It is only through the honesty, integrity, and willingness of participants that we can help to contribute to the health of our communities.
>
> If you are receiving this message, you will not receive an incentive; however, if you think that this e-mail is a mistake, please feel free to call us during regular business hours.

However, lowering incentive may also lower participation rates. In addition, some "fraudsters" may not care about the costs of fraud.

Instead of paying all participants in the study, researchers can alternatively provide a lottery for compensation, whereby a smaller number of participants are randomly chosen to receive a larger amount of compensation. This mechanism can also give researchers time to review and identify fraudulent participants before sending out compensation. But "fraudsters" may take the survey multiple times to increase their chances of winning.[65]

Other prevention methods include stating that participants will not be compensated if they are found by the researchers to have submitted duplicate and/or ineligible entries. Researchers can also monitor whether multiple gift certificates are being sent to one location. In Romine's study, the sales representative from giftcertificates.com was able to provide redemption reports that allowed research staff to confirm when a single email address redeemed excessive certificates.[66]

Investigators can ask participants, too, for a mailing address instead of an email address in order to verify legitimate residential location, detering participants from providing phony email addresses. However, providing personal information, which can also link identification to data, might discourage eligible subjects from participating. Rosser and colleagues allowed participants to choose their method of payment to accommodate respondents' comfort levels with anonymity,[67] yet this method would make identifying "fraudsters" more difficult.

In addition, investigators can delay compensation for initial or follow up portions of the studies, giving researchers time to review and determine which participants are fraudulent before sending out compensation. Providing compensation at follow-up portions of a study rather, or proportionally more, than at baseline may increase response and retention rates, and delayed gratification of compensation may also de-incentivize people from answering a survey multiple times. As discussed below, empirical research is needed to examine the potential effectiveness of these approaches.

Including Interview

Researchers can include an interview component to the study via online written, audio, or video chat (e.g., Skype).[68] Face-to-face interviews may be difficult to arrange as participants may be spread out geographically and even across different states or countries. Furthermore, Skype/videochat interviews may be more effective than written chat or audio-only

interviews not only for potentially facilitating and enhancing qualitative interviews, but perhaps also for screening purposes. Such interviews provide another possible means to deter or detect lying, but may also deter eligible individuals from participating, as anonymity may be less pronounced. Moreover, interviews are not a foolproof system as "good liars" may be hard to detect.[69]

Taking Action against "Fraudsters" Outside the Study

Questions arise as to whether researchers and/or IRBs ever need to report cases of fraud to others, and if so, when and to whom. Researchers could, for instance, communicate with other researchers to share information about specific "fraudsters" — i.e., to make a database. Mentioning the possibility of such a database in the informed consent forms might dissuade "fraudsters" but also may dissuade legitimate participants. However, such a database can potentially be useful. On the other hand, "fraudsters" may create unique fictitious online identities for each study, such that the names, emails, and IP addresses they provide may not be repeated among studies. Nonetheless, as more online studies are conducted, the numbers of "fraudsters" will presumably continue to pose problems, and these other methods may be worth studying for effectiveness. Investigators can assess, for instance, how often they detect identical information from "fraudsters" in different studies.

Once researchers identify fraudulent behavior, they face additional decisions. Questions emerge of whether, in extreme circumstances, researchers may want to file a complaint with the Internet Crime Complaint Center (IC3.gov) — a section of the FBI that deals with Internet crimes[70] — and include a warning in the consent form that reporting may occur. Such a warning could powerfully deter fraudulent behavior, but may frighten eligible participants, who may wonder whether researchers may extend government reporting to include other illicit activities (e.g., drug use). Further scholarly discussion and debate is needed to determine what behaviors, if any, might warrant such action (e.g., if individuals went to great lengths to defraud researchers of government funds).

Certificates of confidentiality (CoCs) from the National Institutes of Health (NIH) are intended to help investigators protect data from involuntary disclosure if subpoenaed by a court. Yet the potential usefulness and limitations of CoCs remain unclear since very few have been challenged in court. This certificate does not cover voluntary or intentional disclosure of information by researchers — e.g., in the case of state reporting if a subject divulges child abuse, or reportable communicative diseases, providing these limitations are included in the informed consent.[71] Hence, this certificate may enable researchers to protect data from subpoenas, but allow researchers to divulge information about fraudulent activity if they think that doing so is necessary.

Cross-Cutting Ethical Concerns

Clearly, ethical considerations arise with each of these approaches. These methods differ in the ethical and logistical issues and the specific nature and degree of tradeoffs they present. Yet across individual strategies, researchers and IRBs confront tensions of how to weigh risks and benefits of each approach — how to include in a study means of checking the validity of subjects and their responses without deterring legitimate subjects from

participating. Two underlying ethical principles conflict here: maximizing the scientific and social benefits of research *vs.* respecting the autonomy of subjects (e.g., by decreasing risks of breaches of confidentiality). It is possible that these two goals cannot both be wholly met simultaneously. That is, effective means of reducing "fraudsters" may inevitably deter some potential subjects from enrolling in a study. However, an optimum balance may be possible to achieve. Specifically, vigorous efforts to significantly reduce or eliminate "fraudsters" can ensure the validity of the data, maximizing its scientific and social benefit. The costs may be that some legitimate subjects do not participate, and that researchers thus need to make additional efforts to recruit necessary sample sizes. However, these additional resources appear justified by the result: optimally valid data. Difficult ethical questions emerge, however, as to whether researchers need to disclose to participants all methods the researchers will use to detect and prevent fraud (e.g., collecting IP addresses; searching for subjects online; and enabling cookies on subjects' computers), and if so, to what degree. On the one hand, such disclosure respects subjects' rights to be informed of all relevant aspects of the study, and may deter "fraudsters." However, legitimate participants may then be deterred from participating as well, and such disclosures may alert "fraudsters" to seek strategies to elude these protections — e.g., creating fake Facebook accounts, listing fake names, etc. Creating a fake online presence may seem to require a significant amount of effort for a "fraudster" and thus disincentivize such behavior, but compensation for some studies with multiple stages over a few years can add up to hundreds of dollars. Bockting, Miner, and Hoefer's study provided each subject a total of $180 if participants successfully completed all tasks,[72] and Rosser et al.'s study provided $80 for completing the pretest, intervention and post-test, and an additional $20–25 for completing each follow-up survey.[73] The overseas currency conversion rate can also attract "fraudsters" abroad more than from the U.S., making foreign "fraudsters" think that these efforts are worthwhile.

Researchers and IRBs have three options here to include in the informed consent documents: 1) all information about these methods, 2) no such information, or 3) general and/or oblique references to such methods. Ethically, disclosing all methods respects subjects' rights most. Disclosure of collection of IP addresses can also be important since, as in any study, breaches of confidentiality may occur, posing risks to subjects. Yet, for the reasons discussed above, these disclosures may threaten, too, to decrease the scientific and social benefit of the study. Hence, it appears that these competing pros and cons can best be balanced via an intermediary approach: disclosing the fact that certain measures will be taken, without divulging the details involved (i.e., not mentioning the specifics, such as collection of IP addresses). At the same time, since risks in any study should be minimized, security protections, such as use of firewalls and encryption of data, are essential.

While these various methods share certain underlying ethical tensions, other ethical issues differ somewhat between these approaches. Specifically, these methods vary in the amount of personal information they obtain and/or their degree of invasiveness – i.e., how much they may be considered to impinge on subject autonomy and/or raise additional concerns. Reporting "fraudsters" to external authorities (with such action presented in the informed consent) is most invasive, and though it may be intended to serve as a deterrent, it may be seen as punitive. Conducting a face-to-face Skype interview and collecting IP addresses is

less invasive, but poses more concerns than storing cookies, which in turn poses more concerns than searching for subjects online.

Discussion

Given the increased possibility of fraud in Internet research, strategies in the form of detection and prevention of such duplicate and fake responses are increasingly crucial, yet also pose challenges. Considering the limitations of various prevention methods, it is imperative that researchers use multiple methods to compensate for the limitations of any one approach, and also monitor for duplicate entries by hand throughout the study.[74] A critical eye throughout the study will enhance early detection of duplications and fraud as well as ensure the quality of the data.

Researchers conducting online studies face difficult questions and tradeoffs in seeking to prevent duplicate and fraudulent participation while maintaining and encouraging recruitment of valid subjects. It is vital that both researchers and IRBs remain acutely aware of the phenomena of “fraudsters” described here, and of means of detecting and preventing these practices. Investigators have several possible means of detecting and preventing such ineligible responses — including requesting specific personal information in the study or examining outside sources such as Facebook, Google Earth or whitepages.com. For each study, researchers must decide the strategy that will be useful for preventing research fraud, what information about subjects to request, how to convey these methods and information in the consent form, and to what extent these strategies may have undesired consequences in deterring eligible subjects.

When researchers publish articles reporting data from their studies, they should include information on how much and in what ways they compensated participants for online studies, methods used for detecting and preventing fraud, and the success of these efforts — i.e., report rates of “fraudster” activity among participants to enhance the field’s abilities to avoid these problems. This information will increase understanding of the phenomenon of fraudulent participants, provide a better overview of the study, and ensure data quality.

Researchers and IRBs may also need to consider notifying IRBs, the Office for Human Research Protections (OHRP) and/or funders of fraudulent activity, as these involve unjustified use of grant funds (i.e., paying “fraudsters”), and can affect the integrity of the data and thus the scientific and social benefit of the study. Adverse events per se involve harm to subjects, and research integrity problems generally concern misconduct of investigators. However, “fraudsters” threaten the integrity of the research results. The advantage of such reporting is that IRBs and/or federal agencies (e.g., OHRP, the Office of Research Integrity, or NIH) can then readily track the extent and severity of the problem. The NIH should consider developing an organization similar to the IC3, or interface with the IC3 to assist in tracking and controlling fraudulent research behavior. The IC3 issues periodic alerts regarding new internet crimes and preventions,[75] and the NIH or OHRP could have a similar listing of new “fraudster” strategies and possibly the IP addresses of “fraudsters” and/or the common usernames they use. Clear criteria defining fraudulent behavior that would warrant such action would be imperative. Efforts to gauge the full

nature and extent of “fraudsters” in these ways can enable researchers, IRBs, and others to then work together as best as possible to detect, prevent, and address this problem in ongoing and future studies.

IRBs need to be flexible concerning detection and prevention of fraudulent behavior. However, IRBs are not designed, either in practice or by statute, to protect researchers, but to protect research subjects. The “fraudster” complicates the definition of human subject in the context of IRB review and human subject research. Researchers cannot always plan in advance how participants will take advantage of an online survey. Kraut et al. suggests that IRBs should have an online/computer expert to assist with Internet research in “both online behavior and technology.”[76] Such an expert could explain to the IRB what is appropriate in the specific study at hand, and can keep the IRB up-to-date on technological advances. As both the Internet and “fraudsters” become more sophisticated and online studies are conducted more frequently, it will indeed be important for the IRB to have online/computer experts to draw on to help facilitate and enhance the conduct of online research, and have IRB members make appropriate decisions to prevent fraud while protecting subjects. Different challenges will emerge over time, and in various kinds of studies aimed at different populations. Researchers and IRBs will need to choose specific strategies for detecting and preventing fraud in individual studies in order to optimally balance protecting both research integrity and subjects.

Future research should test how the structure of online studies and the content of consent forms affect eligible subjects participating in studies, as well as how relevant stakeholders (subjects, researchers, research ethicists and others) view these issues and methods discussed here to prevent “fraudsters,” and the “acceptability and efficacy” of such approaches.[77] Similarly, future studies should build on Bowen et al.’s post-hoc finding that compensation (*vs.* no compensation) increases the number of “fraudsters” and the number of entries these “fraudsters” submit.[78] Studies could also examine prospectively how different rates and structures of compensation and informed consent details affect rates of duplications and/or fraud in a study — e.g., how rates of responses and of “fraudsters” vary between longitudinal studies that offer little or no compensation for the completion of initial surveys or offer equal *vs.* increasing amounts of compensation with completion of subsequent surveys over time. Investigators can examine how participants perceive the methods outlined here (e.g., altering amounts, timing, or types of compensation) and what they feel is an appropriate level of compensation, which could offer important insights. Research could examine, for instance, whether appropriate potential subjects would feel less inclined to participate in studies that used each of the methods mentioned here, and if so, how much so. Future studies could also probe how these decisions might vary based on the population, the research, and the questions posed — e.g., whether a method that proves effective in reducing “fraudsters” by, say, 70% may dissuade 1% or 40% of appropriate subjects. Additional challenges arise since a $20 gift card may be an appropriate amount for U.S. participants, but will be worth a lot more in poorer countries, potentially incentivizing “fraudsters” from abroad. Further investigation on how “fraudsters” identify studies (e.g., through websites such as swagbucks.com) would be valuable as well.

The challenges that researchers and IRBs face in conducting Internet-based research is varied and evolving. As the Internet develops, “fraudsters” too, become more sophisticated. Norms and expectations of web privacy are also changing, highlighting ongoing needs to understanding appropriate and effective means of ensuring privacy, while adequately providing informed consent to a study’s procedures. As the Internet continues to evolve along with online research, so, too, should efforts to detect, prevent, and respond to fraud that may occur. Future research and discussions in this area, and reports on evolving patterns of duplication and fraud, are critical in the growing field of online research.

Acknowledgments

The impetus for this article was an expert meeting about Internet research methods in which all of the authors, with the exception of Jennifer Teitcher, participated. This meeting was held at Columbia University Medical Center, December 14–15, 2012, and was supported by a grant from the National Institute on Child Health and Human Development (9R01HD057595-A1; PI Walter O. Bockting, Ph.D.). The authors would also like to thank Kris Abbate, Patricia Contino, and a National Institute of Mental Health center grant to the HIV Center for Clinical and Behavioral Studies at NY State Psychiatric Institute and Columbia University (P30-MH43520; PI Robert H. Remien, Ph.D.).

Biographies

Jennifer E. F. Teitcher received her B.A. in Criminology from the University of Pennsylvania, Philadelphia, PA.

Walter O. Bockting, Ph.D. is Co-Director of the LGBT Health Initiative in the Division of Gender, Sexuality, and Health. He received his undergraduate and doctoral degree from the Vrije Universiteit, Amsterdam, the Netherlands.

José A. Bauermeister, M.P.H. received his M.P.H. and PhD from the University of Michigan, Ann Arbor, MI, and completed post-doctoral training at Columbia University’s HIV Center for Clinical and Behavioral Studies, New York, NY.

Chris J. Hoefer has a B.S. degree in Family Social Science and Queer Theory from the University of Minnesota, Minneapolis, MN.

Michael H. Miner, Ph.D. received his M.A. in Counseling Psychology from Loyola Marymount University, Los Angeles, CA, and his Ph.D. from St. Louis University, St. Louis, MO.

Robert L. Klitzman, M.D. obtained his B.A. from Princeton University, Princeton, NJ and his M.D. from Yale University, New Haven, CT.

References

1. Pequegnat W, Simon Rosser BR, Bowen AM, Bull SS, DiClemente RJ, Bockting WO, Elford J, et al. Conducting Internet-based HIV/STD Prevention Survey Research: Considerations in Design and Evaluation. AIDS and Behavior. 2007; 11(4):505–521. [PubMed: 17053853]
2. Bauermeister JA, Pingel E, Zimmerman M, Couper M, Carballo-Diéguez A, Strecher VJ. Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data. Field Methods. 2012; 24(3):272–291. [PubMed: 23180978] Bauermeister JA, Zimmerman MA, Johns MM,

Glowacki P, Stoddard S, Volz E. Innovative Recruitment Using Online Networks: Lessons Learned from an Online Study of Alcohol and Other Drug Use Utilizing a Web-Based, Respondent-Driven Sampling (webRDS) Strategy. Journal of Studies on Alcohol and Drugs. 2012; 73(5):834. [PubMed: 22846248] Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying Multiple Submissions in Internet Research: Preserving Data Integrity. AIDS and Behavior. 2008; 12(6):964–973. [PubMed: 18240015] Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M. Psychological Research Online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. American Psychologist. 2004; 59(2):105. [PubMed: 14992637]

3. Bartell AL, Spyridakis JH. Managing Risk in Internet-Based Survey Research. Professional Communication Conference (IPCC), 2012 IEEE International. :1–6.Birnbaum MH. Human Research and Data Collection via the Internet. Annual Review of Psychology. 2004; 55:803–832.Gosling SD, Vazire S, Srivastava S, John OP. Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions about Internet Questionnaires. American Psychologist. 2004; 59(2):93–104. [PubMed: 14992636] Konstan JA, Simon Rosser BR, Ross MW, Stanton J, Edwards WM. The Story of Subject Naught: A Cautionary but Optimistic Tale of Internet Survey Research. Journal of Computer Mediated Communication. 2005; 10(2)Miner MH, Bockting WO, Swinburne Romine R, Raman S. Conducting Internet Research with the Transgender Population: Reaching Broad Samples and Collecting Valid Data. Social Science Computer review. 2012; 30(2): 202–211. [PubMed: 24031157] Reips, J.; Musch, U-D. A Brief History of Web Experimenting. In: Birnbaum, MH., editor. Psychological Experiments on the Internet. San Diego: Elsevier; 2000. p. 61-87.Mustanski BS. Getting wired: Exploiting the Internet for the Collection of Valid Sexuality Data. Journal of Sex Research. 2001; 38(4):292–301.Nosek BA, Banaji MR, Greenwald AG. E-Research: Ethics, Security, Design, and Control in Psychological Research on the Internet. Journal of Social Issues. 2002; 58(1):161–176.Reips U-D. Internet-Based Psychological Experimenting Five Dos and Five Don'ts. Social Science Computer Review. 2002; 20(3):241–249.Reips U-D. Standards for Internet-Based Experimenting. Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie). 2002; 49(4):243–256.Riggle EDB, Rostosky SS, Reedy CS. Online Surveys for BGLT Research: Issues and Techniques. Journal of Homosexuality. 2005; 49(2):1–21. see also Bauermeister, "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra*note 2; Bauermeister, "Innovative Recruitment Using Online Networks: Lessons Learned from an Online Study of Alcohol and Other Drug Use Utilizing a Web-Based, Respondent-Driven Sampling (webRDS) Strategy," *supra*note 2; Kraut, *supra*note 2. [PubMed: 16048891]
4. See Bauermeister "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra* note 2.
5. Swinburne Romine R, Miner MH, Gonzalez C, Hoefer C, Bockting WO. The Effects of Fraudulent Respondents on Internet-Based Research with Hard to Reach Populations: Experience from All Gender Health Online. 2014 unpublished manuscript.
6. Sanger DE, Pelroth N. Chinese Hackers Resume Attacks on U.S. Target. New York Times. May 19.2013
7. See Birnbaum, *supra* note 3; Musch and Reips, *supra* note 3; Reips "Internet-Based Psychological Experimenting Five Dos and Five Don'ts," *supra* note 3, Reips, "Standards for Internet-Based Experimenting," *supra* note 3.
8. See Reips "Standards for Internet-Based Experimenting," *supra* note 3.
9. See Bauermeister "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra* note 2; Gosling, *supra* note 3.
10. See Birnbaum, *supra* note 3.
11. *Id*.; Gosling, *supra* note 3.
12. See Reips, "Standards for Internet-Based Experimenting," *supra* note 3.
13. See Mustanski, *supra* note 3.
14. *Id*, at 297.
15. See Birnbaum, *supra* note 3.
16. See Bowen, *supra* note 2.

17. Manzo, AN.; Burke, JM. Increasing Response Rate in Web-Based/Internet Surveys. In: Gideon, L., editor. Handbook of Survey Methodology for the Social Sciences. New York: Springer; 2012. p. 327-343.
18. Göritz AS. Incentives in Web Studies: Methodological Issues and a Review. International Journal of Internet Science. 2006; 1(1):58–70.
19. Neumann, B.; Göritz, AS. The Longitudinal Effect of Incentives on Participation and Data Quality in Online Panels; paper presented at the General Online Research Conference (GOR) 2010; May 27, 2010; *available at* <http://www.websm.org/db/12/13905/Bibliography/The_longitudinal_effect_of_incentives_on_participation_and_data_quality_in_online_panels/> (last visited February 19, 2015).
20. See Bowen, *supra* note 2; Mustanski, *supra* note 3; Riggle, *supra* note 3.
21. See Konstan, *supra* note 3.
22. See Bauermeister, "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra* note 2.
23. See Bauermeister "Innovative Recruitment Using Online Networks: Lessons Learned from an Online Study of Alcohol and Other Drug Use Utilizing a Web-Based, Respondent-Driven Sampling (webRDS) Strategy," *supra* note 2.
24. See Bowen, *supra* note 2.
25. Bauermeister JA, Yeagley E, Meanley S, Pingel ES. Sexting among Young Men Who Have Sex with Men: Results from a National Survey. Journal of Adolescent Health. 2014; 54(5):606–611. [PubMed: 24361235]
26. See Bowen, *supra* note 2.
27. Moreno MA, Fost NC, Christakis DA. Research Ethics in the MySpace Era. Pediatrics. 2008; 121(1):157–161. [PubMed: 18166570]
28. See Bartell, *supra* note 3; Nosek, *supra* note 3.
29. See Swinburne Romine, *supra* note 5.
30. See Miner, *supra* note 3; Riggle, *supra* note 3.
31. See Nosek, *supra* note 3.
32. King MF, Bruner GC. Social Desirability Bias: A Neglected Aspect of Validity Testing. Psychology & Marketing. 2000; 17(2):79–103.Randall DM, Fernandes MF. The Social Desirability Response Bias in Ethics Research. Journal of Business Ethics. 1991; 10(11):805–817.
33. Von Ahn, L.; Blum, M.; Hopper, NJ.; Langford, J. CAPTCHA: Using Hard AI Problems for Security. In: Biham, E., editor. Advances in Cryptology – EUROCRYPT 2003. Berlin: Springer; 2003. p. 294-311.
34. Id.; The Official CAPTCHA Site, available at <at www.captcha.net> (last visited February 19, 2015).
35. The Official CAPTCHA Site, available at <at www.captcha.net> (last visited February 19, 2015).
36. Stieger S, Reips U-D. What are Participants Doing While Filling in an Online Questionnaire: A Paradata Collection Tool and an Empirical Study. Computers in Human Behavior. 2010; 26(6): 1488–1495.
37. See Miner, *supra* note 3.
38. Sawtooth Software. Pricing and Ordering. available at <http://www.sawtoothsoftware.com/products/pricing-ordering (last visited February 19, 2015)
39. SurveyMonkey. Sign up for your FREE Account. available at <https://www.surveymonkey.com/user/sign-up/?fefla=neco&d=1& (last visited February 19, 2015)
40. Sawtooth Software. Software. available at <http://www.sawtooth.com/index.php/sawtooth/index.php/ (last visited February 19, 2015)
41. SurveyMonkey. Privacy Policy. available at <https://www.sur-veymonkey.com/mp/policy/privacy-policy/ (last visited February 19, 2015)
42. Qualtrics, available at <www.Qualtrics.com (last visited February 19, 2015)
43. See Bowen, *supra* note 2.
44. *Id*.

45. *Id*.; see also Miner, *supra* note 3.
46. See Bowen, *supra* note 2.
47. See Swinburne Romine, *supra* note 5.
48. See Bauermeister "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra* note 2.
49. See Bowen, *supra* note 2.
50. See Bauermeister "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra* note 2.
51. Humphreys L. Tearoom Trade. Society. 1970; 7(3):10–25.Humphreys, L. The Sociologist as Voyeur. In: Golden, MP., editor. The Research Experience. Itasca: Peacock; 1976. p. 101-114.
52. Ohm P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review. 2010; 57(6):1701–1777.
53. See Birnbaum, *supra* note 3; Miner, *supra* note 3.
54. See Bauermeister "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra* note 2; Birnbaum, *supra* note 3; Bowen, *supra* note 2.
55. See Bauermeister "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra* note 2; Bowen, *supra* note 2.
56. See Swinburne Romine, *supra* note 5.
57. Buchanan EJ, Aycock S, Dexter D, Dittrich D, Hvizdak E. Computer Science Security Research And Human Subjects: Emerging Considerations For Research Ethics Boards. Journal of Empirical Research on Human Research Ethics. 2011; 6(2):71–83. [PubMed: 21680978]
58. *Id*.Solove, DJ.; Schwartz, PM. Reconciling Personal Information in the United States and European Union. GW Law Faculty Publications & Other Works. 2013. Paper 956, available at <http://scholarship.law.gwu.edu/faculty_publications/956 (last visited February 19, 2015)
59. Van Gelder MMHJ, Bretveld RW, Roeleveld N. Web-Based Questionnaires: The Future in Epidemiology. American Journal of Epidemiology. 2010; 172(11):1292–1298. [PubMed: 20880962]
60. *Id*.
61. See, for example, <www.paidsurveysonline.com; <www.onlinejunkie.com; <www.ranksurveys.com; and <www.swagbucks.com.
62. See Romine, *supra* note 5.
63. See Bartell, *supra* note 3; Reips, "Standards for Internet-Based Experimenting," *supra* note 3.
64. See Bauermeister "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra* note 2.
65. *Id*.; see also Mustanski, *supra* note 3.
66. See Romine, *supra* note 5.
67. Rosser BRS, Gurak L, Horvath KJ, Oakes JM, Konstan J, Danilenko G. The Challenges of Ensuring Participant Consent in Internet-Based Sex Studies: A Case Study of the Men's INTernet Sex (MINTS I and II) Studies. Journal of Computer Mediated Communication. 2009; 14(3):602–626.
68. See Miner, *supra* note 3.
69. Vrij A. Why Professionals Fail to Catch Liars and How They Can Improve. Legal and Criminological Psychology. 2004; 9(2):159–181.
70. Internet Crime Complaint Center (IC3), *available at* <http://www.ic3.gov/default.aspx (last visited February 19, 2015).
71. National Institute of Health (NIH), Office of Extramural Research. Certificate of Confidentiality Kiosk. *available* at <http://grants.nih.gov/grants/policy/coc (last visited February 19, 2015)
72. See Romine, *supra* note 5.
73. Rosser BRS, Oakes JM, Konstan J, Hooper S, Horvath KJ, Danilenko GP, Nygaard KE, Smolenski DJ. Reducing HIV Risk Behavior of MSM through Persuasive Computing: Results of the Men's INTernet Study (MINTS-II). AIDS. 2010; 24(3):2099–2107. [PubMed: 20601853]

74. See Bauermeister, "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra* note 2; Bowen, *supra* note 2; Konstan, *supra* note 3; Miner, *supra* note 3.
75. IC3.gov. Internet Crime Complaint Center's (IC3) Scam Alerts. *available* at <http://www.ic3.gov/media/2014/140321.aspx (last visited February 19, 2015)
76. See Kraut, *supra* note 2, at 115.
77. See Bauermeister, "Data Quality in HIV/AIDS Web-Based Surveys Handling Invalid and Suspicious Data," *supra* note 2, at 289.
78. See Bowen, *supra* note 2.

J Law Med Ethics. Author manuscript; available in PMC 2015 December 04.

Table 1

Methods of Detecting and Preventing Internet Study Duplication and Fraud and Their Implications

Level of Intervention	Type of Intervention	Method of Detection	Method of Prevention	Pros	Cons	Additional Ethical Issues
Questionnaire/Instrument	Questions in Survey	Inconsistent Reponses	Check for proper/consistent answers	• Indicates level of attention • Can detect "bots"	Subjects may skip questions because of discomfort	
			Include same/similar/strange questions throughout study	Indicates level of attention	Can impact experimental design	
			Include questions of social desirability	Possibly help assess personality traits associated with providing inaccurate responses	• Low, if any, predictability • If "fraudster" not paying attention, then questions of social desirability are not helpful	
	Software for Administering Survey		No back button	Subjects can't easily resubmit survey	• Doesn't prevent, just makes fraud more difficult • Eligible participants may want to change answers upon reflection	
			Change order of questions with each administration	• Indicates level of attention • Can detect "bots"		
			CAPTCHA*	Detects "bots"		
			Collect paradata (i.e., subject's behavior, e.g., time stamp, how mouse moved on the screen)	Examines how subject responding to survey	Programs that allow tracking of paradata are costly	Ethical questions of what we can see with paradata – whether to disclose to participants what we can see of their behavior

Level of Intervention	Type of Intervention	Method of Detection	Method of Prevention	Pros	Cons	Additional Ethical Issues
Tracking Non-Questionnaire Data	Personal Information	Similar/same email, username, password between "different" participants	Contact participant about "red flag," and if no response, remove from study	Clears up misunderstandings	• Could yield a response bias • Could deter eligible subjects • Doesn't stop multiple *dissimilar* email, username, password	Needs to balance protecting integrity of data and subject privacy and confidentiality are particularly important
		Inaccurate/fake address & phone numbers	Researchers request to provide phone number/address to get through registration process	Participants need valid number in order to proceed	"Fraudsters" can create temporary phone numbers	
			Check whether person, address, phone number is valid (through Facebook, whitepages.com, etc.)	• Confirms for consistent information • Deters "fraudsters" and multiple submissions	• Can discourage eligible participants from taking part • Subjects do not always have external validation data to ensure eligibility	
			Ask participants for a website where they are listed (e.g., Facebook)	May deter "fraudsters" and multiple submissions	• Can discourage eligible participants from taking part • May encourage "fraudsters" to provide fake information (e.g., creating fake Facebook account)	
Computer Information	IP Addresses	Same IP as another participant	Check whether IP address is the same or if it is encrypted	Can determine how many times participants took survey and whether	• IP addresses may be shared	• If eligible participant takes survey

J Law Med Ethics. Author manuscript; available in PMC 2015 December 04.

Level of Intervention	Type of Intervention	Method of Detection	Method of Prevention	Pros	Cons	Additional Ethical Issues
				participant fulfills location criteria (i.e., living in US)	(e.g., in US, but not living in US)	multiple times, should researchers trust the initial survey?
				participant fulfills location criteria (i.e., living in US)	(e.g., in US, but not living in US)	• Privacy Issue (is an IP address personal information/ identification?)
					• IP addresses can be encrypted/ scram bled/ fake (e.g., using US IP address abroad)	• Should consent forms mention that researchers are tracking/not tracking IP addresses?
					• Programs to check re-routing IP addresses are costly	
			Block IP address if participant is ineligible	Avoids “fraudsters” from participating	Could be dynamic IP address and not ineligible participant	
	Internet Cookies	Cookies detecting completion of study and multiple attempts access study	Enable cookies	Can detect multiple submissions by tracking the progress/completion of study	• “Fraudsters” can disable cookies	• Administering cookies without people’s knowledge.
					• “Fraudsters” can use different browsers	• How should researchers inform participants of cookies without discouraging eligible participants from taking part?
					• Computers may be legitimately shared (e.g., computer labs or roommates)	
					• If “fraudsters” use multiple usernames, cookies would not be able to detect multiple submissions	
					• Can reveal personal information if someone	

J Law Med Ethics. Author manuscript; available in PMC 2015 December 04.

Level of Intervention	Type of Intervention	Method of Detection	Method of Prevention	Pros	Cons	Additional Ethical Issues
					checks cookies on the computer (e.g., parents seeing th checks cookies on the computer (e.g., parents seeing th checks cookies on the computer (e.g., parents seeing th checks cookies on the computer (e.g., parents seeing th checks cookies on the computer (e.g., parents seeing th checks cookies on the computer (e.g., parents seeing th checks cookies on the computer (e.g., parents seeing th checks cookies on the computer (e.g., parents seeing th checks cookies on the computer (e.g., parents seeing th checks cookies on the computer (e.g., parents seeing th	
	Tracking Survey URL	URL posted in unintended locations	Tracking/Googling URL on Internet	Can see if website where URL located is targeting proper audience	Doesn't prevent "frausters" taking study multiple times	
			Provide link in email to website and track referring URL	• Can control who receives email • Can avoid entries submitted that were not from the email	Researchers don't always know the targeted population	
Study Design	Informed Consent		Break up consent online and only provide compensation information at the end of all the forms	• Subjects need to pay attention; and creates longer process to receive compensation, discouraging "fraudsters" • Reduces "bots" from entering the system	May deter eligible pa1rticipants	
	Compensation	Many gift certificates mailed to same address	Mention that subjects will not be compensated if suspect of fraudulent behavior	Avoids paying "fraudsters" yet keeps incentive		
			Only inform participants of their eligibility for the study after survey	• Avoids paying "fraudsters" yet keeps incentive • Gives researchers time to review and determine "fraudsters" • Delayed gratification		

J Law Med Ethics. Author manuscript; available in PMC 2015 December 04.

Level of Intervention	Type of Intervention	Method of Detection	Method of Prevention	Pros	Cons	Additional Ethical Issues
				may deter “fraudsters” may deter “fraudsters”		
			Ask for mailing address (vs. email address) and verify addresses	Deters ineligible participants if researchers have means to verify addresses	May deter eligible participants (because of need to provide personal information)	
			Check if multiple gift certificates are being sent to one location	Can avoid paying participants if suspected of fraudulent behavior yet keeps incentive	Linking identification to data can threaten confidentiality	
			De-incentivize fraud by paying less and/or emphasizing research and the importance of social/community costs of fraud	Potential “fraudsters” may be persuaded not to skew results	• Deters eligible participants if compensation not large enough • “Fraudsters” may not care about importance of research or costs • Gives people the idea to engage in fraudulent behavior	
			Provide lottery for compensation (do not pay every person)	Gives researchers time to review and determine “fraudsters” before compensating	• Lottery may not be enough of incentive for eligible participants to take part • “Fraudsters” may take survey more often to increase chances of winning	
	Including Interview	See whether subject already participated and/or is lying on responses	Audio Interview	• May deter “fraudsters” from participating	• Fraud can be hard to detect (“good liars”)	Needs to balance protecting integrity of data and subject privacy and confidentiality are particularly important

J Law Med Ethics. Author manuscript; available in PMC 2015 December 04.

Level of Intervention	Type of Intervention	Method of Detection	Method of Prevention	Pros	Cons	Additional Ethical Issues
				• Another means to detect lying	• Lose anonymity – could discourage eligible participants	
			Skype/"face-to-face" Interview			
IRBs	IRB Structure		Having an online/computer expert as a member of the IRB	• Has ability to assess the study at hand and find appropriate balance to protect subjects and ensure data quality • Can be up-to-date as technology and "fraudsters" advance to understand how to best prevent "fraudsters"	Does not deter "fraudsters" from taking survey multiple times	
	Have PIs Report Information on "Fraudsters" to IRB		IRBs can follow and monitor to make appropriate decisions for current and future studies	May deter "fraudsters" from participating	• May deter legitimate participants • "Fraudsters" can create new names, emails, IP addresses for each study to avoid detection as a "fraudster"	
Broader Regulatory and other Entities	Reporting Information on "Fraudsters"		PIs create "fraudster" list for other PIs and share information	• PIs can have a list as a reference to easily remove "fraudster" entries • May deter "fraudsters" from participating	"Fraudsters" can create new names, emails, IP addresses for each study to avoid detection as a "fraudster"	Possible harm of individuals are incorrectly classified as "fraudsters" and reported externally? Need to ensure that characterization as "fraudster" is accurate

J Law Med Ethics. Author manuscript; available in PMC 2015 December 04.

Level of Intervention	Type of Intervention	Method of Detection	Method of Prevention	Pros	Cons	Additional Ethical Issues
			Reporting fraudulent behavior to Internet Crime Complaint Center (IC3.gov), OHRP or funders	May deter "fraudsters" from participating	• May deter legitimate participants (some may wonder if researchers will extend reporting to include other illicit activities) • "Fraudsters" can create new names, emails, IP addresses for each study to avoid detection as a "fraudster"	

* Completely Automated Public Turing test to tell Computers and Humans Apart