Guidance on Fraud Prevention Regarding Use of Survey Instruments

Contents

Purpose/Background

The purpose of this guidance is to assist research teams by offering recommended strategies to preserve data integrity and mitigate the risk of financial loss or fraudulent submissions / data contamination while using survey instruments to collect data from research subjects.

Many studies conducted by Johns Hopkins researchers include online surveys that seek to gather data from eligible research participants. Online survey instruments serve as cost efficient and scalable tools which collect data from identifiable or anonymous participants. Data received in surveys can be contaminated by duplicate or fraudulent submissions (malicious or unintentional) that contain inaccurate, irrelevant, or fabricated information. Contamination adversely impacts the quality of data and may cause difficulties for researchers interpreting survey outcomes. Participation in surveys may be remunerated by gift cards or other awards of value. These remunerations can inadvertently incentivize and create opportunities for fraud by participants (eligible or ineligible) who submit data for the sole purpose of receiving payment. These submissions cause financial difficulties for research and undermine the value and reliability of the collected data.

JH Researchers have experienced fraud attempts against previous surveys. As a result, the JH IRBs and compliance offices requested the development of guidance to mitigate this risk.

Responsibility

The level of effort required by researchers to mitigate risk is variable and subjective due to differences among surveys. These include differences in the types of surveys or survey instruments, the scope and diversity of target audiences, the sensitivity of the subject matter, the data being collected, and the type and value of payment. All JHU/JHMI affiliates involved in the collection, storage, or processing of survey data have a responsibility to ensure that data are well managed and that the interests of both Johns Hopkins and survey participants are protected.

This guidance serves as a collection of recommended practices which may be applied in an *a la carte* manner as appropriate. Research teams should use their best judgement in choosing the extent to which the guidance is applied to their unique cases.

In some cases, the value of guidance is derived from applying and evaluating different controls in conjunction with each other. For example, submissions of multiple surveys in a short window of time may be expected behavior, but receipt of these from the same IP address might suggest malevolence.

Adding additional controls in surveys may introduce bias or other undesirable outcomes. Some of the proposed controls may discourage participation by respondents who are less tech savvy, are more privacy conscious, have lower literacy, or have disabilities that are not addressed by careful survey design, resulting in disproportionate representation of these subgroups. This risk of inadvertent bias should be thoughtfully measured against the risk of fraud and/or survey contamination, and in consideration of the target population.

Institutional Review Boards or other JHU/JHMI authorities may encourage review and adoption of practices as appropriate from this guidance.

Definitions

Researcher	**Definition**
Sensitive Data	Any data that contain information that should be reasonably protected from disclosure due to legal, ethical, privacy, or policy considerations. These data include all PHI, PII, student and employee data, proprietary data, data related to socially or culturally defined groups or any data that when disclosed would cause harm including, but not limited to survey respondents, Johns Hopkins Institutions, its affiliates, or partners. JH policies that define types of sensitive data include: https://policies.jhu.edu/doc/fetch.cfm/US6qdR8L
Survey	Survey research is defined as "the collection of information from a sample of individuals through their responses to questions"(Check & Schutt, 2012).
Survey Instrument	Electronic or online platform used to solicit responses to questions from anonymous, random, or targeted participants.
PII	Personally Identifiable Information (PII) is information that can be used to identify an individual, whether on its own or in combination with other personal or identifying information that is linked or linkable to an individual. https://policies.jhu.edu/doc/fetch.cfm/ukMsXgOC
PHI	Protected Health Information, (PHI): individually identifiable health information from medical records or any individual’s health, treatment, or payment information, relating to past, present, or future physical or

	mental condition. https://hpo.johnshopkins.edu/doc/fetch.cfm/8Z5MUkrH
Payment/Incentives/Remuneration	Anything of value that is provided to a survey respondent as incentive, compensation, or other reward that is related to their participation in the survey. Electronic payment is any payment that exists or is transmitted through digital technologies. This may include gift card serial numbers, debit card numbers, PayPal, Venmo, crypto-currency, or other digital payment.
Control	Step taken to mitigate risk. In this document, controls refer to steps taken to mitigate the risk of fraudulent submissions to surveys.
IP Address	A numerical identifier associates an internet connection with its source. IP addresses can be linked to a specific device or network and may be used to identify a person.

Guidelines

The following categories of risk mitigation may be applicable for design, recruitment, and/or implementation of online surveys:

General / Survey Design

- Survey Instrument selection - Selecting the correct survey instrument is a key step in the design and deployment of a secure and effective survey. JHU/JHMI maintain license agreements with vetted enterprise survey platforms. Free internet survey platforms and subscription-based platforms that have not been vetted for information security should not be used for human participant research, PII/PHI or other HIPAA-covered information. This includes SurveyMonkey and Google Forms. Use of these platforms risks security as well as data ownership and control by the P.I. and JHU. Recommended survey platforms include REDCap and Qualtrics.
    - There is a central instance of Qualtrics available at https://uis.jhu.edu/qualtrics/ with "limited" licenses for staff and "licensed" features for students and several divisions and departments. Check with your department or division's IT or research office for local Qualtrics license availability.
    - JH employees are able to request a REDCap project at https://redcap.jhu.edu/. Although a license is not required, there is a monthly fee to receive JH REDCap support (based on the complexity and funding for each individual research project).
- Implement "Security Scan Monitor" (see instructions for this fraud feature in Qualtrics) that prevents email scanners from inadvertently starting a survey session. Similar features may be available in other survey instruments ("Security Scan Monitor," 2022)
- Implement "Expert Review" (feature in Qualtrics) that helps to ensure that surveys collect data of the highest quality. (*Expert Review Functionality*, 2022).
- Remove the back button to prevent retake of survey.
- Perform periodic web-browser searchs for the survey URL to identify if it has been republished.
- Design the survey with random answer choices, skip patterning, and selection of default answers.

Participant Verification

The following strategies can help to positively identify survey respondents to prevent BOTs, duplicate submissions, and submission of data by ineligible participants:

- Remove back button to prevent retake of survey.

- Implement CAPTCHA / ReCAPTCHA or other Bot detection technology. If using Qualtrics, this feature is "Bot Detection" (See instructions in "Bot Detection," 2022). In REDCap, add CAPTCHA in Survey Settings or the REDCaptcha External Module. Note that the effectiveness of CAPTCHA to prevent bot use is variable and subject to the evolution of technology and hacking techniques (See: **Safeguarding REDCap Public Surveys: Tips to Prevent Fraud.docx** (Carey & Babicheva, 2024).
- Implement "RelevantID" (feature Qualtrics, see instructions) that measures likelihood of duplicates based on metadata in responses ("Relevant ID," 2022)
- Request mailing address in the survey to allow for post survey address verification. Note that this control may result in collecting more PII than is necessary and should only be considered for high-risk instances (e.g., posting links to social media, offering high remuneration, etc.).
- Conduct survey in person / over the phone / one-on-one.
- Create a unique survey link for each participant. Application of this is determined by the survey instrument being used.
- Perform IP address verification to assess geographic consistency with expected location of respondent, or to identify duplicate submissions. Note that:
    - Duplicate IP addresses may be a result of separate submissions from participants of the same household, institutions, or across shared computers (library, for example).
    - Geolocation based on IP address may not be consistent for respondents who use VPN's.
    - IP address is a PII identifier recognized by DHS and their collection may be prohibited in some instances.
    - In REDCap a module IP-Encrypt allows collection of encrypted IPs, masked to researchers but allowing comparison of IP addresses to detect duplicates.
- Implement pre-distributed passwords for invited respondents to access the survey for high-risk surveys.
- Review database of respondents to identify duplicate identifiers, if collected.
- Create links to survey that are one-time use to prevent sharing and/or reuse of links.
- Avoid posting links to surveys on publicly accessible sources including social media.
- Add a response limit (in REDCap or Qualtrics) for surveys with known participant caps to prevent extra or automated responses, or monitor and increase incrementally an initial cap to halt automated "survey farm" entries.
- Use a screening survey to determine eligibility of participants as a prerequisite to providing a link to the actual survey.
- Include a question that would be difficult for people outside of the target audience to answer, when possible for the population being surveyed, ideally not easily found by a search engine; for example, something only JHU students might tend to know, or particular to a health condition common to the surveyed group.

Input Validation

The following strategies help to prevent the acceptance of survey submissions that are blatantly fraudulent, inaccurate, irrelevant, duplicate, or fabricated:

- Add fields to the survey that are invisible to humans but would be answered by bots. Note that these items may show up for participants using assistive technologies for people with vision or hearing impairment. (*Hidden Question Traps for Bots*, 2022)
    - In REDCap, add meaningless "honeypot" questions using the @HIDDEN-SURVEY action tag. REDCap prevents these from being "painted" to the screen. Interactively, a human won't see a honeypot questions but a bot can access and try to answer them. If a survey is completed with values in a honeypot field, it was NOT completed by a human (CHOP Research Institute, 2024).

- Perform statistical analysis of responses to identify outliers (Bauermeister et al., 2012; Morrison, 2019)
- Add a question at the end of the questionnaire asking respondents if they looked up any answers online or received assistance to complete the survey.
- Measure time taken to complete survey, or individual sections and then assess to identify anomalies that indicate automation.
- Include at least one question that asks for a text-based qualitative response.
- Include a duplicate question with one requiring a text response, the other multiple choice. Then compare differences.
- Include a simple validation question. For example, "To prove that you are not a robot, select response B."
- Review responses for consistency across questions. For example, age misaligned with year of birth.
- Use a URL shortener (Bitly, REDCap url customized URL) (ITHS, 2016) to obfuscate the link so that bots may not recognize it as a survey. Do not use the word "survey" in the URL.
- Include a duplicate question or similar congruent question and compare responses for inconsistencies, (e.g. DOB and ask later for age in years.)
- Compare all responses to identify duplication across submissions.
- Review responses to identify anomalous time of day submissions that might indicate a submission from an unexpected location based on time zone.
- Review submission times to identify groups of submissions entered within statistically unlikely period or at regular intervals (suggesting automation).
- Include a tiered response question asking for a typed response to a previous multiple-choice question. For example, ask "What was your response to question x (type your response)."

Financial Fraud Identification and Prevention

The following strategies can mitigate the risk of fraud being perpetrated against survey instruments:

- Do not implement mechanisms that automatically issue compensation based on survey completion unless other controls have been implemented to mitigate the risk of financial loss.
- Implement "Fraudulent Detection Fields" (Qualtrics) that embed fraud detection fields into surveys to facilitate post submission fraud analysis (“Adding Fraudulent Detection Fields to the Survey Flow,” 2022) (“Prevent Multiple Submissions,” 2022).
- Include a statement in the informed consent form or description that participants will not be paid if evidence of fraud is found.
  - Example language: *In order to preserve data integrity, this study may include safeguards against fraudulent responses and the collection of data from automated bots. If you agree to participate in this study, you acknowledge that our team may take steps to verify the authenticity of responses. This verification process may be conducted prior to issuing compensation. If responses cannot be verified, compensation may not be issued.*
  - Note: if there is a planned verification process, consider including the timeline for the verification process to the above (i.e., the verification process will be completed within 7 days of survey response).
- Perform post survey assessment to identify if multiple compensation/incentives are requested for the same destination (mailing/email address).
- Carefully consider the value and type of incentive/compensation offered. Higher values also incentivize fraud.
- Structure the compensation as a lottery rather than offering to pay each participant (Habib & Jha, 2021; O’Neil & Penrod, 2001).

- Send incentive/compensation/reward through US Mail rather than to email address if the compensation is of a significant value.
- Inform prospective participants that payments may not be sent immediately to validate submissions and complete fraud detection analysis.

Contacts

Support and guidance regarding data and privacy protection are available throughout JHU/JHMI, depending on the context of the support that is needed. Questions related to data protection for research that is under the oversight of an IRB should be channeled through the appropriate IRB. General data protection questions should be channeled through the appropriate divisional Cybersecurity or IT office. The following are some of these support channels:

- JHU/JHMI IT ~ (410) 516-HELP
- JHU/JHMI Risk Management ~ ITRisk@jhu.edu
- BSPH IT ~ 410-955-3781 or https://my.jhsph.edu/Resources/Help/Pages/Default.aspx
- BSPH Cybersecurity ~ bsph_cybersecurity@jhu.edu
- JH REDCap Administrators Redcap@jhu.edu
- JH Qualtrics site https://uis.jhu.edu/qualtrics/
- ResearchIT@Johns Hopkins University https://researchit.jhu.edu/ for assistance with secure storage and research platforms including SAFER Desktop, SAFESTOR, and DISCOVERY
- JHU Data Services ~ https://dataservices.jhu.edu for data management planning, privacy & sharing guidance
- JHU IRBs
    - JHM IRB: jhmeirb@jhmi.edu
    - JHSPH IRB: BSPH.irboffice@jhu.edu
    - Homewood IRB: hirb@jhu.edu

Approval

This guidance has been reviewed and approved by:

- Darren Lacey, Chief Information Security Officer, Johns Hopkins University
- JHU IRBs (JHM IRB, JHSPH IRB and Homewood IRB)

Related Resources

Bibliography:

The following reference material informed this guidance or may be helpful to researchers who are planning or designing surveys.

Adding Fraudulent Detection Fields to the Survey Flow. (2022). *Qualtrix Fraud Detection*. https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/#SurveyFlow

Bauermeister, J. A., Pingel, E., Zimmerman, M., Couper, M., Carballo-Diéguez, A., & Strecher, V. J. (2012). Data Quality in HIV/AIDS Web-Based Surveys: Handling Invalid and Suspicious Data. *Field Methods*, *24*(3), 272–291. https://doi.org/10.1177/1525822X12443097

Bot Detection. (2022). *Qualtrix Fraud Detection*. https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/#BotDetection

Carey, Scott M., and Viktoriya Babicheva. (2024). Detecting and Preventing BOT and Fraudulent Survey Responses: A Comprehensive Overview. UCONN Health: Clinical Research Center. https://health.uconn.edu/clinical-research-center/wp-content/uploads/sites/50/2024/05/Avoiding-and-Detecting-Bots-and-Fraud_REDCapSurveys.pdf. Alternative source: https://portal.redcap.yale.edu/news/safeguarding-redcap-public-surveys-tips-prevent-fraud.

CHOP Research Institute. (2024), Action Tag Spotlight: Hide Fields with @HIDDEN. Accessed December 3, 2024. https://research.chop.edu/announcements/action-tag-spotlight-hide-fields-with-hidden

ExpertReview Functionality. (2022). https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/quality-iq-functionality/

Habib, D., & Jha, N. (2021). AIM against survey fraud. *JAMIA Open*, *4*(4), ooab099. https://doi.org/10.1093/jamiaopen/ooab099

Hidden question traps for bots. (2022). Qualtrics Community. https://community.qualtrics.com/XMcommunity/discussion/6152/hidden-question-traps-for-bots

ITHS (2016, Oct. 24). How to Create Unique, Custom Survey URLs in REDCap. https://www.iths.org/news/redcap-tip/how-to-create-unique-custom-redcap-survey-urls/.

Morrison, J. (2019, May 30). Assessing Questionnaire Reliability. *Select Statistical Consultants*. https://select-statistics.co.uk/blog/assessing-questionnaire-reliability/

O'Neil, K. M., & Penrod, S. D. (2001). Methodological variables in Web-based research that may affect results: Sample type, monetary incentives, and personal information. *Behavior Research Methods Instruments & Computers*, *33*(2), 226–233. https://doi.org/10.3758/BF03195369

Prevent Multiple Submissions. (2022). *Qualtrix Fraud Detection*. https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/#PreventBallotBoxStuffing

Relevant ID. (2022). *Qualtrix Fraud Detection*. https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/#RelevantID

Security Scan Monitor. (2022). *Qualtrix Fraud Detection*. https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/

Other References

9 Best Methods to Stop Bots From Submitting Forms to Prevent Spam, Fake Users. (n.d.). Retrieved March 18, 2022, from https://www.ipqualityscore.com/articles/view/70/best-methods-to-stop-bots-from-submitting-form-spam

Ballard, A. M., Cardwell, T., & Young, A. M. (2019). Fraud Detection Protocol for Web-Based Research Among Men Who Have Sex With Men: Development and Descriptive Evaluation. *JMIR Public Health and Surveillance*, *5*(1), e12344. https://doi.org/10.2196/12344

Communities, E. (2019, December 11). How to spot and stop fraudsters when collecting research data online. *Evolving Communities*. https://evolvingcommunities.co.uk/how-to-spot-and-stop-fraudsters-in-research/

Lawlor, J., Thomas, C., Guhin, A.T., Kenyon, K., Lerner, M.D., Drahota, A., 2021. Suspicious and fraudulent online survey participation: Introducing the REAL framework. Methodological Innovations 14, 20597991211050467. https://doi.org/10.1177/20597991211050467

Marjanovic, Z., Struthers, C.W., Cribbie, R., Greenglass, E.R., 2014. The Conscientious Responders Scale: A New Tool for Discriminating Between Conscientious and Random Responders. *Sage Open* 4, 2158244014545964. https://doi.org/10.1177/2158244014545964

Johns Hopkins Medicine Institutional Review Board

Bloomberg School of Public Health Institutional Review Board

JHU Homewood Institutional Review Board

Authors and Contributors:

The following subject matter experts contributed to the development of this guidance:

- Michelle Campbell, School of Public Health Office of Information Technology
- Scott Carey, School of Medicine Institute for Clinical and Translational Research
- David Fearon, JHU Data Services, Sheridan Libraries
- Andre Hackman, School of Public Health Biostatistics Center
- Jacky Jennings, School of Medicine Bayview Pediatric Unit
- John Manchester, School of Public Health Cybersecurity
- Sarah Olson, School of Medicine Bayview Pediatric Unit
- Katherine Ready, JH University Information Systems
- Jason Schnell, JH University Information Systems
- Valerie Smothers, School of Medicine Administrative Executive Office
- Rose Weeks, School of Public Health Department of International Health

Example Scenarios / Case Studies

The following examples illustrate the types of problems that this guidance may help to alleviate:

Case 1

Scenario / Problem

We aimed to recruit young adults and health department employees. In-person recruitment during COVID-19 was challenging, so we posted ads on Craigslist, specifying in the flyer our inclusion criteria as well as the reimbursement rate of $20 for each of four activities. The study screening form, consent, and questionnaire forms were hosted on Qualtrics. We had to pause study activities because several participants who initially joined the study were fraudulently representing themselves as being based in the United States by including U.S. zip codes on the screening form.

Solution / Outcome

The Craigslist ads for the study were being viewed and shared outside the U.S. We identified fraudulent participants by inspecting the Qualtrics form downloads for participants' IP address location, looking for VPN use, duplicate IP addresses, or users outside the U.S. (VPN & Proxy Detection Tool https://ip.teoh.io/vpn-detection). Using a generic Google email address (e.g., studystaff@gmail.com) to prevent staff harassment, we informed such users that they were ineligible for the study and could not be compensated. Following this, we established a layered approach to fraud prevention for new online study instruments: screening, consent, and questionnaires. For instance, for our new Qualtrics screening form, we turned on built-in security features including Bot detection, security scan, and RelevantID. We also posted a notice at the top of the screening form stating that people had to be residents in the United States to be eligible for the study. For each potential participant, we reviewed IP addresses to determine not only that the participant was in the U.S. and that the zip code entered on the screening form matched the IP location. For consent forms and questionnaires, we disabled anonymous links and sent each participant unique links.

The applied controls reduced the number of prospective participants who were fraudulently posing as eligible and made it easier to identify ineligible participants.

Risks and Best Practices

While social media platforms (e.g. Facebook, Craigslist) may be utilized for convenience sampling, the use of these platforms for recruitment may increase the risk of fraud. More targeted recruitment vehicles may reduce the risk of fraud while increasing the likelihood of reaching the target population. Recommend separating the link to recruitment survey from the actual survey link. Link to final survey should only be provided to vetted and accepted respondents. Publishing reimbursement value is often necessary but doing so will always invite attempts to defraud the survey. Heightened awareness and diligence are prudent when compensation is published.

Case 2

Scenario / Problem

Dr. Jones' study was investigating sleep patterns of high school students during the COVID shutdowns when all classes were online for extended periods. They created a public survey and offered a $5 Amazon Music gift card to incentivize participation. The survey included CAPTCHA to mitigate the risk of fraud. The goal was to send the survey to 500 randomly selected students from a list of email addresses of students across 3 schools. A link to a generic survey was sent to each targeted student on a Monday afternoon. The survey link was unexpectedly shared with untargeted students at the 3 schools as well as students from other schools. The link went viral among the student population and quickly generated several thousand survey responses. The survey did not indicate that the compensation was limited to invited participants, resulting in liability questions for the survey team.

Solution / Outcome

Because the survey was launched early during a work week, one of the study team members was monitoring activity and identified an anomaly, seeing an unexpected number of survey submissions. The survey link was shut down. While many submissions were received from uninvited students, the use of CAPTCHA likely prevented BOT submissions.

Risks and Best Practices

When compensation is offered, it should be assumed that the link will be shared widely, even among ineligible or untargeted people. Links to surveys are quickly found by BOTS, web crawlers and other automated tools. Plain language indicating who is eligible to receive compensation can reduce unwanted submissions and unambiguously allows teams to withhold payment from ineligible participants. Since this survey had a defined list of target participants and a known maximum number of expected responses, individual links could have been sent to each invitee and a response limit could have been imposed. These features are available in REDCap and Qualtrics.

Case 3

Scenario / Problem

Dr. Smith's study was investigating the impact of COVID on childcare for seasonal immigrants from Mexico. Study details, an offer of a $20 gift card, and a survey link were posted on Facebook pages that are targeted to immigrants to the United States. After a few days, a large spike in responses was observed. Upon closer review, the following anomalies were noted:

- Some responses were coming in the middle of the night.
- Many of the participant names were Slavic and unlikely to be of Hispanic origin.
- Many of the email addresses were atypical (random characters followed by @gmail, for example).
- There were clusters of submission times vs the more typical pattern they saw with surveys that appeared more legitimate.
- Some responses made no sense and were suspected of being created by BOTS.

Solution / Outcome

Study team reached out for help interpreting the results which were evaluated for anomalous response times, unexpected names, atypical email addresses, and clustered response intervals. Submissions deemed fraudulent were discarded and not compensated.

Risks and Best Practices

While social media is a necessary and appropriate platform for reaching some target populations, it should be used with extreme caution. Social media posts should be assumed to also reach unintended recipients, including BOTS, hackers, fraudsters, and people who would submit fake responses for mischievousness or malevolence. This study intended to specifically recruit seasonal immigrants from Mexico but published to social media that targeted all immigrants. More focused targeting may have eliminated the unexpected responses from Slavic populations. CAPTCHA and other participant verification guidance would help to eliminate BOTS and unintended respondents. Plain language indicating who is eligible to receive compensation can reduce unwanted submissions and unambiguously allows teams to withhold payment from ineligible participants.

Case 4

Scenario / Problem

Dr. Lee's study was looking at how inflation has impacted single parents at various levels of education. A public survey link was published to Facebook and Twitter. QR Codes were posted at grocery stores and fast-food restaurants. Compensation/incentive was not offered. The survey included 25 questions and was expected to take 8 minutes. Validity questions were not included. A rudimentary version of CAPTCHA that simply asked for a "Not a robot" response was used. Within 10 days, over 50k responses were received. Concerns were raised about the possibility of BOTS or hacktivists being responsible for many of the responses.

Solution / Outcome

While the study team identified anomalies in response rates and submitted data, they were limited in their ability to evaluate all submissions to determine the validity of the data. Certain combinations of responses were illogical and could be excluded, but questions remained about the remaining data.

Risks and Best Practices

While compensation was not offered and did not incentivize fraudulent responses for this study, there are other factors that influence people to respond to surveys with malicious intent. These include factors as mundane as pranksters having fun, or programmers practicing their skills at BOT automation. CAPTCHA technology is constantly evolving. CAPTCHA features embedded in REDCap or Qualtrics should be used. For surveys on other platforms, CAPTCHA that asks questions related to a displayed image are more difficult for BOTS to bypass. Other participant verification and input validation guidance may have prevented some of the invalid submissions or helped to identify them.