REDCap Survey Security / Integrity:

Detecting and Preventing BOT and Fraudulent Survey Responses

A Comprehensive Overview

Scott M. Carey	Viktoriya Babicheva, MPH
Sr. Systems Engineer	Research Implementation Specialist
REDCap Administrator	REDCap Administrator
Institute for Clinical and Translational Research (ICTR)	Boston College, William F. Connell School of Nursing
Johns Hopkins University, School of Medicine	

REDCap: Survey Security

Detecting and Preventing BOT and Fraudulent Responses

As technology improves, the ability to misuse it also improves. This is particularly true when it involves public survey links and / or opportunities for electronic compensation.

It is incumbent on investigators to understand such threats and to take appropriate steps to mitigate risk, ensuring that the highest quality data is collected and any compensation funds are distributed appropriately.

The goal of this presentation is to educate PI's and study teams on the risks associated with using public survey links and to promote best practices for mitigating those risks.

REDCap: Survey Security

REDCap Public Surveys: A Blessing and a Curse!

Common

- MANY projects use public survey links.

Useful

- They are often the primary method used for study outreach / recruitment.

Vulnerable to Fraud

- They require extra attention and an understanding of the risks and mitigations.

REDCap: Survey Security

Many (most?) PI's and study teams are unfamiliar with the risks that come with using a public survey link.

Let's look at a few situations where this unfamiliarity resulted in major issues.

REDCap: Survey Security

Scenario 1

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 a CAPTCHA to mitigate the risk of fraud. The goal was to send the survey to 500 randomly selected email addresses of students across 3 schools.

A link to a public survey was sent to 500 students on a Monday afternoon. The survey link was unexpectedly shared with other 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 compensation was limited to invited participants, resulting in liability questions.

REDCap: Survey Security
Scenario 2

Dr. Smith's study was investigating the impact of COVID on childcare for *seasonal worker* immigrants from Mexico. The study overview, an offer of a $20 gift card, and a survey link were posted on Facebook pages that targeted seasonal worker immigrants to the United States. The use of a CAPTCHA was not employed. After a few days, a large spike in responses was observed. Upon closer review, the following anomalies were noted:

- Several 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 (e.g., random characters followed by @gmail).
- There were clusters of submission times vs the typical distribution that was more commonly experienced.

REDCap: Survey Security

Scenario 3

Dr. Adam's study was reaching out to MS patients to participate in a 2-year study. A public survey was created to identify potential study candidates. It was, at it's core, an "I'm interested" survey where the study was described and interested individuals could provide their name and contact information. A follow-up call would be made by a study coordinator to determine eligibility. They did not implement a CAPTCHA and the survey was pretty much "wide open". The link was posted on a website and also on social media, as it was seen as a great tool for outreach.

The study went live on a Thursday. Over the following weekend, thousands of responses had been received. Visually, most were junk. The remainder proved difficult to sort out, as many "seemed" real, but when reaching out, the phone numbers were not legitimate. It was difficult to identify valid responses, as the information being collected was so minimal. It was also determined that there were "waves" of incoming data, suggesting bots. Additionally, many of these waves occurred at unexpected times (middle of the night).

REDCap: Survey Security

Keep these 3 scenarios in mind as we delve into this discussion regarding the use of Public Surveys.

Okay, let's jump in!

REDCap: Survey Security

bot (noun)

- A software program that imitates the behavior of a human, as in participating in a chat, or performing automated tasks on the Internet.

fraud (noun)

- A deception practiced in order to induce another to give up possession of property or surrender a right.

REDCap: Survey Security

Motivations for Survey Fraud:

Financial Gain
- Taking advantage of a reward

Disruption / Harm
- Causing damage to the target

The Challenge
- Simply because it can be done

REDCap: Survey Security

Methods of Survey Fraud:

Eligibility Fraud:
- Survey completion by those ineligible to participate.

Multiple Identity Fraud:
- Using multiple identities / email addresses to receive multiple compensations.

Double-Dipping Fraud:
- Simply completing surveys multiple times to receive multiple rewards.

REDCap: Survey Security

The Environments of Fraud:

Public Survey Links:
- Using a public survey link vs. sending unique links to targeted recipients.

Participation Reward / Compensation:
- Studies offering compensation for participation.

REDCap: Survey Security

The fight against survey fraud must happen on multiple fronts.

The Home Front

Bots

Bad Actors

REDCap: Survey Security

The Home Front

REDCap: Survey Security

The Home Front

- Make a plan
- Monitor activity
- Avoid fully automated compensation

REDCap: Survey Security

Make a Plan

"Remember, if you fail to prepare, you are preparing to fail." - Rev. H. K. Williams, 1919

When public surveys are to be used in a project, the PI should be <u>proactive</u> in making plans to limit risk / exposure.

- Determine if public surveys are actually "necessary" to accomplish the goals of the study.
- Use the information provided in this presentation to develop a plan.
- Meet with a REDCap Administrator to review the plan and discuss strengths / weaknesses.

REDCap: Survey Security

Monitor Activity

"You can't manage what you can't measure"

While REDCap can help track surveys with very helpful information related to fraud, someone has to "mind the store".

- Determine what methods of monitoring will be employed.
- Designate that responsibility to the appropriate team member.
- _Schedule_* reviews of the monitoring results (particularly important in the early stages).

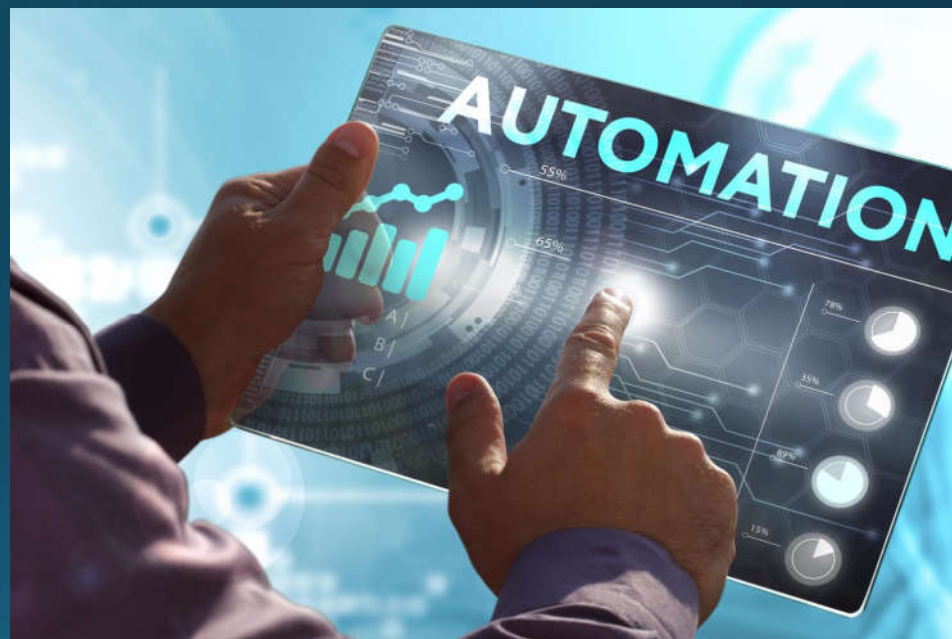* If it's not scheduled, it's unlikely to occur.

REDCap: Survey Security

Avoid "full" Automation

While REDCap has excellent automation tools, automation is NOT always your friend. Particularly, when it involves compensation.

- Build in break-points.
- Put a human in the chain to review / approve compensation.
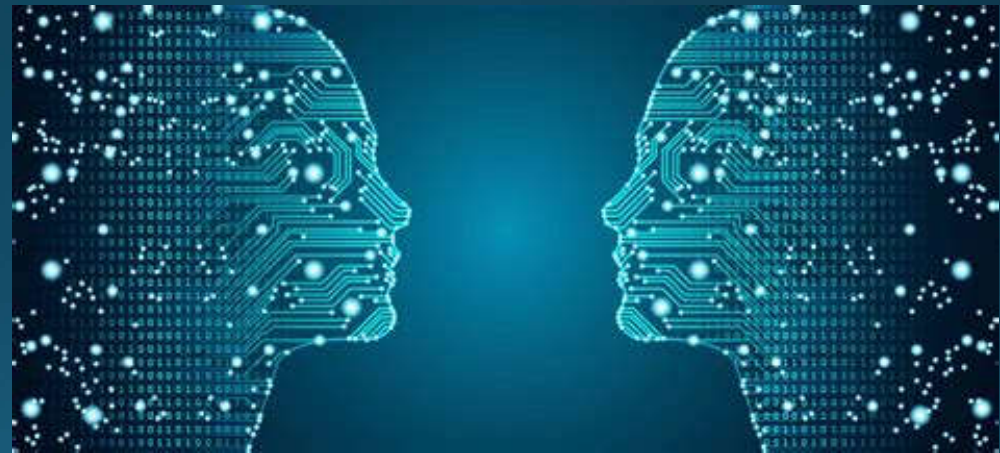
REDCap: Survey Security

Bots

REDCap: Survey Security

Defending Against Bots

- Use CAPTCHA's
- Consider a "Response Limit"
- Implement Challenge questions
- Add "Honeypot" questions
- Include "repetition"
- Paradata collection analysis

REDCap: Survey Security

CAPTCHA

"CAPTCHA, an acronym for 'Completely Automated Public Turing Test to Tell Humans Apart', refers to various authentication methods that validate users as humans, and not bots, by testing users with a challenge that is simple for humans but difficult for machines. CAPTCHAs prevent scammers and spammers from using bots to fill out web forms for malicious purposes."
(source: ibm.com)

In simple terms, a CAPTCHA can help prevent BOTs from accessing public surveys.
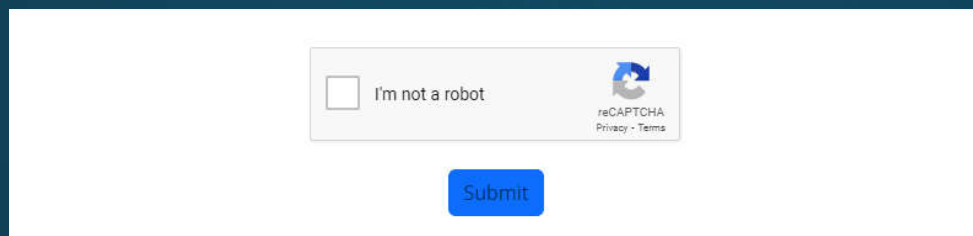
REDCap: Survey Security

CAPTCHA

There are multiple approaches to adding a CAPTCHA to your REDCap public survey. Let's take a look at a few…

IMPORTANT: If you cannot find any of the CAPTCHA options described on the next few slides, reach out to your REDCap administrator to determine if any are currently available or could be made available.

REDCap: Survey Security

The I'm not a robot CAPTCHA is familiar to many. There are two approaches to adding this CAPTCHA to your REDCap public survey.

Method 1: Enable the **I'm not a robot** reCAPTCHA on the Survey Settings page.

Method 2: Use the **REDCaptcha** External Module

*If you cannot find any CAPTCHA options, reach out to your REDCap administrator.

REDCap: Survey Security

Consent Form

Please type in the text exactly as displayed

5bf4kv

Submit

Consent Form

Please solve this math problem:

$9 + 7 - 6$

Submit

NEDCaptcha is an external module that has multiple implementations.

Option 1: Scrambled and masked letters/numbers (provides multiple complexity settings).

Option 2: Answer a basic math question (again, provides multiple complexity settings).

Reminder: If you cannot find any CAPTCHA options, reach out to your REDCap administrator.

JOHNS HOPKINS
UNIVERSITY

REDCap

BC
CONNELL SCHOOL
OF NURSING

REDCap: Survey Security

Response Limit
(survey settings)

- Using the response limiter can help prevent an unexpected wave of bot or invalid human survey responses.
- It can prevent an unanticipated financial liability.
- It can be increased in increments to help avoid a wave of unintended responses.

Survey Access:

⊘ **Response Limit (optional)**
(Maximum number of responses to collect. Prevents respondents from starting the survey after a set number of responses have been collected.) ?

`100`   (e.g., 150) If left blank, the response limit will not be enforced.

Will include `partial and completed responses`

Custom text to display to respondent on survey when limit is reached:

Paragraph — B I U 𝒮 ⟨/⟩ 🖾 @ ☰ ☰ ☰ ☰ ↶ ↷
☰ ☰ ⇤ ⇥ ⊞ ⊞ ⊞ A ✎ Q ⟨⟩ Ix ⤢

Thank you for your interest; however, the survey is closed because the maximum number of responses has been reached.

JOHNS HOPKINS UNIVERSITY

REDCap

BC CONNELL SCHOOL OF NURSING

REDCap: Survey Security

Challenge Questions

By adding a small set of "challenge" questions at the start of a public survey, it may be possible to catch and halt a good portion of bot submissions.

Implementing such questions can contribute to better data.

JOHNS HOPKINS UNIVERSITY

REDCap

BC CONNELL SCHOOL OF NURSING

REDCap: Survey Security

Challenge Questions:

- **Option 1:** Add them to the start of the public survey and use branching logic to hide the remainder of the questions unless all challenge questions are completed correctly (consider using a hidden calculated field to create a "score").

- **Option 2:** The challenge questions can be used to create an "Eligibility" survey (the first survey completed). Based on the answers, limit access to the subsequent survey(s)

  - In the Survey Settings, set the "Conditional Auto-Continue" logic to continue to the next survey ONLY if all of the challenge questions were answered correctly.

REDCap: Survey Security

Honeypot Questions

"*In computer terminology, a honeypot is a computer security mechanism set to detect, deflect, or, in some manner, counteract attempts at unauthorized use of information systems.*"
(source: Wikipedia)

In REDCap, this can be achieved by adding meaningless survey questions using the @HIDDEN-SURVEY action tag. This doesn't remove them from the page. Instead, REDCap prevents them from being "painted" to the screen. Interactively, a human won't see a honeypot questions. However, a bot would still see the questions and try to answer them. If a survey is completed with values in a honeypot field, it was NOT completed by a human.

REDCap: Survey Security

"Honeypot" Questions

Adding a couple of honeypot questions is simple.

- Add a radio button field.
- Add a textbox field.
- Add the @HIDDEN-SURVEY action tag.
- Setup an alert / notification to send an email to the appropriate person if / when honeypot questions are answered.
- Prevent continuing if these fields are valued.
- Review and determine appropriate next steps.

REDCap: Survey Security

Repetition

Asking a couple of similar / identical questions using different sentence structures can help identify BOT responses.

Adding repetition questions is also very simple in REDCap.

In the example below, the question is less obvious to a BOT, but easy for a human to understand and answer.

Please enter this HIGHLIGHTED portion of the date provided above (mmdd`YYYY`)

Note: this is being asked to help identify fraudulant automated responses.

REDCap: Survey Security

Paradata

"The paradata of a data set or survey are data about the process by which the data were collected.[1][2] Paradata of a survey are usually "administrative data about the survey."[3]

Example paradata topics about a survey include the times of day interviews were conducted, how long the interviews took, how many times there were contacts with each interviewee or attempts to contact the interviewee, the reluctance of the interviewee, and the mode of communication (such as phone, Web, email, or in person).[4] Thus there are paradata about each observation in the survey. These attributes affect the costs and management of a survey, the findings of a survey, evaluations of interviewers, and inferences one might make about non-respondents.

Paradata information can be used to help achieve the goals of a survey. For example, early responses may be mainly from one type of respondent, and the collectors knowing this can focus on reaching the other types so the survey has good coverage of the intended population. Thus survey efforts can be dynamically responsive to the paradata.[5]

In principle a survey's metadata includes its paradata."
(source: Wikipedia)

In REDCap, this can be achieved by using survey metadata to analyze various details related to survey completion.

REDCap: Survey Security

REDCap Paradata

REDCap captures certain metadata related to survey completion. Exporting and analyzing these data can help identify potential BOT submissions. These data include:

- Survey Start Time
  - Look for clusters of surveys occurring at the same or similar times.
  - Look for patterns (is there a bump in submissions at a similar time every day/night?).

- Survey Duration Time
  - Enter a few test surveys to create a range of "reasonable" times for completion.
  - Are surveys being fully completed in an unrealistic timeframe?

- IP Address
  - Outside of eConsent, collecting IP addresses is generally considered inappropriate (PHI)
  - However, JHU has a module* that allows for the collection of an *__encrypted__* IP address.
  - The encrypted IP addresses allow the study team to identify multiple submissions from the same location, without exposing the actual IP address.
  - If necessary, a REDCap administrator can decrypt (with appropriate IRB authorization).
  - This approach has been approved by JH Legal.

* This tool (@IP-ENCRYPT) is available to other institutions as an External Module.

REDCap: Survey Security

Bad Actors

REDCap: Survey Security

Bad Actors

Let's look again at some motivations for fraud…

- Financial Pressures
- Rationalization
- Opportunity

NOTE: Sometimes, as in scenario 1 (students sharing a public link), it's not about people being nefarious. Sometimes, it's the result of a project plan that just wasn't thought through.

What causes someone to commit workplace fraud?

The Fraud Triangle

Fraud - such as the intentional misuse of funds - is often the result of pressure, rationalization and opportunity.

PRESSURE
I can't pay my bills!

RATIONALIZATION
I'm just borrowing the money…

OPPORTUNITY
I got away with it once, now I know how.

REDCap: Survey Security

Defending Against Bad Actors (use multiple techniques)

- Highlight Surveillance
- Disclose Consequences
- Add a Response Limit
- Repeat Questions
- Cross-Reference Questions
- Include Challenge Questions
- Include Open Ended Questions
- Conscientious Responders Scale
- Targeted Survey Distribution
- Use Smart Incentives
- Paradata Collection / Analysis

REDCap: Survey Security

Highlight Surveillance And Disclose Possible Consequences

- Inform survey takers that activity is being monitored.
- Hint a possible consequences.
- Reach out to your legal department for appropriate wording.

COVID Outcomes Survey

NOTE: Responses are monitored for fraud. Submitted surveys deemed to be questionable or fraudulent will not be compensated. Attempts to acquire inappropriate compensation through survey fraud may be prosecuted.

1) First Name

2) Last Name

3) Date of Birth   M-D-Y

JOHNS HOPKINS
UNIVERSITY

REDCap

BC
CONNELL SCHOOL OF NURSING

REDCap: Survey Security

Response Limit
(survey settings)

- Avoid a fraudulent "survey farm" attack by adding a response limit.
- It can prevent an unanticipated financial liability.
- Increase incrementally, as needed.

Survey Access:

🚫 **Response Limit (optional)**
(Maximum number of responses to collect. Prevents respondents from starting the survey after a set number of responses have been collected.) [?]

`100`   (e.g., 150)   If left blank, the response limit will not be enforced.

Will include `partial and completed responses`

Custom text to display to respondent on survey when limit is reached:

Paragraph   —   **B** *I* U 𝒮 </>   ▤ 📎   ≡ ≡ ≡ ≡   ↺ ↻

☰ ☷   ⇤ ⇥   ▦ ▦ ▦   A ⌄ ✎ ⌄   Q ‹› Iₓ   ⛶

Thank you for your interest; however, the survey is closed because the maximum number of responses has been reached.

REDCap: Survey Security

Repeat Questions

- Fraudulent respondents are less likely to pay a great deal of attention to the actual questions.
- Ask similar questions using different wording on different survey pages, making it difficult for the respondent to recall what was entered previously.
- Flag records with inconsistent answers.

REPEAT IT

REDCap: Survey Security

Cross-Reference Questions

- Ask different, but related questions that should be consistent / congruent.
- Examples:
  - Ask for DOB and later ask for their age in years.
  - Consider adding past diagnosis and treatment questions and review / evaluate for validity. For example, a breast cancer patient having certain illogical treatment combinations would be a red flag!

Consistency Is

REDCap: Survey Security

Challenge Questions

- Ask questions that verify eligibility.
- Examples:
  - Age eligibility.
  - Able to participate through the entire study.
  - Specific health questions that reveal study eligibility.
- Use the Conditional Auto-Continue feature to allow only eligible respondents to continue to the next survey.

NOTE: Using challenge questions is a good practice and should be considered for most studies where a public survey is used for recruitment.
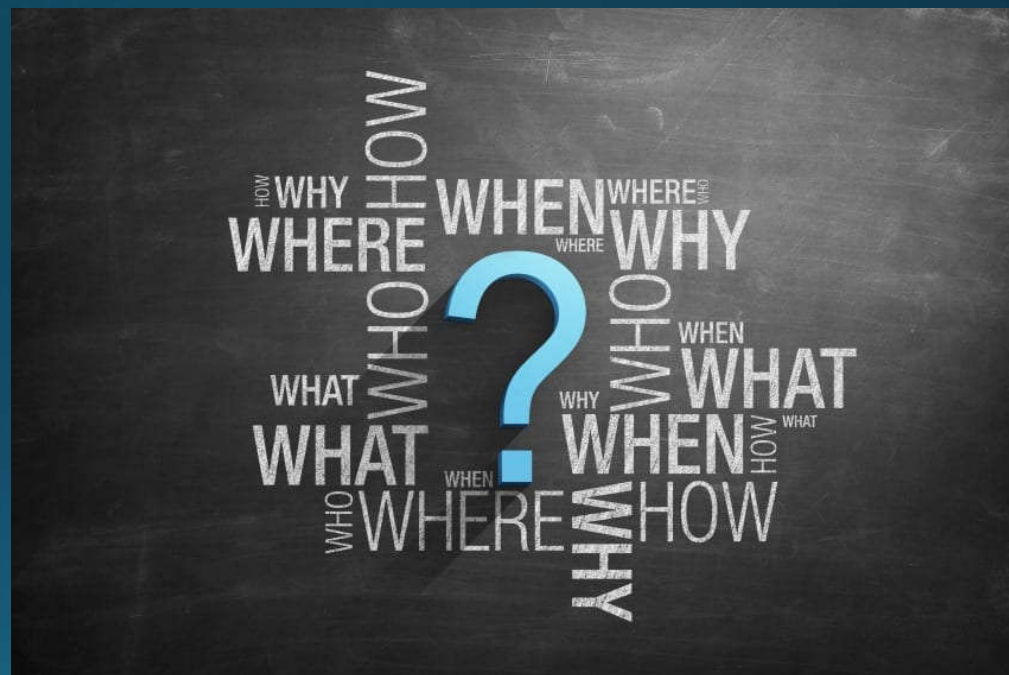
REDCap: Survey Security

Ask Open-Ended Questions

- Fraudulent respondents are less likely to take the time necessary to answer open ended questions.
- Asking questions specific to the study make this even more difficult for a fraudster to "game".

Be creative…
- What activities lessen your symptoms?
- Describe your first symptoms?
- What have others noticed about your condition?

REDCap: Survey Security

Conscientious Responders Scale
(more here)

The Conscientious Responders Scale attempts to determine if the respondent is being thoughtful about their responses vs. just blowing through their answers without much thought to their responses.

To see an example, click here.

REDCap: Survey Security

Targeted Survey Distribution

- Avoid social media. It is the most likely to attract fraud / bots.
- If social media is the only option, use a targeted "group" vs. a public posting.
- Consider using a list-serv address to target a very focused group of individuals.
- If a list of target recipients is available, use REDCap's _Survey Distribution Tools_*.
- Alternatively, create records for each targeted individual and _use Automated Survey Invitations_* (ASI's).

* These options generate survey links unique to the individual, making it impossible for them to be used multiple times.
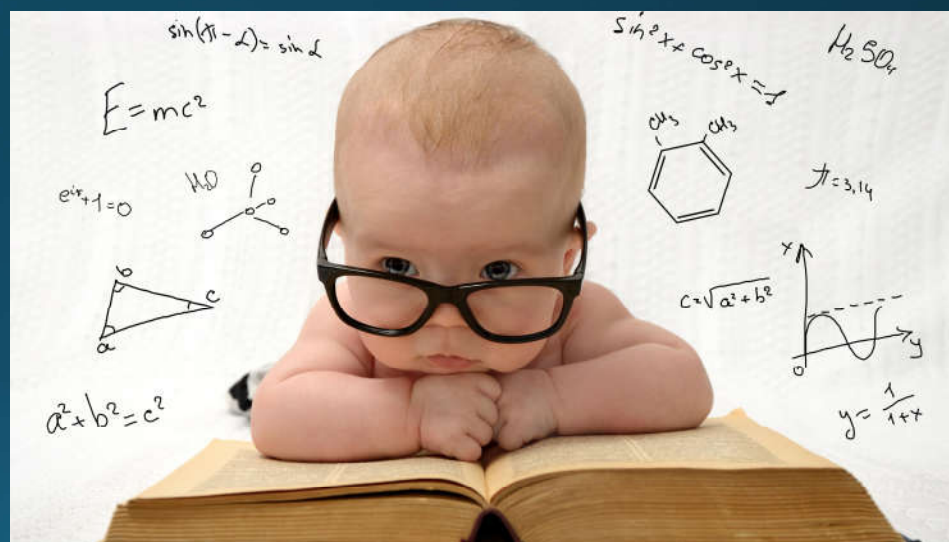
REDCap: Survey Security

Use Smart Incentives

Fraudsters generally look for a quick hit. They are typically not committed enough to jump through hoops.

Again, be creative...
- Wait a day or two and then send a follow-up "compensation claim" survey link to them.
  - Completing it documents "intent".
  - It also requires time and remembering what it was about. It becomes "too complicated" for many fraudsters.
- Make it time dependent or dependent on completing a certain number of tasks or completion of the study.
- Avoid rewarding laziness. If compensation is being given, it's appropriate to have it be "earned".

REDCap: Survey Security

REDCap Paradata

As discussed previously, REDCap captures certain metadata related to survey completion. Exporting and analyzing these data can help identify potential fraudulent submissions.

REDCap: Survey Security

Controlled Access ➜ Controlled Continuance

If you remember nothing else, remember this...

Public survey links may let someone onto your front porch, but that doesn't mean you have to let them into your home!

Treat the initial (public) survey as your front porch. Use the techniques discussed in this presentation to screen who is permitted to continue.

REDCap: Survey Security

Let's take another look at our three opening scenarios

REDCap: Survey Security

Scenario 1 (students sharing a public survey link with their peers):

The Good:
- They used a CAPTCHA.
- Attempted to use a "targeted" pool of study participants.

The Bad:
- Didn't consider the target population or the possibility the link would be shared among peers.
  - This should have been anticipated!
- Didn't add any language indicating that participation in the survey was restricted to the target recipients.
- They used a public survey link when it wasn't necessary.

Mitigation:
- Add a "Response Limit" to limit the total number of responses that can be submitted (survey settings).
- Rather than using a public survey link, could have used REDCap's "participant list" (sends a unique link to each student).
- Include language indicating only those who originally received the link are eligible.

JOHNS HOPKINS UNIVERSITY

REDCap

BC CONNELL SCHOOL OF NURSING

REDCap: Survey Security

Scenario 2 (survey targeting Hispanic migrate workers):

The Good:
- They knew their audience and could look at the data and tell something was not right.
- Attempted to use a "targeted" pool of study participants.

The Bad:
- No CAPTCHA was used.
- Didn't use any tools to track IP addresses, which may have been helpful.

Mitigation:
- Add a CAPTCHA!
- Add a "Response Limit" to limit the total number of responses that can be submitted (survey settings).
- Include language indicating who is eligible to participate and eligible for compensation.
- Capturing an encrypted IP address would have been helpful in identifying possible fraud (@IP-ENCRYPT).
- Other fraud prevention methods described herein.

REDCap: Survey Security

Scenario 3 (wide open survey for MA patients with thousands of responses)

The Good:
- The study team was REALLY nice! (other than that, there isn't much that was "good" about this situation)

The Bad:
- No CAPTCHA
- Using social media is begging for problems.
- Very few question were included in the survey, making it VERY easy for a BOT to complete.
- No way to determine if there were blocks of responses from the same location.

Mitigation:
- Add a "Response Limit" to limit the total number of responses that can be submitted (survey settings).
- Use the encrypted IP tracking tool (@IP-ENCRYPT).
- Use honeypot questions.
- Use challenge questions.
- Use a CAPTCHA!
- Other methods described herein.

REDCap: Survey Security

So... what tools should I use for my project?

Honestly, it depends...

- How is survey being used?
- Who is the target audience?
- What's at stake?
- What are the available resources for evaluating survey results for fraud?
- Who's minding the store?
- Would it help to talk to a REDCap Administrator?

As was stated at the beginning. The PI is responsible for ensuring the quality of the data and the appropriate distribution of participation compensation. It's important that the PI and study team understand what is at stake and what steps can be done to prevent survey fraud.

JOHNS HOPKINS
UNIVERSITY

REDCap

BC
CONNELL SCHOOL
OF NURSING

REDCap: Survey Security

Being proactive requires MUCH less effort than being reactive!

Naïveté is not an acceptable excuse. It is the PI's **_responsibility_** to anticipate fraud / bots and to actively (_pro_actively) take measures to mitigate those risks.

The information provided in this presentation will greatly reduce the risks associated with using public surveys. But the time to consider the risks and mitigation is **_PRIOR_** to releasing a public survey.

REDCap: Survey Security

Additional Resources

- 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(3). https://doi.org/10.1177/2158244014545964

- Babicheva, Viktoriya (2023). **Preventing Bots and Fraudulent Participation in REDCap Surveys. REDCapCon 2023.** REDCapCon 2023 Slides, Videos, & Notes (vanderbilt.edu). (access limited to REDCap Administrators)

- 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(3). https://doi.org/10.1177/20597991211050467

REDCap: Survey Security

Contributors

With <u>much</u> gratitude, the following groups and individuals warrant recognition for their valuable contributions to this presentation.

- JHU Working Group: "JHU IRB Guidance on Fraud Prevention for eSurveys"
- JHU study team members (323) completing a "Survey Usage" questionnaire
- JHU REDCap Advisory Committee
- Viktoriya Babicheva: REDCapCon 2023 Presentation
- Reviewers:
  - Leila Deering – Marshfield Clinic – Marshfield, WI
  - Greg Neils – Mass General Brigham – Boston, MA

REDCap: Survey Security

Questions?

(time permitting)

Contact Information:
- Scott Carey (Johns Hopkins University)
  - scarey@jhu.edu
- Viktoriya Babicheva (Boston College – Connell School of Nursing)
  - babichev@bc.edu