REDCap

Research Electronic Data Capture

# Survey Security

## Detecting and Preventing BOT and Fraudulent Survey Responses

THE UNIVERSITY OF CHICAGO MEDICINE & BIOLOGICAL SCIENCES

CENTER FOR RESEARCH INFORMATICS

Version 14.5.15 - © 2024 Vanderbilt University

## INTRODUCTION

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 document is to educate Primary Investigators and study teams on the risks associated with using public survey links and to promote best practices for mitigating those risks. And the time to consider the risks and mitigation is PRIOR to releasing a public survey. Questions that should be asked when creating your survey are:

- 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?
- Would it help to talk to a REDCap Administrator?

## DEFINITIONS

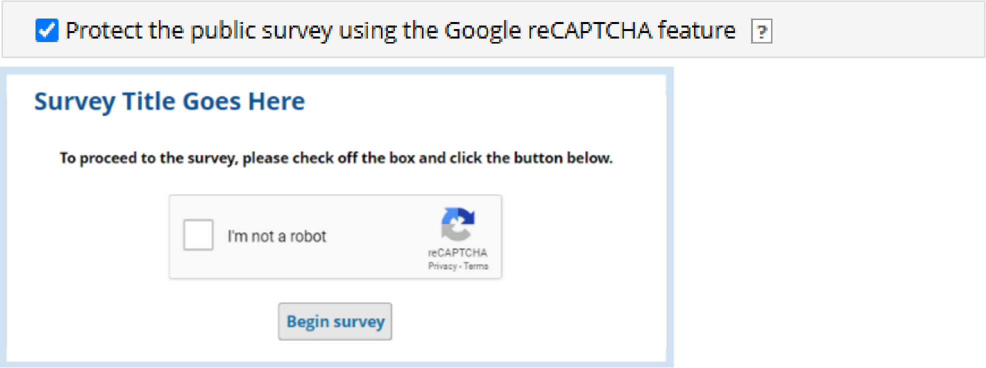
**bad actor** (*noun*) - alleged perpetrators of cyberattacks and other malicious online activity (i.e. bots, hackers, trolls, cybercriminals).
**bot** (*noun*) - A software program that imitates the behavior of a human, as in participating in a chat, or performing automated tasks on the Internet.
**fraud** (*noun*) - A deception practice to induce another to give up possession of property or surrender a right.
**cybercrime** (*noun*) – Fraud that takes place over the internet or on email, including crimes like identity theft, phishing, and other hacking activities designed to scam people out of money.

## DISCUSSION

### 1. Are there ways to ensure that participants responding to a survey are not bots?

Yes, here are several ways to help curb bots:

- Enable the Google reCAPTCHA on the Survey Distribution Tools page. The Google reCAPTCHA feature can be enabled to help protect your public surveys from abuse from 'bots', which are automated software programs that might enter trash data into your survey. A 'captcha' is a turing test to tell humans and bots apart. It is easy for humans to solve, but hard for bots and other malicious software to figure out. By enabling Google reCAPTCHA on your public survey, you can block automated software while helping welcome your survey participants to begin your survey with ease. Below is an example screenshot of what the reCAPTCHA might look like on your survey.

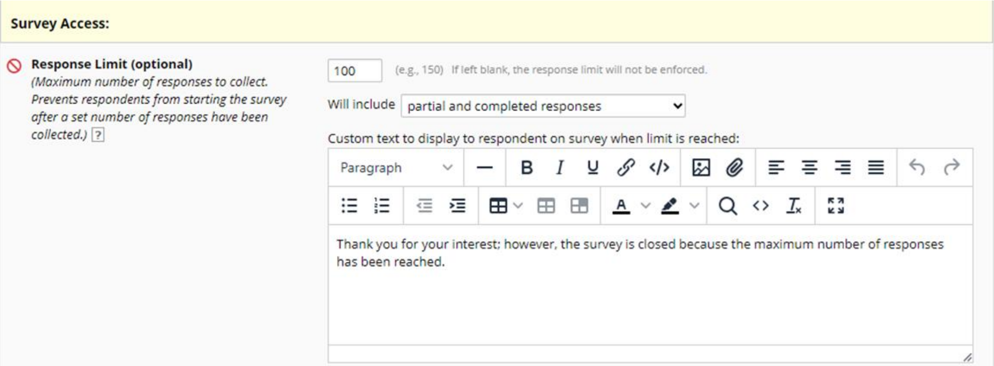


☑ Protect the public survey using the Google reCAPTCHA feature

Survey Title Goes Here

To proceed to the survey, please check off the box and click the button below.

- Consider a “Response Limit”. 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.)* ?

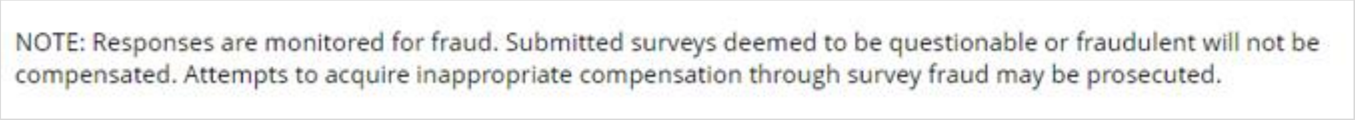
100 (e.g., 150) If left blank, the response limit will not be enforced.

Will include: partial and completed responses

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

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

- Inform survey takers that activity is being monitored and list possible consequences.

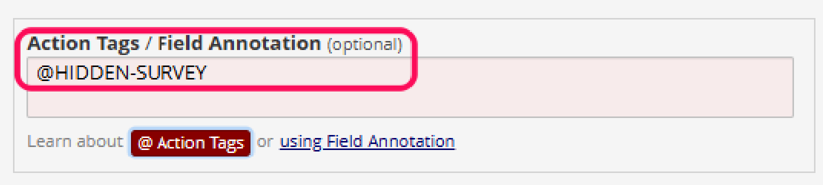

NOTE: Responses are monitored for fraud. Submitted surveys deemed to be questionable or fraudulent will not be compensated. Attempts to acquire inappropriate compensation through survey fraud may be prosecuted.

- When compensation is involved, avoid advertising survey on social media. It is the most likely to attract bots, bad actors, or fraud.

- When compensation is involved, build in breakpoints; put in a step for human review at intake and compensation approval.

- For e-consent, create an intake survey to screen candidates. Review all intake surveys and only send the consent survey to eligible participants.

- Add hidden "honeypot" questions to the survey. If a question is hidden, REDCap prevents it from being displayed on the survey screen. Interactively, a human won’t see a honeypot questions. However, a bot program 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. Questions can be hidden on surveys with the use of the @HIDDEN-SURVEY action tag in the field’s Action Tab box:

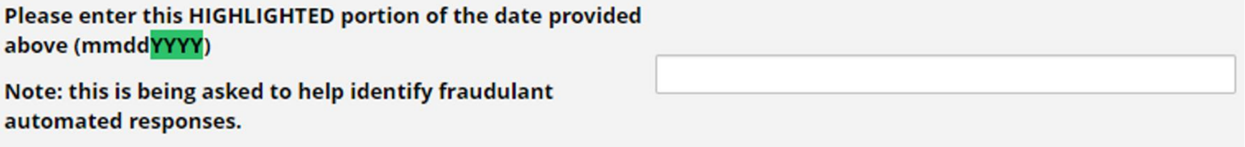

**Action Tags / Field Annotation** (optional)

@HIDDEN-SURVEY

Learn about @ Action Tags or using Field Annotation

*As an added measure, set up an Alert/Notification to send an immediate email to the study team when honeypot questions are answered so the team can take remediative action.

- Add "visual challenge" questions throughout the survey. A bot cannot "see" or "read" this question to answer it correctly, but humans can. Disregard any surveys that have an incorrect answer. For example:


**Please enter this HIGHLIGHTED portion of the date provided above (mmddYYYY)**

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

- Ask different, but related questions that should be consistent/congruent. Flag records with inconsistent answers. For example, ask for their birthdate and later ask for their age in years.

- Use “smart incentives”. Fraudsters generally look for a quick hit. They are typically not committed enough to jump through hoops. 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.

### 2. Is there a way to limit IP addresses to US residents?

Unfortunately, there is no REDCap feature that will block IP addresses. In general, network administrators can use hardware or software tools to prevent access from certain IP addresses, but you would need to define the IP addresses being blocked. However, the use of proxy IP addresses, dynamic IP addresses which change frequently, and other bypass techniques make it difficult to completely block IP addresses. In short, IP address blocking is a complex and almost impossible task.

### 3. For existing survey responses, is there a way to determine whether the participants were real or bots?

There are websites where you can check the IP addresses, location, and hostname manually. There are also online bot checkers that perform a bot detection test on any IP address. However, because of the reasons in question 2 above, and because today’s bad bot traffic is almost indistinguishable from legitimate human traffic, this is also not a foolproof process.

## QUESTIONS?

For further help on this topic for your specific REDCap survey project(s), please contact CRI REDCap Support at redcap@uchicago.edu.

## ACKNOWLEDGEMENTS

The following individuals warrant recognition for their content used in this document.

Scott M. Carey
Senior Systems Engineer
REDCap Administrator
Institute for Clinical and Translational Research (ICTR)
Johns Hopkins University, School of Medicine

Viktoriya Babicheva, MPH
Research Data Consultant and Acquisition Analyst
REDCap Administrator
Boston College, Research Services ITS