Methods for Remote Data Collection

JULIE MITCHELL

INFORMATICS & DATA COORDINATING CENTER

JANUARY 25, 2024

Agenda

Oregon Clinical and Translational Research Institute

- Electronic Data Capture Considerations
- Types of Survey Distribution
- Fraud & Bot Prevention
- Additional Considerations
- Scenarios
- Accessibility
- Participant Experience

Electronic Data Capture Considerations

Oregon Clinical and Translational Research Institute

- What type of device will be used to complete the survey?
- Where/ when will the survey be taken?
- What does internet /connectivity look like?
- How many devices will be trying to utilize the same connection at the same time?

Type of Survey Distribution Methods

Oregon Clinical and Translational Research Institute

OHSU

Public vs. Private Survey links

- Public links send allow surveys to be anonymous if no personally identifying information is collected
- Private (unique) links send allow only specific participants to complete. Most times, it restricts the number of times a link can be used. Link is automatically associated to a specific record.

Type of Survey Distribution Methods: *Private Survey Links*

Oregon Clinical and Translational Research Institute

- Link can be sent by text message or email
- Targets specific audience with unique links

Type of Survey Distribution Methods: *Public Survey Links*

- Common
- Useful
- Vulnerable to Fraud
  - Avoid posting links on social media

Type of Survey Distribution Methods: *Anonymous Surveys*

Oregon Clinical and Translational Research Institute

- Must be a public survey link
- Can not capture any PHI/HIPAA data
- No linking data over time
- Consider best places to post/distribute link
- If providing some type of survey compensation must use 2 separate surveys - *Refer to Fraud & BOT section*

Fraud & Bots

Oregon Clinical and Translational Research Institute

Fraud

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

Robot (bot)

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

↑ incentive = ↑ effort

Fraud & Bots

Oregon Clinical and Translational Research Institute

technology ↑ = ability to misuse ↑

- Study investigators are responsible for:
  - Understanding such threats and taking appropriate steps to mitigate risk
  - Ensuring that the highest quality data is collected
  - Compensation funds are distributed appropriately

Fraud & Bots:
Methods of Survey Fraud

Oregon Clinical and Translational Research Institute

- Eligibility Fraud
- Multiple identity Fraud
- Double-Dipping Fraud

Fraud & Bots: Identifying Fraud

Oregon Clinical and Translational Research Institute

- Abnormally fast completion rate
- Inconsistent responses
- Duplicate responses (identical or with slight modification)
- Irrelevant responses
- Responses to hidden questions
- Unusual navigation

Reference: Teitcher, J. E., Bockting, W. O., Bauermeister, J. A., Hoefer, C. J., Miner, M. H., & Klitzman, R. L. (2015). Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. *Journal of Law, Medicine & Ethics*, *43*(1), 116-133.
https://lifespan.ku.edu/online-surveys-and-data-collection-tools

Fraud & Bots: Human Intervention

Oregon Clinical and Translational Research Institute

Make a plan

- Study team should be proactive in making plans to limit risk/fraud
- Determine if public survey link is necessary
- Meet with software administrator (if available)
- Include details in applicable study materials

Fraud & Bots: Human Intervention

Oregon Clinical and Translational Research Institute

Monitor activity

- Determine how & what to monitor
- Determine responsible team member/ designee
- Schedule reviews – initially frequently after launch

Fraud & Bots: Human Intervention

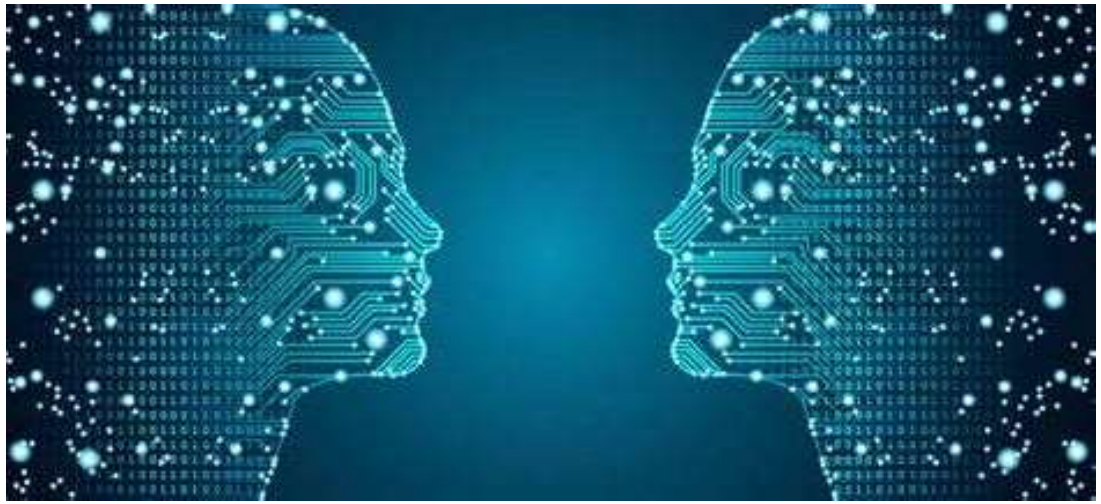
Oregon Clinical and Translational Research Institute

- Avoid fully automated compensation
  - Build in break-points
  - Put a human in the chain to review/approve compensation

Methods to Defend Surveys Against Bots

Oregon Clinical and Translational Research Institute

- Use of CAPTCHA's
- Set a "Response Limit"
- Implement Challenge questions
- Add "Honeypot" questions
- Include "Repetition" questions
- Paradata collection analysis

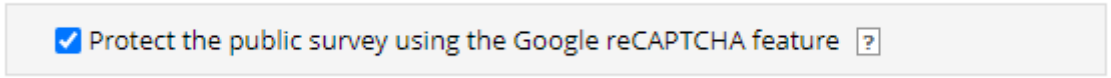


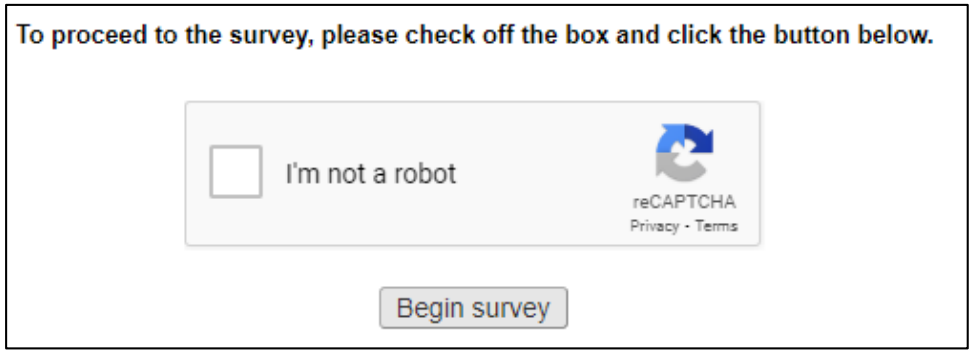
Defending Surveys: CAPTCHA

Oregon Clinical and Translational Research Institute

CAPTCHa, an ancronym for 'Completely Automated Public Turing Test to Tell Humans Apart', refers to various authentication methods that validate users as humans, and not bots. - source: ibm.com

In simple terms, a CAPTCHA can help prevent BOTs from accessing public surveys.

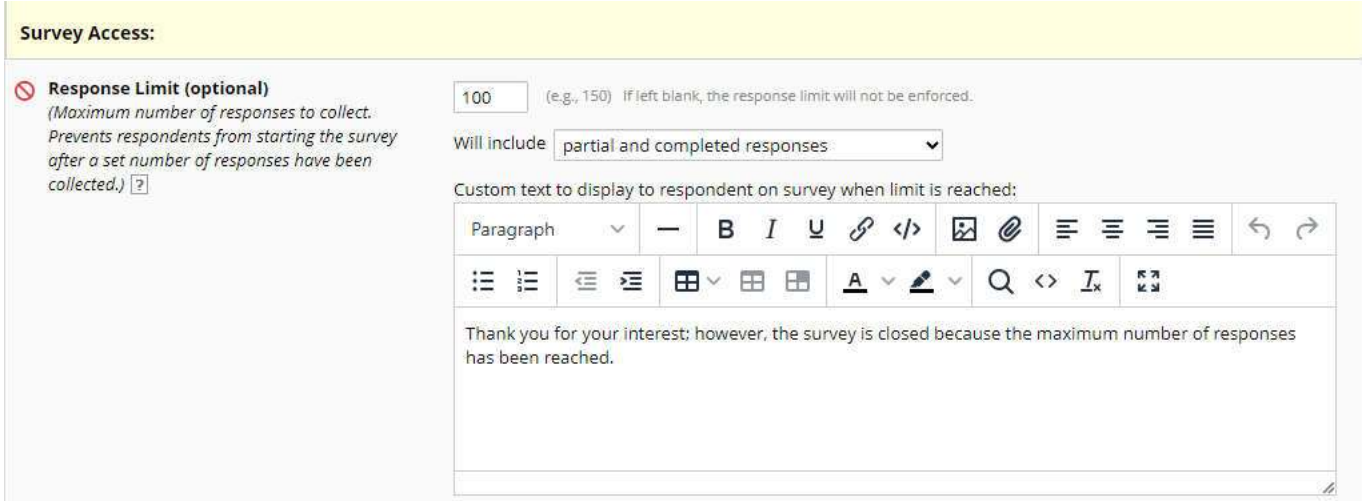




Defending Surveys: Response Limit

Oregon Clinical and Translational Research Institute

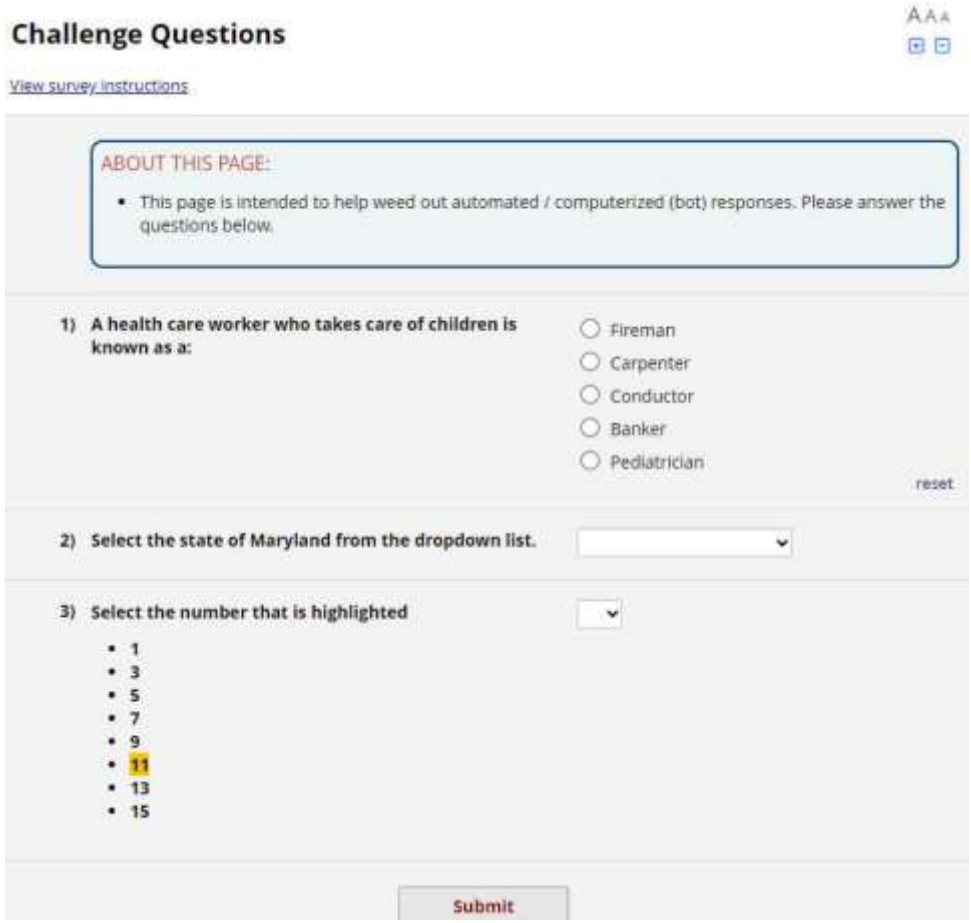
- Avoid a fraudulent "survey farm" attack by adding a response limit
- Limits financial liability
- Increase in increments to help avoid a wave of unintended responses



Defending Surveys: Challenge Questions

Oregon Clinical and Translational Research Institute

- Add a small set of "challenge" questions at the start of a public survey to help catch bot submissions
- Set logic to continue to the next survey ONLY if all of the challenge questions were answered correctly

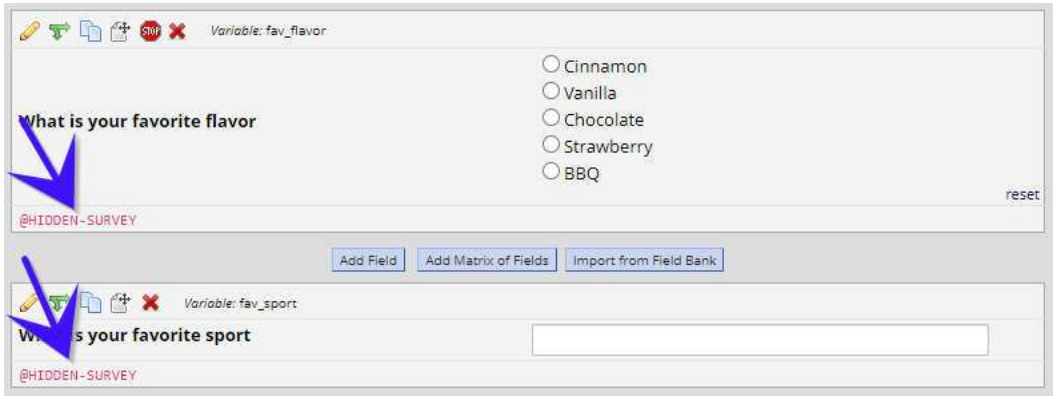


Defending Surveys: Honeypot Questions

Oregon Clinical and Translational Research Institute

"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)

- Add meaningless survey questions that are hidden on a survey (in REDCap use @HIDDEN-SURVEY action tag).
- This allows the question to be hidden on the screen but BOTS typically see the question.
- Prevent continuing on if honeypot questions have a response



Defending Surveys: Repetition

Oregon Clinical and Translational Research Institute

- Ask a few similar / identical questions using different sentence structures on different pages/surveys can help identify BOT responses on different pages/surveys.
- Fraudulent respondents are less likely to pay a great deal of attention to the actual questions.
- Flag records with inconsistent answers.

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

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

Defending Surveys: Paradata

Oregon Clinical and Translational Research Institute

"Paradata" is the operational or metadata saved about the process by which the data was collected.

- Paradata can affect the cost and management of a survey, the interpretation of results, evaluations of interviewers, and inferences about non-respondents

(source: Wikipedia)

Examples:

- When a survey was conducted
- Where (IP address) – requires IRB & participant approval
- Duration
- Number of contacts with each interviewee
- Attempts to contact the interviewee
- Reluctance of the interviewee
- Mode of communication

Methods to Defend Surveys Against Bad Actors

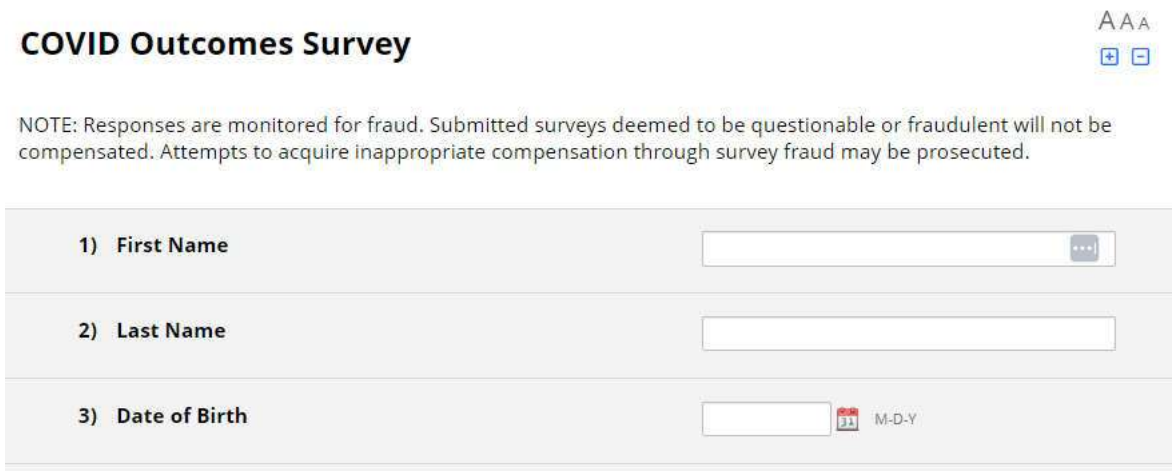
Oregon Clinical and Translational Research Institute

- Highlight Surveillance and Disclose Consequences
- Include Open Ended Questions
- Conscientious Responders Scale
- Targeted Survey Distribution
- Use Smart Incentives

Defending Surveys: Surveillance & Consequences

Oregon Clinical and Translational Research Institute

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



COVID Outcomes Survey

AAA

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

Defending Surveys: Ask Open-Ended Questions

Oregon Clinical and Translational Research Institute

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*

Defending Surveys: Conscientious Responders Scale

Oregon Clinical and Translational Research Institute

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.

- Five-item embeddable validity scale that differentiates between conscientious and indiscriminate responding

https://pubmed.ncbi.nlm.nih.gov/29914343/

Defending Surveys: Smart Incentives

Oregon Clinical and Translational Research Institute

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.
- 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".

Additional Considerations: Surveys with PHI

Oregon Clinical and Translational Research Institute

- PI's have a responsibility to ensure data captured is kept secure and confidential
- If 'Save & Return Later' functionality and ask for PHI, surveys should be password protected.
- PHI shouldn't be piped directly into a survey

Additional Considerations: Study materials

Oregon Clinical and Translational Research Institute

- Add text to consent and top of surveys that specify ~
  - *Participants may be contacted by telephone or email to confirm eligibility AND/OR Participants will not be compensated if suspected fraudulent or duplicate enrollment.*
- HIPAA identifiers to assist in Fraud & Bot prevention:
  - IP Address (check with IRB first- some only let IP address at Consent, others require it to be encrypted) - The IP address is stored totally separate from the study data captured in REDCap.
  - Email and/or Phone
  - Name
- Do not specify that data will be deleted from the system after study completion. Data can NOT be deleted from back-ups and overall system logs instead specify when/where data will be used
- Consent form and protocol should specify if survey invitations will be distributed via text messaging as this is done with a 3rd party vendor. (For text messaging in REDCap, Twilio, Inc. must be specified)

Additional Considerations: Best practices

Oregon Clinical and Translational Research Institute

- Ask survey questions to eliminate respondents that are NOT in your target audience.
- Require survey responses (if able to do so)
- Limit PHI in messages sent to participants with a private survey link

Scenario 1

Oregon Clinical and Translational Research Institute

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 by email 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.

The Good:

- Used CAPTCHA
- Attempted to use a "targeted" pool of study participants

The Bad:

- Didn't consider the possibility the link would be shared
- Didn't add any language or filters indicating participation was restricted to the target recipients
- Used a public survey link when it wasn't necessary

Mitigation:

- Add a "Response Limit"
- Use a unique link to each student
- Include language indicating only those who originally received the link are eligible

Scenario 2

Oregon Clinical and Translational Research Institute

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.

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

Mitigation:

- Add a CAPTCHA
- Add a "Response Limit" to limit the total number of responses
- Include language indicating who is eligible to participate and receive compensation
- Capturing an encrypted IP address

Scenario 3

Oregon Clinical and Translational Research Institute

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 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*).*

The Bad:

- No CAPTCHA
- Using social media to distribute public survey link
- Very few questions were included in the survey
- 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
- Use honeypot questions
- Use challenge questions
- Use a CAPTCHA

Fraud & Bot Prevention

Oregon Clinical and Translational Research Institute

- Time to consider the risks and mitigation is PRIOR to releasing a survey.
- PI's responsibility to anticipate fraud / bots and to actively take measures to mitigate risk.

Being proactive requires MUCH less effort than being reactive!

Accessibility: Electronic Surveys

Oregon Clinical and Translational Research Institute

- Bold
- Font size – at least 18 point or 14 point bolded
- Adequate line spacing 1/2 the height of the text and line spacing 1.5 times the height between paragraphs
- Static Images have a text-based description to convey content
- Audio or video have text based transcript to convey content
- Font & background colors (contrast ratio of at least 3:1)
- If your survey contains text fields that require a keyboard to appear on screen - does it block the question(s) on the screen?

Participant Experience

Oregon Clinical and Translational Research Institute

Survey Settings and Design Options

- Make radio and checkbox fields bigger
- Use radio boxes instead of dropdowns for any field where response options are not standardized
- Consider page breaks to limit scrolling
- Provide examples/instructions
- Set expectations of survey length
- Rich text editor
- Field embedding
- Survey notifications
- Confirmation acknowledgement or email
- Hiding and skipping questions & forms
- Determine when action is required by the participant or a study team
- Documents, images, videos

Participant Experience

Oregon Clinical and Translational Research Institute

OHSU

Test, Test, Test

Thank you

Oregon Clinical and Translational Research Institute

- Viktoriya Babicheva at Boston College
- Scott Carey at Johns Hopkins University

Questions?

Email to mitcjuli@ohsu.edu