Best Practices in Online Research

Design and Review in a World of Bots, Hackers, and Evolving Technology
Presenters

Christina Wright, M.A., CIP
Virginia Commonwealth University

Chiara Acquati, PhD, LMSW
University of Houston
Learning Objectives

- Identify signs of online survey compromise, differentiate fake from real research participants, and modify procedures to protect from further compromise through real-life case study examples
- Discuss how the IRB and principal investigators can work together to address survey compromise in terms of reporting requirements, compensation provision, and maintaining data confidentiality
- Share best practices for design of online research and for cybersecurity monitoring of survey data
- Describe best practices for data interpretation post-compromise
Online Research Methods

• Research methods that utilise the internet as medium for research
• Sometimes referred to as internet mediated research
• Includes research methods designed to investigate both online & offline phenomena
• surveys, experiments, interviews, observation, document analysis, virtual ethnographies, web analytics & content analysis, simulations, GIS
## Benefits & Perils of Online Research

<table>
<thead>
<tr>
<th>Investigators</th>
<th>Potential Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Can help achieve adequate participant numbers for data analyses;</td>
<td>• Less burdensome;</td>
</tr>
<tr>
<td>• Can allow a national or international scope;</td>
<td>• May heighten comfort with provision of sensitive information (decrease personal risk)</td>
</tr>
<tr>
<td>• Can access harder-to-reach populations;</td>
<td>• May facilitate participation in research (balanced power relationship)</td>
</tr>
<tr>
<td>• Can lead to more representative sampling;</td>
<td>• DEI consideration</td>
</tr>
<tr>
<td>• Expedite the data collection process;</td>
<td></td>
</tr>
<tr>
<td>• Cost-effective</td>
<td></td>
</tr>
<tr>
<td>• Reduced social desirability effects</td>
<td></td>
</tr>
</tbody>
</table>
Benefits & Perils of Online Research

- Inherent bias in internet-accessed samples
- Digital literacy
- Reduced levels of control over procedures
- Identity verification
- Building rapport with participants
- Technical competence and institutional support
- Ethical considerations
## Study Design Considerations

### Conceptualization
- Research question and target population, including key characteristics and type of data
- Benefit vs risks of online research
- Design and layout: organization of questionnaires/measure
- Software: contract, ownership of data, incentive distribution
- IRB submission: Informed consent forms; surveys, scripts for social media, communication with participants, contingency plan, monitoring plan

### Recruitment
- Community partners and key stakeholders
- FB pages, private groups, events, FB Ads, Creation of Institution level website for online research efforts
- Social Media posting, template, QRC
- Study information survey
- Prevent indexing, multiple submission
- Phone Screening
- Piloting
- Incentives: exercise caution

### Data Management
- Distinguish between study information survey and full questionnaire
- Password protected
- Individualized link, created for the participant unique e-mail
- Set expiration date
- Develop standard procedure to distribute survey access/info
- Limit accessibility: US, State, City level
- Consider HIPAA compliant tools for qualitative research and team sharing in post-COVID19 research
Example 1: Scripts and Images

A STUDY IS SEEKING 120 WOMEN DIAGNOSED WITH EARLY-STAGE BREAST CANCER TO BETTER UNDERSTAND HOW COVID-19 HAS IMPACTED THEIR DECISIONS AND EXPERIENCES OF CARE.

TO BE CONSIDERED FOR THIS STUDY, YOU MUST:
- Willing to participate in an online survey that will take about 30-40 minutes;
- Have received a diagnosis of early-stage breast cancer in January 2020 or later;
- Be willing to complete the survey at time of the study;
- Be at least 18 years of age;
- Have access to a computer, smartphone, tablet, or other device to complete an Internet-based survey;
- Be able to read and understand English.

THE STUDY INCLUDES:
- 120 women will complete an online survey that will take about 30-40 minutes;
- Survey participants will receive a $10 Amazon gift card;
- All survey participants will be invited to participate in a semi-structured interview, interview participants will receive a $40 Amazon gift card.

Please contact us:
Dr. Chiara Agostini
(713) 743-4343
cagostini@uh.edu
or e-mail us:
care@tmih.org

This project has been reviewed by the University of Houston Institutional Review Board, which approved it.
Example 2: Defensive Survey Tools
Example 2: Defensive Survey Tools

Prevent multiple submissions
End the survey for respondents and display a custom message, redirect to a different website, or flag the data for respondents trying to take your survey multiple times.

- **Action**: Redirect to URL
- **Website URL**: https://www.qualtrics.com/support/

Bot detection
We'll look for bots that might be taking your survey and flag their responses with an embedded data field (reCAPTCHA).

- **Action**: Off

Security scan monitor
Prevent security scanners from accidentally starting surveys when they test your link (reCAPTCHA).
Is this within the IRB’s purview?

Risks v. Benefits

- Even minimal risk to participants should be offset by at least minimal potential benefit of the research.
- For many studies, particularly minimal risk survey studies, the only benefit is scientific value, which may be lost if the data is compromised.
- Data compromise may also introduce risk to participants of not being compensated appropriately, or potential loss of confidentiality if research data is hacked and additional identity verification is needed.
Impact on IRB review

- The IRB submission should include a detailed recruitment plan, especially for social media recruitment
  - Consider targeted advertising rather than public blasts
  - Consider use of closed or moderated specific topic groups or professional accounts
  - Avoid advertising on more public-facing platforms like Twitter or Reddit
- Describe defensive study designs and data monitoring
  - Examples: screening to verify identity, attention checks, validity questions
- Consider adding language in advertisements and consent forms warning participants that compensation may be withheld if verification checks are not passed, that additional follow-up may be needed to verify identity, etc.
Incentives/Compensation Information

Compensation may attract bots/scammers
- Consider the risk of data compromise against benefits of compensation
- Ensure minimal focus on compensation in study advertisements
  - Avoid images like dollar signs, money bags, etc.
  - Avoid providing the $ amount
- Consider if compensation should be reduced or removed
- Consider multi-step screening and compensation procedures
- Include a time-frame for the expected compensation to be distributed
Collecting IP addresses or contact information may be worth it for identity verification

The threat of compromised data validity may outweigh the risks of collecting identifiers from participants.
IRBs and researchers can work together to ensure appropriate confidentiality and privacy measures are in place to handle identifiable data.
Work with Information Security or IT for guidance on preferred secure platforms.

If surveys will be anonymous, reviewing for targeted recruitment plans, defensive survey designs, and minimized compensation plans are even more important as there will be no way to go back to verify identity.
How YOUR IRB can HELP!

1. Train IRB staff and reviewers on what to look for:
   - Develop and disseminate institutional best practices.
   - Consider creating Social Media Recruitment Guidance: Work with your institution’s stakeholders to develop this; Create template management plans to help researchers develop their strategy.

2. Educate the research community:
   - Example: Encourage institutionally supported online accounts that can be used to disseminate research advertisements and survey links via secured and monitored accounts.

3. Work proactively with institutional stakeholders to raise awareness and set researchers up for success.
Things are going wrong...
What to do?
# Strategies to Identify and Prevent Fraudulent Data

## Software-enabled tools
- Software with fraud prevention detection capabilities
- Screening questions, CAPTCHA
- IP address and GIS (cluster)
- Time stamps, compare with estimated completion time
- Track methods for survey completion (Link, QRC)
- Quota

## Duplicate or unusual responses, Bots
- Survey completion time (estimate and range)
- Patterns in answers: straight lining
- Exact response provided multiple times
- Content of the response: irrelevant, too general, not applicable to the target population, duplicate from website
- Zip code and GIS do not match, cluster of responses from one area of the country
- Identify a source beyond the list of organizations/contacts
- Hidden items = bots

## Experience and IRB collaboration
- Remuneration Letter;
- Request acknowledgement of eligibility/exclusion criteria
- Set a % of the survey with mandated answers
### Example 3 – Fraudulent Data Detection

#### Table

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start Date</td>
<td>End Date</td>
<td>Status</td>
<td>IPAddress</td>
<td>Progress</td>
<td>Duration (in)</td>
<td>Finished</td>
<td>Recorded Date</td>
<td>ResponseID</td>
<td>RecipientLastName</td>
<td>RecipientFirstName</td>
<td>RecipientEmail</td>
</tr>
<tr>
<td>11/20/19 12:06</td>
<td>11/20/19 12:06</td>
<td>0</td>
<td>100</td>
<td>36</td>
<td>TRUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/20/19 12:06</td>
<td>11/20/19 12:06</td>
<td>0</td>
<td>100</td>
<td>24</td>
<td>TRUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/20/19 12:06</td>
<td>11/20/19 12:06</td>
<td>0</td>
<td>100</td>
<td>36</td>
<td>TRUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/20/19 12:06</td>
<td>11/20/19 12:06</td>
<td>0</td>
<td>100</td>
<td>24</td>
<td>TRUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/20/19 12:06</td>
<td>11/20/19 12:06</td>
<td>0</td>
<td>100</td>
<td>26</td>
<td>TRUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/20/19 12:06</td>
<td>11/20/19 12:06</td>
<td>0</td>
<td>100</td>
<td>35</td>
<td>TRUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/20/19 12:06</td>
<td>11/20/19 12:06</td>
<td>0</td>
<td>100</td>
<td>36</td>
<td>TRUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/20/19 12:06</td>
<td>11/20/19 12:06</td>
<td>0</td>
<td>100</td>
<td>23</td>
<td>TRUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/20/19 12:06</td>
<td>11/20/19 12:06</td>
<td>0</td>
<td>100</td>
<td>22</td>
<td>TRUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Full Record

- **Record at:** September 21, 2017
- **Record at:** September 21, 2017

**Location ID:** 33456

**Q_RelatedResponseField:** Location ID

**Q_RelatedResponseFieldValues:** 33456

**Location Data**

- **Location:** (40.7599061, -73.985298631)
- **Source:** Google Maps

---

**Full Record**

- **Show full record:**

**Status**:  

**Priority**:  

**Location**:  

**Full Record Fields**:  

---

**Close**
Reporting to the IRB

The IRB will need to consider if the event involved any of the following:

- increased risk or harm to participants
  - Ex: risk to data confidentiality, inability to compensate, etc.
- increased risk to scientific integrity of the study
  - Ex. loss of valid data; inability to separate good data from bad, etc.
- noncompliance with the approved protocol
  - Ex. procedures intended to mitigate compromise not followed; compensation not provided as promised in the consent; altering recruitment strategies or data management plan without IRB approval.
Corrective Actions

- Balance the risk of losing data or not compensating valid participants against the risk of stopping the entire study and/or loss of research funds to pay scammers.
  - Can the PI compensate all (including scammers) and still salvage the research?
  - Can the PI reasonably determine who should not be compensated?
- If the survey will be re-posted, what changes are needed to the approved protocol to avoid a repeat incident?
- How will the event and any changes to compensation or screening be communicated to participants, and how will complaints be handled?
- Case study example –
  - Corrective action plan for survey of registered nurses
Takeaway for Investigators

• Enhance Software: contract, ownership of data, incentive distribution, fraud prevention detection capabilities
• Assess recruitment methods that will mitigate the problem
• Establish data collection and management methods that will reduce the opportunity for fraudulent submissions
• Monitor the data for Duplicate or unusual responses, Bots
Takeaways for IRB Review and Institutional Support

• Provide a detailed recruitment plan to help the IRB make an appropriate risk/benefit determination
• Describe how you will implement defensive study designs such as screening to verify identity, attention checks, validity questions
• Consider if collecting identifiers is more effective for deterring bad actors, and explain this in the IRB submission
• Evaluate compensation plans
• Encourage training for IRB members/staff and Research Investigators
• Encourage the development of institutional resources