When Research Becomes All About the Bots: A Case Study on Fraud Prevention and Participant Validation in the Context of Abortion Storytelling

Michaela Krawczyk
mikrawcz@iu.edu
Indiana University
Bloomington, Indiana, USA

Katie A. Siek
ksiek@iu.edu
Indiana University
Bloomington, Indiana, USA

ABSTRACT

Effective fraud prevention and participant validation are essential for ensuring data quality in today's highly-digitized research landscape. Increasingly sophisticated bots and high levels of fraudulent participants have generated a need for more complex and nuanced methods to combat fraudulent activity. In this paper, we share our experiences with fraudulent survey responses, which we encountered in our work around abortion storytelling, and the multi-stage protocol that we developed to validate participants. We found that effective fraud prevention should start early and include a variety of flagging methods to encourage holistic pattern-searching in data. Researchers should overestimate the amount of time they will need to validate participants and consider asking participants to assist in the validation process. We encourage researchers to be transparent about the interpretive nature of this work. To this end, we contribute a Participant Validation Guide in supplemental materials for community members to adapt in their own practices.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; • **Social and professional topics** → User characteristics.

KEYWORDS

fraud, fraud prevention, digital survey, recruitment, asynchronous remote community

ACM Reference Format:
Michaela Krawczyk and Katie A. Siek. 2024. When Research Becomes All About the Bots: A Case Study on Fraud Prevention and Participant Validation in the Context of Abortion Storytelling. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3613905.3637109

1 INTRODUCTION

Researchers who utilize digital methods and tools have experienced increasing levels of fraudulent activity in recent years, such as high volumes of automated "bot" responses on digital recruitment surveys [12, 13, 16] or intentional deception from participants during remote interviews [14, 15]. We define "fraudulent activity" as the intentional submission of data from people who are ineligible for a study, typically in the pursuit of compensation. This umbrella term includes automated "bot" responses engineered by people with disingenuous or even malicious intent, as well as people (eligible or not) who attempt to participate in a study multiple times. These cases present a major problem for modern researchers because such fraudulent activity compromises the validity of data and its subsequent findings.

Fraud prevention is a dynamic process, requiring iterative methods across study design and participant interactions. CAPTCHA scoring and surface-level attention checks (e.g., type the word "blue" to prove you're human) are not enough to effectively weed out automated responses, nor intentionally misleading participants [1, 5, 12, 17, 18]. We are in dire need of more nuanced and multi-pronged techniques to ward off increasingly-complex tools used for fraudulent research behavior. Health and social science researchers have published their experiences and compiled fraud-related considerations across multiple types of research, from digital survey data [1, 5, 6, 12–14, 16–18] to qualitative remote interviewing [14, 15, 17]. One overarching theme is that there is no "one size fits all," best protocol for fraud prevention. Rather, effective participant and data validation requires a combination of diverse techniques [6, 12, 16], each with their own affordances, drawbacks, and ethical considerations. It is crucial that researchers become more transparent about the methods they use to ensure the validity of their data—from study design, to recruitment, to data cleaning.

We contribute to this dialogue by sharing our own experiences with fraudulent activity while designing and implementing a study on abortion storytelling. The goal of this work was to explore how digital tools can support people in writing and sharing their abortion experiences. We encountered fraudulent activity through high volumes of automated responses on two digital surveys, an initial interest survey and the recruitment survey for a 5-week asynchronous remote community (ARC) workshop. In between these surveys, we spent 5 months parsing through fraudulent data and developing a multi-step protocol that helped us transition from 127 Intake Survey responses to 26 validated participant invitations for the ARC workshop. Our understanding of fraud prevention and participant validation continued to evolve through study design, data analysis, and paper writing. Reexamining study data with an awareness of fraudulent participation will be a continuous learning process for the CHI research community as technologies develop that widen

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3637109

remote research opportunities and increase the capabilities of tools that allow others to commit research fraud.

In this case study, we present our experiences with fraud prevention and participant validation in the context of abortion storytelling. We provide our methodological strategies while designing and implementing our recruitment survey, the protocol we used to validate participants, and a collection of lessons learned. Although these strategies were developed in the context of abortion storytelling, the nature of fraudulent activity, good survey design, and effective data cleaning are widely applicable.

We encourage fellow researchers to use our methods as a baseline and make adaptations that better fit the nuances of their own studies and topic spaces. To this end we include a Participant Validation Guide (see supplemental materials), formatted as a table of questions and considerations, which evolved from our original protocol, lessons learned, and external research into this space of fraud prevention. The HCI community can use our findings to become aware of the changing landscape of remote participant engagement and to inform the development of their own work. This case study also contributes to the broader CHI culture of transparency [9] and mutual benefit as we fight for the validity of our data.

2 OUR INTRODUCTION TO FRAUDULENT ACTIVITY

Our first foray into fraudulent activity stemmed from an initial interest survey we created on Google Forms in the Summer of 2022. The goal of the digital survey was to collect demographic data and gauge general interest in abortion storytelling, with the goal of informing a full study design. We received 981 survey responses in under four hours—many of which seemed immediately suspect (an experience in which we are not alone [5, 12, 13, 16]). We suspended data collection and spent the next two months parsing through the collected responses and iterating on methods to verify participants. These tactics helped us clean our data from 981 survey responses to 15 validated entries, although our subjective intuition said that only 1–3 responses were truly authentic. We improved upon these participant validation methods while recruiting for the full asynchronous remote community (ARC) study.

3 ARC WORKSHOP: PARTICIPANT VALIDATION METHODS

In this section, we detail our methodology for recruiting and validating participants in preparation for the asynchronous remote community (ARC) workshop. Our multi-step protocol, as shown in Figure 1, included two surveys and two rounds of manual coding, helping us narrow from 127 survey submissions to 26 validated participant invitations. Once the study began, 22 participants accepted the invitation, and 17 participants were active upon joining. We also discuss our investigation into one of these remaining participants who we later suspected to be fraudulent, and our decision to keep their responses in the final dataset.

3.1 About the ARC Workshop

The goal of an asynchronous remote community (ARC) is to understand the unique experiences and design needs of a particular group by engaging participants in activities over multiple weeks. Previous ARC studies have worked with people living with HIV [11], rare diseases [10], and miscarriage experiences [8]. Our team designed and implemented an ARC study centered around abortion experiences. We call it an "ARC Workshop" because a main goal of our study was to help participants iterate on their written abortion story, thereby giving it the spirit of a writing workshop. All materials were approved by the Indiana University Institutional Review Board. We engaged 17 participants with abortion experiences over a span of five weeks between January and February 2023. All participants joined a private Slack workspace, where we assigned 14 activities (2–3 activities per week) that aimed to promote participant interactions, spark conversation about abortion, and prompt story iteration.

3.2 Recruitment

We recruited participants through physical flyers, the All-IN-4-Health research volunteer registry, and social media postings, including: Instagram, Twitter/𝕏, and paid Facebook advertisements. We also contacted people who had publicly posted their abortion stories on these platforms with one of the following hashtags: #ShoutYourAbortion, #YouKnowMe, #IHadAnAbortion, and #abortion. These IRB-approved recruitment methods directed people to the digital Intake Survey.

Participants needed to be 18+ and had personally obtained an abortion before June 24, 2022. We defined "abortion" as the intentional termination of a pregnancy and requested experiences prior to this date due to the overturning of Roe v Wade and the resulting criminalization of abortion in some U.S. states. We offered an incentive of $16/week via digital Amazon gift card (up to $80 total). In response to some discussion about the relationship between fraudulent activity and the amount of information displayed in recruitment methods [1, 3, 17], all of our postings listed the inclusion criteria and a vague compensation statement (see supplemental materials for recruitment flyer). The incentive format and amount was not specified until page four of the Intake Survey, which was only accessible after the two screening questions.

3.3 Intake Survey

The Intake Survey (see supplemental materials) followed a similar structure as our initial interest survey. It included screening questions, informed consent materials, and questions about: demographics, abortion experience, abortion storytelling experience, social media usage, and a digital version of the Individual Level Abortion Stigma (ILAS) scale [4]. The Intake Survey was open for 3 weeks and received 127 responses.

3.3.1 Survey Design. The first design change we made between our initial interest survey and the Intake Survey for our ARC Workshop was switching to the Qualtrics platform, so we could access its fraud detection features. We activated the following features: (1) Prevent multiple submissions, (2) Bot detection (reCAPTCHA), (3) Security scan monitor, (4) RelevantID, (5) Prevent indexing, and (6) [Deactivate] Anonymize responses (allow IP tracking). Collecting non-anonymized responses became especially important for later-stage participant validation because it allowed us to flag responses with duplicate IP addresses and, therefore, geographic coordinates.

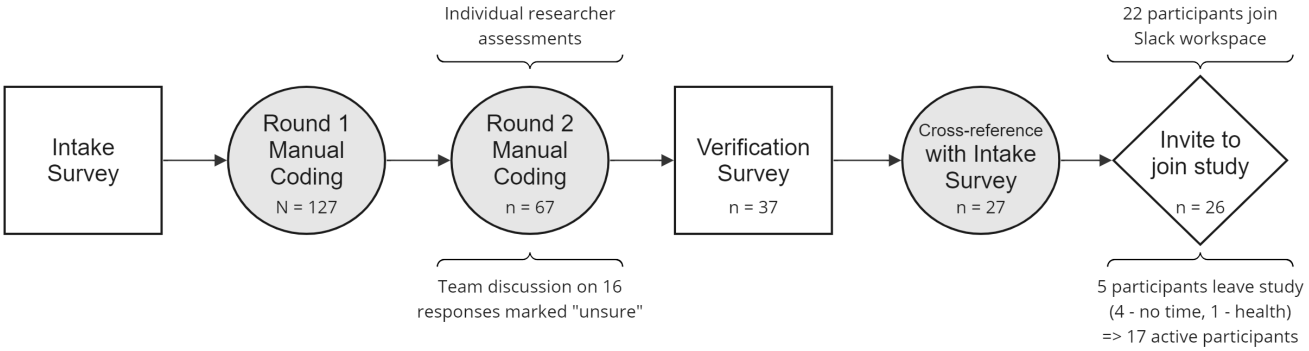
Individual researcher assessments
22 participants join Slack workspace
Intake Survey
Round 1 Manual Coding N = 127
Round 2 Manual Coding n = 67
Verification Survey n = 37
Cross-reference with Intake Survey n = 27
Invite to join study n = 26
Team discussion on 16 responses marked "unsure"
5 participants leave study (4 - no time, 1 - health) => 17 active participants

Figure 1: Flowchart of participant validation methods during recruitment for the ARC Workshop.

It also allowed us to cross-reference IP locations with the geographic region that respondents reported in their survey submissions.

In terms of front-facing survey content, we included many open-ended questions, with the goal of creating more opportunities for researchers to identify bot activity. For example, we asked participants to type out their abortion year, instead of picking from a dropdown list, and we provided multiple free-response opportunities (e.g., "What are you hoping to get out of this experience?"). Only one of these free-response questions required a response to move forward; however, a majority of survey respondents (both validated and suspect) filled them out.

Another tactic that we implemented was cross-referencing. During the Intake Survey, respondents provided the year of their first abortion, age receiving that abortion, and current age. We mathematically validated survey responses—someone's current age subtracted by their abortion age should align with the current year subtracted by their abortion year (±2 years to account for birthdays and human error).

We also conducted "consistency checks" where researchers ask about a topic in multiple ways or locations, creating opportunities for self-contradiction. For instance, our Intake Survey asked respondents about their preferred social media platforms in three different locations. First, it prompted respondents to select all social media platforms that they used from a checkbox list. Later, a Likert-scale question asked them to rate the same list of platforms according to the likelihood that they would post their story there. Finally, an open-ended question asked which platform they would most likely use to share their story and why. We asked about social media usage once more in our post-study Feedback Survey, allowing us to also check for consistency across multiple survey submissions.

3.4 Manual Coding

We investigated the validity of Intake Survey responses through two rounds of manual coding. For Round 1, we rejected or flagged responses according to survey metadata. For Round 2, we investigated the remaining survey responses more closely and qualitatively. In Round 1, 67 survey responses passed, and 37 responses passed Round 2. We provide an in-depth list of questions and considerations to assist others through this process in the Participant Validation Guide (see supplemental materials).

3.4.1 Round 1 Coding. We exported the 127 Intake Survey responses to a secure Google spreadsheet. We began our investigation by flagging the following items:

(1) Response failed any Qualtrics fraud protection feature
  - reCAPTCHA score < 0.8
  - RelevantID Duplicate Entry = true
  - RelevantID Fraud Score > 30

(2) Duplicate IP address
(3) Duplicate geolocation (latitude/longitude derived from IP address)
(4) Invalid abortion year math (difference of more than 2 years)

Table 1: Number of Intake Survey responses flagged in Round 1 manual coding (N = 127).

Round 1 coding item	# flagged responses (%)
Failed Qualtrics reCAPTCHA scoring at 0.5 benchmark	6 (4.7)
Failed Qualtrics reCAPTCHA scoring at 0.8 benchmark	7 (5.5)
Failed Qualtrics RelevantID	51 (40.2)
Duplicate IP address	20 (15.7)
Duplicate IP geolocation	58 (45.7)
Invalid abortion year math	6 (4.7)

We rejected almost 50% of survey responses for failing items 2–4, as shown in Table 1. We flagged, rather than rejected, items that only failed the Qualtrics fraud protection features because shortly after the implementation of the Intake Survey, we suspected that their accuracy was limited. For context, 117 Intake Survey responses (92%) scored above the Qualtrics-recommended reCAPTCHA benchmark for "most likely human" (0.5)—a lower level of detection than we had expected. We also saw similar reCAPTCHA scores between responses from people we knew to be real and responses that were highly suspect.

Review of external literature revealed similar observations on the limitations of both CAPTCHA and IP tracking, due to increasingly sophisticated bots [16, 17], inaccurate geolocation technology [1, 5, 16, 17], and consumer behaviors, such as device sharing [1, 3, 17] and VPN usage [1, 18]. Despite these limitations, we continued to rely heavily on both IP tracking and Qualtrics metadata tools because rejecting survey responses with duplicate IP addresses or low RelevantID scores cleaned more survey submissions than any other techniques in our protocol (see Table 1). Future work might investigate researcher attitudes toward CAPTCHA and IP tracking, with the goal of developing better fraud prevention tools that address current limitations and ethical concerns.

3.4.2 Round 2 Coding. 67 survey responses passed Round 1 of manual coding. The second round of coding was split into two parts—individual assessments and team discussion. For individual assessments, both researchers examined the remaining survey responses individually and categorized their validity as "valid," "invalid," or "unsure." 16 total survey responses were marked as "unsure" by either researcher. Following these individual-level assessments, researchers discussed the 16 "unsure" responses. During both parts of this process, researchers took a holistic approach by examining Round 1 flags and the following list of considerations:

(1) Do their short-answer and free-response questions make sense?
(2) Do their Likert scale responses make sense? (e.g., They did not put the same value for all 8 sets of Likert scales)
(3) How long did their survey response take compared to other submissions/estimated time?
(4) Does their IP geolocation match their reported location (U.S. State) within reason?
(5) Does their IP geolocation make sense? (e.g., not in the middle of the ocean)
(6) Can we verify their existence online? (e.g., Social media, LinkedIn, personal website, organization webpage, etc.)
(7) Does their email address pull up a profile image on Google Sheets? (Only 1 out of 21 survey responses with a profile image attached to their email address was determined invalid through other considerations)

Of these 16 "unsure" responses, we validated 3, rejected 9 according to the above criteria, and ended with just 4 participants that we remained unsure about. At this stage of coding, many decisions became fairly subjective. Researcher judgment was heavily influenced by participants' online presence and the similarity of their IP address to their reported location. Combining the "valid" and "unsure" responses, we ended with 37 survey responses that passed Round 2 coding.

3.5 Verification Survey

Our last step towards participant validation, shown in Figure 2, was a short secondary Verification Survey (see supplemental materials) that we sent to the 37 respondents who passed Round 2 Coding, so we could cross-reference information from the Intake Survey. We activated all available Qualtrics fraud detection features. Participants entered their name, age, and first abortion year. The final question asked them to type out a short statement on respecting fellow participants. We received 27 responses and rejected one person whose abortion year was off by 14 years between the two surveys and who took only six minutes to complete the Intake Survey (compared to an average of 15 minutes, excluding two respondents who took 3 and 8 hours). We invited the 26 verified participants to enroll in the study.

3.6 Email Communication

Throughout this recruitment process, we sent emails to all 127 Intake Survey respondents with their study status. For respondents flagged as fraudulent, we sent the following message.

> "Hello! Thank you for your interest in our study on abortion storytelling. We received your response to the pre-survey; however, our system has flagged your submission as a potential bot/computer. As a result, you are currently ineligible to join the study. If this is incorrect, please reach out to one of our researchers! There's a lot we still don't understand about automated responses to surveys and bot protection. [Researcher contact information]"

Several studies wrote about their communication of suspected fraudulence to survey respondents [1–3, 5, 13, 16]. Unlike most of these studies who report receiving little to no response, we received over 30 messages in response to this "suspected bot" email. Some email responses were helpful to our participant validation process, such as addresses that bounced back as "Failed Delivery" or messages that contained major grammatical errors and, sometimes, outright hostility (examples below).

> "Is not dear Krawczyk, am not a robot, am a human being and my name is [name], your computer can malfunction, and you don't depend on electronic device because it can disappoint."

> "What you're saying is fake and a lie am a real person and ready to attain to share!"

For most messages, however, it was impossible to tell (e.g., *"Hi, Thank you for your response. However, I am a real person and this must be a mistake. Thanks."*). One reason is that short-form text is much harder to vet for authenticity than longer passages. Another, larger, reason is that we did not provide respondents with any tangible action points to "prove" their authenticity. As a result, our outreach was unintentionally performative. In future studies, we would go beyond encouraging participants to "reach out" and, instead, offer next steps for respondents, such as calling researchers, setting up an additional screening, or providing a website that would let us verify their existence outside the survey.

3.7 Suspected Fraudulence Post-Recruitment

After completing the 5-week ARC Workshop, we began a combination of quantitative and qualitative analyses on all exported participant data, including: Slack messages, iterations of written abortion stories, and responses to a Feedback Survey. During this process, we began to suspect one participant (P15) as potentially fraudulent, namely because they participated very little and late, and they used an exclamation mark for every sentence in their abortion story (excerpt below).

> "There I was 24 and pregnant, by a guy all my family said was no good but I was in love! Told him I was pregnant and he told me it wasn't his, I was completely heartbroken!"

Looking back, we marked P15 as "unsure" in Round 2 coding of the Intake Survey because they only took 8 minutes to complete the survey and their IP geolocation was slightly different from their

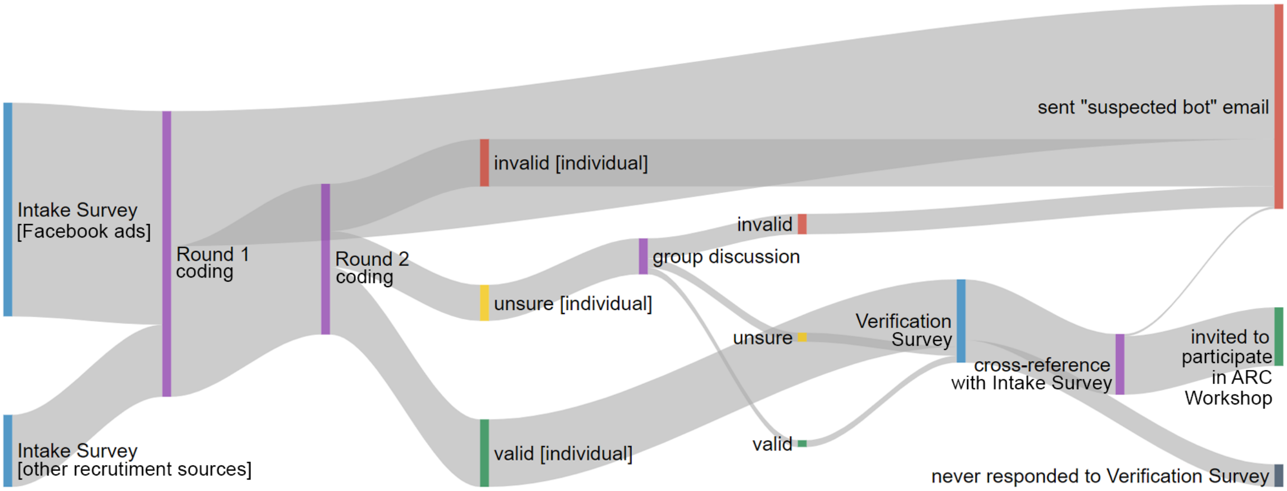
Intake Survey [Facebook ads]
Intake Survey [other recrutiment sources]
Round 1 coding
Round 2 coding
invalid [individual]
unsure [individual]
valid [individual]
sent "suspected bot" email
invalid
group discussion
unsure
valid
Verification Survey
cross-reference with Intake Survey
invited to participate in ARC Workshop
never responded to Verification Survey

Figure 2: Flow of survey respondents through our participant validation methods while recruiting for the ARC Workshop. We began with 127 Intake Survey responses and ended with 26 validated participant invitations.

reported location. They also reported two abortion years with a difference of 1 year between the Intake and Verification Survey (e.g., 2018 vs. 2019); however, so did 8 validated participants, including people we knew to be real. Here are the post-recruitment items that we took into consideration while investigating P15 as potentially fraudulent. We indicated each item as ● = suspect, ◐= slightly suspect, or ○ = not suspect.

- Level of participation
  - ◐ Stopped responding after they posted their abortion story for Activity 2
  - ○ Took the Feedback Survey at the end of the 5 weeks
- Timing of participation
  - ◐ Completed their activities almost an entire month after other participants
  - ● Did not respond to researcher check-ins
  - ○ Offered a plausible reason for their limited participation (health)
- Adherence to activities
  - ◐ Provided extraneous detail in their introduction compared to other participants (e.g., description of pets)
  - ○ Story content was plausible and related to abortion
- Quality of responses
  - ● Highly irregular punctuation and some grammatical mistakes
  - ○ Responses still made sense
- Consistency checks
  - ◐ Rated activities in the Feedback Survey that they did not complete
  - ○ Reported social media usage lined up between the Intake Survey and Feedback Survey

In the end, we decided to keep P15 in the dataset. Our investigation had reached the point of such researcher subjectivity that we revisited the overarching goals of our study and our own academic morals. We were not comfortable removing a participant's data according to a "gut feeling," and if we based our decision off the tangible evidence we had—limited participation, late posting, punctuation and grammar, and minor inaccuracies between surveys—, then we needed to apply those same thresholds to the other 16 participants, resulting in the loss of perfectly eligible and insightful data. Additionally, we had to acknowledge our own potential biases (participant writing abilities will vary greatly; thus, we should not judge too many exclamation points!) and expectations for participant behavior (participants should prioritize their health over study participation).

Our team decided that the value of our study and its findings was rooted in the themes we developed from participants' stories and activity responses, not necessarily the total success of our recruitment protocol. P15's data contributed so little to our thematic findings that they essentially only existed in our demographics table. As such, we decided that keeping them in our final participant pool was less threatening to our overall data quality than removing them (and others) would be. Note that this subjective, value-laden, and research-goal-oriented conversation will, and should, look different for every research team. We encourage all teams to have similar discussions, perhaps even before designing their study, while determining the level of exclusivity that they want to instill in their own methods.

4 LESSONS LEARNED

We offer a collection of lessons learned while iterating on our participant validation tools that informed our Participant Validation Guide (see supplemental materials). Our lessons contain observations about the nature of effective fraud prevention, things we would have done differently, and considerations for future work.

L1: Effective fraud prevention begins early. *To be effective, researchers should build fraud prevention and participant validation* *techniques into the study design, agree on flagging protocols prior to participant interactions, and revisit data throughout the study.* We frequently referenced data collected several months prior to inform later discussions and validation decisions, such as looking at Intake Survey responses while investigating P15. To support adaptive fraud prevention, teams may need to plan and budget for validation tools and labor hours [6, 7].

If a team is using a digital survey, then—in addition to our survey design considerations in section 3.3—we recommend *creating survey links that map to specific recruitment methods, so the team can track where data is coming from.* In the event that one recruitment method is leading to high levels of fraudulent activity, the team can shut down response intake [13, 14, 16, 17]. Our team divided public recruitment into two sources (Facebook Ads and Other); however, we wish we had tracked survey links with more specificity. For context, 20/32 (62.5%) survey responses collected from Facebook ads were determined to be valid, while the survey tied to all other recruitment options, including physical flyers, Instagram, and Twitter/𝕏, received only 12 (13.3%) valid responses out of 90 submissions. Beyond trackable links, it is also important to check survey data frequently to identify fraudulent data accumulation [7] and shut down compromised sources.

Another validation technique we recommend is to *first collect data from people the research team know to be real, so the convenience sample data can be used for comparison later on [16].* Our team sent a copy of our Intake Survey to a group of university colleagues, which gave us a final round of external feedback and a collection of submissions that helped inform our interpretation of actual participant metadata.

L2: Effective fraud prevention requires a combination of diverse tactics and a holistic, pattern-searching approach. Early in our study, while weeding out bots from survey data, it became apparent that a single tactic, such as CAPTCHA, or a single flagging technique was not sufficient. Rather, it was the slightly suspect free-response question in combination with an improbable Likert scale and/or a lower reCAPTCHA score that helped us feel confident in our rejection of a survey respondent. This holistic approach was important at the participant-level and the dataset-level because it allowed us to search for broader patterns [5, 12, 14]. Sometimes, a free-response question was not suspicious until we noticed that 20 other respondents, some with duplicate IP addresses, answered those questions with the exact same phrasing. In short, researchers cannot rely on automated metadata for fraud protection. *Research teams should implement a variety of flagging techniques (e.g., metadata, close-ended, open-ended, mathematically verifiable, etc.) to create more opportunities for cross-referencing and better recognize fraudulent patterns.*

L3: Overestimate the amount of time that participant validation will take. We were surprised by the level and speed of fraudulent activity that we received on our initial interest survey; however, a bigger surprise was the amount of time and effort it took to validate our data afterwards. Both times that we disseminated digital surveys to the public, we estimated around 4 weeks for data collection, cleaning, and analysis. For the initial interest survey, we spent around 2 months designing new methods and validating participants. For ARC Workshop recruitment, almost 3 months passed between posting the Intake Survey (Oct 12, 2022), distributing the Verification Survey to respondents who passed Round 2 coding (Nov 15, 2022), and launching the ARC Workshop with validated participants (Jan 09, 2023). Granted, we intentionally delayed the study's start date to account for winter holidays.

The validation process should go faster for those who anticipate fraud prevention and build flagging mechanisms into their participant interactions; however, time will need to be budgeted for possible iteration on validation techniques (e.g. familiarizing oneself with a dataset to notice sophisticated bot patterns). In our case, we administered a Verification Survey to corroborate information and contacted participants suspected of fraudulence. These follow-up activities required additional time for participants to respond and then researcher time to clean and cross-reference the data. We encourage the research community to continue developing tools and techniques to make validation more efficient while still prioritizing the well-being of participants and the integrity of the scientific process.

L4: Consider asking participants to assist in their own validation. A huge irony of digital survey research is that the same mechanisms which allow participant anonymity—widening the opportunity for study of geographically-separated and stigmatized communities—also allow increased levels of fraudulent activity. Consequently, there is major tension between effective fraud prevention and participant privacy [16, 17]. As Jones et al. identified, the challenge is figuring out the minimum amount of information needed to assess eligibility, so we can avoid inquiring beyond it [7]. We do not have a solution to this problem; however, we provide tactics that we will try in future studies. One limitation of our participant validation protocol is the performative nature of our "suspected bot" emails because we did not provide any tangible action points for people to respond with. Another limitation is the subjective and non-exhaustive method of searching for people's identities online as a form of verification.

In the future, we plan to *explore strategies that ask participants to assist in their own validation,* such as requesting an organizational website or social media handle that can verify their existence outside of a survey. This technique is supported by other studies that have requested personal information (e.g., phone numbers or mailing addresses) [1, 3, 6, 13]. Of course, this technique faces limitations, especially while operating in more stigmatized spaces, such as abortion; however, we think there may be something lucrative in this concept of collaborative validation where researchers emphasize the need for participant assistance in ensuring the quality of their data while studying something that both parties, presumably, have a stake in.

L5: Researcher subjectivity is inevitable. What do you do when you've run out of "objective" markers, but still feel like a survey response is fraudulent? Shortly after starting our journey into fraud prevention, we reached a place where all next steps required a high level of interpretation (e.g., how many grammatical errors in an open-response survey should constitute a fraudulent response?). Even harder—when, exactly, should a response become "fraudulent"? Jones et al. wrote that "researchers should avoid being overly influenced by their expectations of what the data 'should' look like [and that], ideally, response screening would be facilitated by someone who is blinded to the study hypothesis and aims" [7]. When this is not feasible, how do you safeguard the validity of your data while checking your own biases? In response to this complex question, we developed the Participant Validation Guide to help us get closer to an objective measure before bringing in subjective considerations. Moreover, we want to note that abortion is a highly stigmatized experience, and we decided to err on the side of losing eligible participants, rather than risk the integrity of our study or trust with participants and the research community.

In terms of tangible advice *for checking biases, we recommend getting frequent external feedback on both the research team's survey instruments and data validation protocols.* We presented our Participant Validation Guide and anonymized data from P15 to our lab and received instrumental feedback on whether we could remove P15 from the final dataset (we could not) and how to improve future participant validation protocols (e.g., adding action points to our "suspected bot" emails). Moreover, we agree with Ridge et al.'s assessment that it is important to create a culture of open dialogue where team members can express uncertainty about participant validity and interrogate the methods being used to determine fraudulence [14].

5 LIMITATIONS

Some limitations to our participant validation methods are scattered among our Lessons Learned, such as our treatment of email communication and the heavy-handedness of our weeding-out process due to the stigmatized nature of the research topic. Our suggested validation techniques are shaped by the nature of our own data collection methods, namely digital recruitment surveys and qualitative asynchronous engagement—resulting in the exclusion of other valuable tactics geared towards methods such as real-time interviewing [6, 14, 15, 17] or completely asynchronous studies that do not always offer the ability to reach out to participants suspected of fraud [7]. As such, our validation protocol will have varying levels of applicability to other work—although we suspect that our overall thought process and lessons learned will have wider generalization.

Perhaps the biggest limitation of this paper is its ability to inform and strengthen the very types of fraudulent behavior that we aim to prevent. This ethical dilemma is akin to Teitcher et al.'s adjacent question of how much information should researchers disclose to participants in terms of validation methods (e.g., IP tracking), without assisting fraudulent responses or dissuading eligible participants? They ultimately advocate for an intermediary approach that informs participants of validation measures without going into the specifics [17]; however, such disclosure levels belong to the personal discretion of individual research teams and their IRB consultants. Future work might explore ways for researchers to achieve this balance between informing the community about trends in fraudulent activity without also empowering said activity.

6 CONCLUSION

Online recruitment and study facilitation tools provide researchers with the ability to recruit diverse populations who otherwise would not participate in research; however, they also increase the possibility of fraudulent participants engaging in studies. In this case study, we detail our experiences and our holistic, mixed methods approach to identifying fraudulent participants through metadata, cross-referencing, consistency checks, and triangulation through follow-up surveys. We provide the HCI community with five lessons learned and encourage researchers to plan time into their study schedule to accommodate extra care in study design, data cleaning, and validation. Our goal was to not only raise awareness and share techniques, but contribute to the CHI culture of transparency.

ACKNOWLEDGMENTS

We want to thank all the participants who allowed us to engage with their stories during this work, as well as the Indiana University Cox Scholars Program, which helped fund Michaela during the beginning stages of this project. We also want to thank our research lab colleagues who provided valuable feedback and support, especially: Colin LeFevre, Cristina Bosco, Zaidat Ibrahim, and Yuxing Wu. This research is based upon work supported by the National Science Foundation (NSF) under Grant No. IIS 1852294 and the NSF Graduate Research Fellowship under Grant No. 2240777. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

REFERENCES

[1] April M. Ballard, Trey Cardwell, and April M. Young. 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 (Feb. 2019), e12344. https://doi.org/10.2196/12344

[2] Jose Bauermeister, Emily Pingel, Marc Zimmerman, Mick Couper, Alex Carballo-Diéguez, and Victor J. Strecher. 2012. Data Quality in web-based HIV/AIDS research: Handling Invalid and Suspicious Data. *Field methods* 24, 3 (Aug. 2012), 272–291. https://doi.org/10.1177/1525822X12443097

[3] Anne M. Bowen, Candice M. Daniel, Mark L. Williams, and Grayson L. Baird. 2008. Identifying multiple submissions in Internet research: preserving data integrity. *AIDS and behavior* 12, 6 (Nov. 2008), 964–973. https://doi.org/10.1007/s10461-007-9352-2

[4] Kate Cockrill, Ushma D. Upadhyay, Janet Turan, and Diana Greene Foster. 2013. The Stigma of Having an Abortion: Development of a Scale and Characteristics of Women Experiencing Abortion Stigma. *Perspectives on Sexual and Reproductive Health* 45, 2 (2013), 79–88. https://doi.org/10.1363/4507913 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1363/4507913.

[5] James Dewitt, Benjamin Capistrant, Nidhi Kohli, B. R. Simon Rosser, Darryl Mitteldorf, Enyinnaya Merengwa, and William West. 2018. Addressing Participant Validity in a Small Internet Health Survey (The Restore Study): Protocol and Recommendations for Survey Response Validation. *JMIR Research Protocols* 7, 4 (April 2018), e7655. https://doi.org/10.2196/resprot.7655

[6] Jillian V. Glazer, Kirsten MacDonnell, Christina Frederick, Karen Ingersoll, and Lee M. Ritterband. 2021. Liar! Liar! Identifying eligibility fraud by applicants in digital health research. *Internet Interventions* 25 (Sept. 2021), 100401. https://doi.org/10.1016/j.invent.2021.100401

[7] Abigail Jones, Line Caes, Tessa Rugg, Melanie Noel, Sharon Bateman, and Abbie Jordan. 2021. Challenging issues of integrity and identity of participants in non-synchronous online qualitative methods. *Methods in Psychology* 5 (2021), 100072. https://doi.org/10.1016/j.metip.2021.100072

[8] K. Cassie Kresnye, Mona Y. Alqassim, Briana Hollins, Lucia Guerra-Reyes, Maria K. Wolters, and Katie A. Siek. 2020. What to Expect When You are No Longer Expecting: Information Needs of Women who Experienced a Miscarriage. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, Atlanta GA USA, 85–96. https://doi.org/10.1145/3421937.3421995

[9] Amanda Lazar, Ben Jelen, Alisha Pradhan, and Katie A. Siek. 2021. Adopting Diffractive Reading to Advance HCI Research: A Case Study on Technology for Aging. *ACM Transactions on Computer-Human Interaction* 28, 5 (Aug. 2021), 32:1–32:29. https://doi.org/10.1145/3462326

[10] Haley MacLeod, Ben Jelen, Annu Prabhakar, Lora Oehlberg, Katie Siek, and Kay Connelly. 2017. A Guide to Using Asynchronous Remote Communities (ARC) for Researching Distributed Populations. *EAI Endorsed Transactions on Pervasive Health and Technology* 3, 11 (July 2017), 152898. https://doi.org/10.4108/eai.18-7-2017.152898

[11] Juan F. Maestre, Haley MacLeod, Ciabhan L. Connelly, Julia C. Dunbar, Jordan Beck, Katie A. Siek, and Patrick C. Shih. 2018. Defining Through Expansion: Conducting Asynchronous Remote Communities (ARC) Research with Stigmatized Groups. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174131

[12] Rachel Pozzar, Marilyn J. Hammer, Meghan Underhill-Blazey, Alexi A. Wright, James A. Tulsky, Fangxin Hong, Daniel A. Gundersen, and Donna L. Berry. 2020. Threats of Bots and Other Bad Actors to Data Quality Following Research Participant Recruitment Through Social Media: Cross-Sectional Questionnaire. *Journal of Medical Internet Research* 22, 10 (Oct. 2020), e23021. https://doi.org/10.2196/23021

[13] Mandi Pratt-Chapman, Jenna Moses, and Hannah Arem. 2021. Strategies for the Identification and Prevention of Survey Fraud: Data Analysis of a Web-Based Survey. *JMIR cancer* 7, 3 (July 2021), 100072. https://doi.org/10.2196/30730

[14] Damien Ridge, Laurna Bullock, Hilary Causer, Tamsin Fisher, Samantha Hider, Tom Kingstone, Lauren Gray, Ruth Riley, Nina Smyth, Victoria Silverwood, Johanna Spiers, and Jane Southam. 2023. 'Imposter participants' in online qualitative research, a new and increasing threat to data integrity? *Health Expectations* 26, 3 (2023), 941–944. https://doi.org/10.1111/hex.13724 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/hex.13724.

[15] Jacqueline Roehl and Darci Harland. 2022. Imposter Participants: Overcoming Methodological Challenges Related to Balancing Participant Privacy with Data Quality When Using Online Recruitment and Data Collection. *The Qualitative Report* 27, 11 (Nov. 2022), 2469–2485. https://doi.org/10.46743/2160-3715/2022.5475

[16] Margaret R. Salinas. 2023. Are Your Participants Real? Dealing with Fraud in Recruiting Older Adults Online. *Western Journal of Nursing Research* 45, 1 (Jan. 2023), 93–99. https://doi.org/10.1177/01939459221098468

[17] Jennifer E. F. Teitcher, Walter O. Bockting, José A. Bauermeister, Chris J. Hoefer, Michael H. Miner, and Robert L. Klitzman. 2015. Detecting, Preventing, and Responding to "Fraudsters" in Internet Research: Ethics and Tradeoffs. *Journal of Law, Medicine & Ethics* 43, 1 (April 2015), 116–133. https://doi.org/10.1111/jlme.12200 Publisher: Cambridge University Press.

[18] Ziyi Zhang, Shuofei Zhu, Jaron Mink, Aiping Xiong, Linhai Song, and Gang Wang. 2022. Beyond Bot Detection: Combating Fraudulent Online Survey Takers. In *Proceedings of the ACM Web Conference 2022 (WWW '22)*. Association for Computing Machinery, New York, NY, USA, 699–709. https://doi.org/10.1145/3485447.3512230