Review

Imposters, Bots, and Other Threats to Data Integrity in Online Research: Scoping Review of the Literature and Recommendations for Best Practices

Isabella B Strickland[1]; Amy K Ferketich[1,2], PhD; Alayna P Tackett[2,3], PhD; Joanne G Patterson[1], PhD; Nicholas J K Breitborde[4], PhD; Jade Davis[1]; Megan Roberts[1,2], PhD

[1]College of Public Health, The Ohio State University, Columbus, OH, United States
[2]Center for Tobacco Research, The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, OH, United States
[3]Division of Medical Oncology, Department of Internal Medicine, College of Medicine, The Ohio State University, Columbus, OH, United States
[4]Psychiatry and Behavioral Health, College of Medicine, The Ohio State University, Columbus, OH, United States

Corresponding Author:

Megan Roberts, PhD
College of Public Health
The Ohio State University
1841 Neil Ave
Columbus, OH 43210
United States
Phone: 1 6142924647
Email: roberts.1558@osu.edu

Abstract

Background: Threats to data integrity have always existed in online human subjects research, but it appears these threats have become more common and more advanced in recent years. Researchers have proposed various techniques to address satisficers, repeat participants, bots, and fraudulent participants; yet, no synthesis of this literature has been conducted.

Objective: This study undertakes a scoping review of recent methods and ethical considerations for addressing threats to data integrity in online research.

Methods: A PubMed search was used to identify 90 articles published from 2020 to 2024 that were written in English, that discussed online human subjects research, and that had at least one paragraph dedicated to discussing threats to online data integrity.

Results: We cataloged 16 types of techniques for addressing threats to online data integrity. Techniques to authenticate personal information (eg, videoconferencing and mailing incentives to a physical address) appear to be very effective at deterring or identifying fraudulent participants. Yet such techniques also come with ethical considerations, including participant burden and increased threats to privacy. Other techniques, such as Completely Automated Public Turing test to tell Computers and Humans Apart (reCAPTCHA; Google LLC), scores, and checking IP addresses, although very common, were also deemed by several researchers as no longer sufficient protections against advanced threats to data integrity.

Conclusions: Overall, this review demonstrates the importance of shifting online research protocols as bots and fraudulent participants become more sophisticated.

Online J Public Health Inform 2025;17:e70926; doi: 10.2196/70926

Keywords: review; fraud; data integrity; bots; online data collection; PRISMA

Introduction

Recent years have witnessed a shift in research protocols, with many studies that were previously conducted in-person being moved online [1]. This shift has had several benefits for researchers in terms of easier sampling, broader reach, and better access to historically marginalized populations [2,3]. However, the shift has also ushered in critical concerns about threats to data integrity. While concerns about data integrity have always existed, even with in-person studies, there has been a notable increase in the number and types of threats to

data integrity in online studies since the COVID-19 pandemic [4,5].

As outlined in Table 1, the types of threats to data integrity in online research come in several forms. First, satisficers (also known as speeders, straightliners, and cheaters) are individuals who rush through surveys with little care for the accuracy or thoroughness of their responses. While satisficers also exist with study protocols administered in-person or by mail [6,7], this threat is challenging to monitor online. Next, there are repeat participants (also known as duplicate participants). These are individuals who complete screener surveys or study protocols multiple times. Motivations for repeat participation vary: some people may be curious about what happens if they complete the survey with different answers [8]. Others may try to complete the survey multiple times to get extra compensation [9]. Whatever the motivation, this behavior can have serious consequences for data integrity and research findings [10]. And, again, it is often easier for individuals to engage in these behaviors with online research studies.

Table 1. Types of threats to data integrity in online research.

Type of threat	Other terms	Definition
Satisficers	Cheaters, straightliners, speeders, and careless participants	Inattentive participants who speed through surveys often not paying attention to questions and responding thoughtlessly
Repeat participants	Duplicate participants	Participants who attempt to complete a study more than once out of curiosity or a desire for additional remuneration
Bots	Chatbots and artificial intelligence respondents	Computer algorithms deployed on studies in order to gain compensation without human effort for completion
Fraudulent participants	Imposters, scammers, bad actors, and lying participants	Participants who lie about their identity or otherwise attempt to deceive researchers often with the intent of gaining study compensation

Another growing threat to data integrity is bots (also known as chatbots and artificial intelligence respondents). Bots are automated computer programs that people can create to randomly and methodically complete online surveys, usually numerous times, to gain compensation without having to take the time and effort to manually complete the survey [11]. Tools such as a Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA; or the newer version, reCAPTCHA; Google LLC) can help prevent bots from infiltrating surveys; however, more advanced bots can bypass these measures [12]. Bots can very quickly complete surveys and, if they are not properly blocked, can compromise results and force time-consuming and expensive relaunches of research projects [13,14].

Finally, there are fraudulent participants (also known as imposter participants, scammers, and lying participants) [4,5,15]. Unlike bots, fraudulent participants are real people who complete a study protocol. However, they lie about themselves to qualify for study participation. For example, Pellicano et al [4] describe a situation where fraudulent participants posed as either people with autism or parents of children with autism during online, qualitative interviews. In this particular example, several clues aroused the researcher's suspicion, such as keeping cameras off, inconsistent responses between prescreening and the interview, similarities in voices and mannerisms across interviews, and repeated inquiries about payments. Fraudulent participants have been detected in many domains of research but seem, concerningly, to have the largest impact on research on small populations that are often historically minoritized or otherwise vulnerable [10]. One study interviewing research participants found that, on average, 55% of participants who had participated in some sort of research fraud reported fabricating information to qualify for studies [16]. Several studies have found that these participants often respond differently than authentic participants, potentially influencing the results of research studies or weakening the effects detected [17-19].

While many researchers have published concerns or potential solutions to these various threats to data integrity, there has not yet been a review or synthesis of the literature. This research gap makes it difficult and time-consuming for researchers designing new online studies to decide on best practices. Due to the research interests of the authors, we were particularly motivated to identify research methods being used to address threats to online data integrity in the medical and public health domains. Therefore, we used PubMed to conduct a scoping review of methods that address contemporary threats to online data integrity, with keywords that focus on bots and fraudulent participants. Our objectives were to catalog and evaluate the most common research methods used to address these threats and discuss the ethical considerations raised about the techniques. Ultimately, this review aims to expand and centralize knowledge on addressing threats to data integrity in online studies, with the goal of aiding researchers in developing robust online methodologies.

Methods

Search Strategy

This search was conducted using Covidence (Veritas Health Innovation), an online tool used to conduct and organize literature reviews. We searched PubMed for the terms "fraud* OR imposter* OR scam* OR bot OR bots." To be eligible for review, papers needed to be published in or after 2020 because we were interested in methods for addressing

the recent threats to data integrity that have emerged since COVID-19. Additional eligibility criteria were: discussing an online study, using human participants, being written in English, and having at least one paragraph dedicated to discussing threats to data integrity.

Our PubMed search yielded an initial 10,681 publications. After eliminating 189 duplicate texts using Covidence, 10,492 publications remained to be screened. Our first phase of screening checked all abstracts and eliminated those not pertaining to online research. Of the 10,492 publications, 196 were retained. In the second phase of screening, the full article was checked for inclusion criteria. Ultimately, 90 articles met criteria for inclusion in this review [2-4,11,13,15,20-103]. This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines.

Data Extraction

A codebook was developed through an interrogative process. First, a codebook was created a priori according to our research questions. Additional codes and subcodes were added through an inductive process. Two team members (IBS and JD) independently reviewed and coded each of the 90 articles [2-4,11,13,15,20-103]. A senior team member (MR) reviewed their interrater agreement and resolved all discrepancies.

The following information was extracted and coded from the final 90 articles [2-4,11,13,15,20-103]: article type (eg, original research and commentary), type of data collection (eg, qualitative and quantitative), methodology of the study (eg, survey and qualitative interview), country where the study was conducted, recruitment methods (eg, social media and survey service platform), type of suspected threat to data integrity (eg, bots and fraudulent participants), the estimated prevalence of compromised data, and techniques mentioned for addressing threats to data integrity (eg, authenticating personal information and attention checks).

Techniques to address threats to data integrity were additionally sorted into 3 categories. “Very effective” techniques were those that authors of the reviewed articles, especially in the most recent publications, deemed to be successful at identifying poor-quality data, such as bots and fraudulent participants. “Somewhat effective” techniques were those considered capable of identifying a proportion of poor-quality data but that had drawbacks preventing them from being used alone. “No longer effective” techniques were those deemed by the authors of the reviewed articles as being no longer sufficient in addressing threats to data integrity.

When synthesizing the data, we computed the most common types of threats to data integrity, the most common recruitment methods, and the estimated prevalence of threats to data integrity. We also narratively reviewed how authors discussed the adverse effects of threats to data integrity. Next, to address our study objectives, we described all the proposed techniques to address threats to data integrity that we uncovered in this review. Finally, we narratively reviewed how authors discussed the ethical considerations raised by the techniques.

Results

A total of 90 studies were included in the review (Figure 1). The most common type of threat to data integrity documented by researchers was bots (n=59) followed by fraudulent participants (n=51), repeat participants (n=42), and satisficing participants (n=17). The most common recruitment method was social media advertising, followed by using online survey service platforms, such as MTurk (Amazon Web Services, Inc; Table S1 in Multimedia Appendix 1 [2-4,11,13,15,20-103]). Many studies used more than one recruitment method.

The estimated prevalence of threats to data integrity ranged from approximately 1% to 99%. Implications for data validity and reliability were commonly discussed. Other adverse effects included the heavy, and often wasteful, use of resources needed to address fraud. For example, some researchers with a high prevalence of participant fraud described having to end their study and start over, wasting valuable time and resources. Some articles even discussed how dealing with high proportions of imposter participants can be difficult to handle emotionally as researchers. As one research team expressed after finding around 90% of their study participants to be fraudulent: “It is disheartening to encounter issues related to fraud during research. Our team experienced significant demoralization related to this occurrence” [20].

All of the articles provided information on methods to improve the integrity of data, either by (1) preventing the collection of poor-quality data in the first place or (2) identifying and removing poor-quality data if collected (or both; Table S2 in Multimedia Appendix 2 [2-4,11,13,15,20-103]). As cataloged in Table 2, we identified 16 techniques, representing a wide variety of methods. Techniques deemed to be “very effective” included authenticating personal information, such as requesting to see participants’ IDs over a video call, which eliminates the potential for bots and helps identify fraudulent participants. As another version of this technique, Hardesty et al [21] mailed the study incentives to participants’ street addresses (rather than sending the incentive electronically) because they observed that fraudulent participants were providing false addresses to meet geographically based eligibility criteria. A related technique deemed “very effective” was including background-related questions that could be easily answered by participants in the target population (or by partners in dyad research) but that are not widely known by other groups. For example, in a community study on narcolepsy, data were excluded from participants who reported unlikely symptomology [22]. A final “very effective” technique concerned data checking: cross-checking for inconsistent answers (eg, between screening and a baseline survey). Of note, we observed that these very effective techniques were used across a variety of study designs, including both quantitative and qualitative studies.

Figure 1. Flow chart of the review process for article selection.

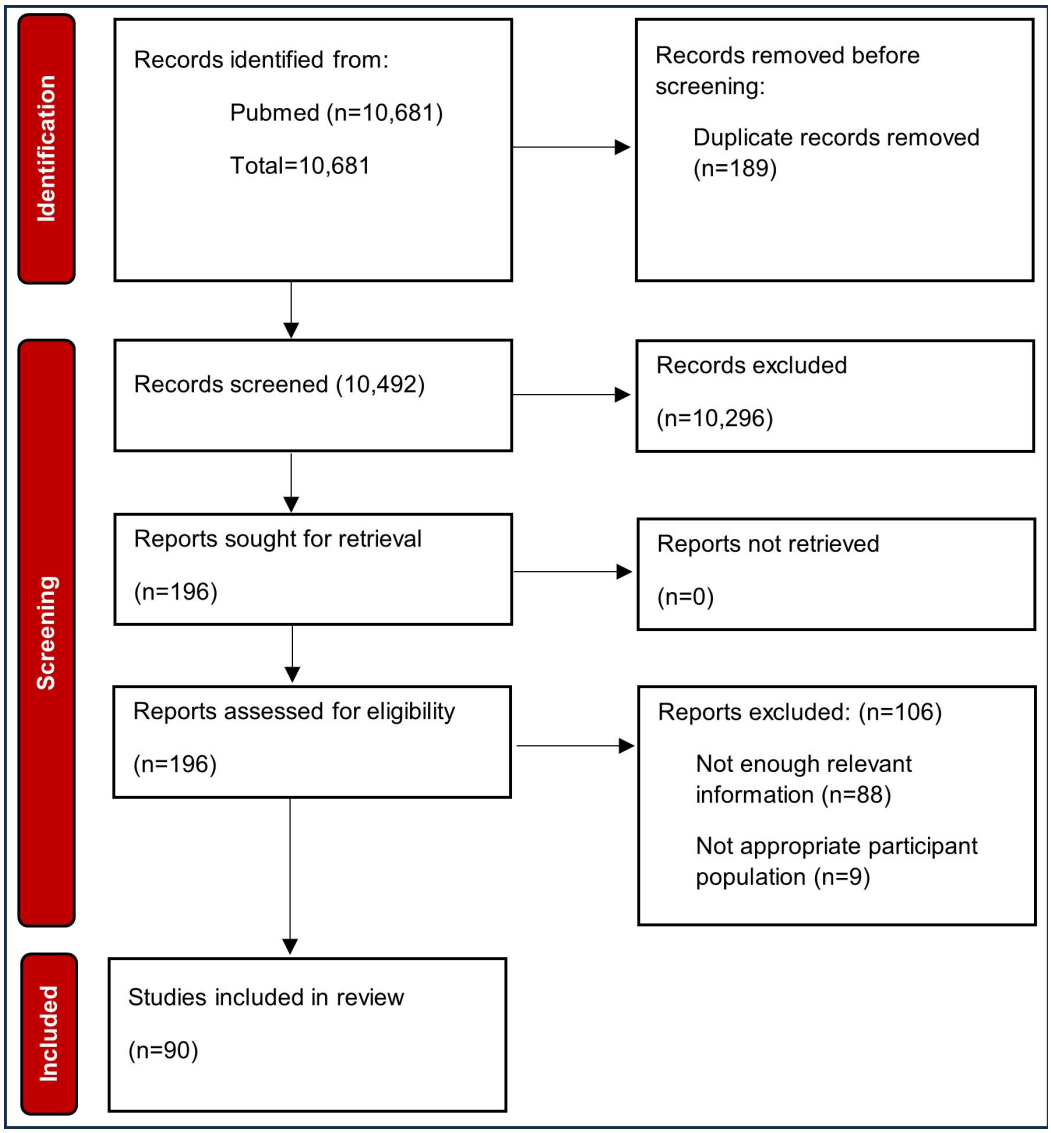

Table 2. Techniques to address threats to data integrity in online research and their frequency of being mentioned across the 90 articles examined in this scoping review.

Effectiveness ranking and technique	Description	Example	Freq.[a]
Very effective			
Authenticate personal information	Checking IDs, emails, addresses, zip codes, and phone numbers for authenticity. This could include using third-party services to verify identities, requiring video calls at enrollment (verification step), or mailing incentives to the provided street address.	"Include a preinterview briefing over videoconferencing or telephone to go through eligibility criteria and the consent process. Researchers could forewarn potential participants about this aspect in the consent form." [4]	48.9%
Background-related questions	Including questions about information that would be easily answered by participants in target population but is not widely known by other groups.	"To reduce fraudulent responses, the study investigators added 4 military validation questions to confirm history of military service prior to the study survey. These questions were developed and piloted with service members and veterans of varying components and across branches." [23]	20.0%
Cross-check inconsistent answers	Checking for inconsistent or contradictory answers across survey items to detect fraud or inattention.	"By identifying inconsistencies in data collected at screening and survey data, the team could identify potentially fraudulent or ineligible participants." [24]	51.1%

Effectiveness ranking and technique	Description	Example	Freq.[a]
Somewhat effective			
Attention checks	Including survey questions that request specific answers or that may only have one reasonable answer. This screens for satisficing and basic logical reasoning.	"The attention checks consisted of the following: (1) embedded on the Grit scale—"Select 'Somewhat like me' for this statement," (2) embedded on the Beck Depression Inventory—"1 – Select this option," and (3) embedded on the Borderline Personality Inventory—"Select 'Yes' for this statement.'" [25]	33.3%
Camera-on requirement	Requiring participants to turn on their camera, even if just for a moment, as often fraudulent participants will leave theirs off.	"Participants were not using their cameras for the Zoom sessions because they refused to turn on their camera or they stated there were internet issues." [26]	10.0%
Check for dataset duplicates	Checking a dataset for duplicate names, emails, etc, across participants for duplicate replies.	"SAS programs were run to check the newly submitted record against all previous baseline questionnaires to check for duplicates of email addresses, mobile numbers, IP addresses, mailing addresses, social media handles, and preferred names." [24]	46.7%
Post hoc testing	Conducting statistical analysis of data for unreasonable response patterns and notable outliers that may be indicative of fraud.	"Interactive visualization can improve data quality by facilitating the identification of issues such as missing data, outliers, duplicates, pattern or constraint violations, and data inconsistencies." [27]	16.7%
Watchful of a large number of responses	Checking the timestamps on survey submissions. A flurry of responses or sign-ups can often be an indication of bots or fraud.	"Before launching the DIP, various indications of fraudulent activity were noted. These include…a rush of survey time stamps…found in the same 1- to 15-min period." [28]	30.0%
Screen for low response rates	Excluding data from participants who complete less than a certain percentage of a study, as they may be satisficing.	"Frequently examine the data for any patterns such as large blocks of blank question." [29]	8.9%
Changing payment protocols	Intentionally not emphasizing and not automating participant payments. For example, not including the payment amount in recruitment materials.	"To maximize reach and limit fraud, gift cards could be manually distributed via text or email after each survey is verified." [30]	21.1%
Not paying fraudulent respondents	Informing potential participants in the consent form that fraudulent participants will not be compensated. This helps researchers not waste money on fraudulent participants.	"On the consent form, participants were also informed that "…we have put in place a number of safeguards to ensure that participants provide valid and accurate data for this study. If we have strong reason to believe your data are invalid, your responses will not be approved or paid and your data will be discarded." [31]	11.1%
No longer effective			
IP address or Geolocation	Examining participant IP and geolocation to see if they match location requirements of study and screening for duplicate IP addresses.	"[Researchers used] other survey platform features to track IP addresses, geolocation, latitude and longitude, and participants' postal codes when they discovered that geographic markers or indicators did not match the participants' stated location of residence." [32]	65.6%
(re) CAPTCHA	Including tests that can help to screen out bots by providing challenges that theoretically only humans can complete.	"Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) script was created and implemented into the Google Form." [33]	45.6%
Timing checks	Checking for unusually fast or slow response times, which can indicate bots or satisficing.	"We noticed a large proportion of responses with improbably fast completion times (as well as those with particularly long completion times, eg, 4220 min)." [34]	45.6%
Open-ended questions	Including open-ended questions and reviewing the responses. This	"Another indicator of data quality is suspicious responses to open-ended	37.8%

Effectiveness ranking and technique	Description	Example	Freq.[a]
	can help assess attention as well as check for bots who may incoherently respond.	questions. For example, when given an open response box to report thoughts or ask questions at the end of the survey, responses written in all caps, one-word responses seemingly unrelated to the prompt, restatements of parts of the question, or nonsensical phrases." [35]	
Honeypots	Incorporating questions into only the code of a survey, such that they are not visible to the human eye. These questions would only be answered by bots.	"We added a honeypot question as a second line of defense against bots. Honeypots are survey questions hidden from rendering on the screen using custom JavaScript code." [36]	14.4%

[a]Freq.: frequency of articles mentioning this technique.

Several techniques were deemed "somewhat effective," as they had notable benefits as well as limitations. For example, attention check questions help detect satisficers and some types of bots but are ineffective against fraudulent participants. Many studies noted suspicion when a large number of surveys were completed at once, as that can indicate their study has been "discovered" by bots or fraudulent respondents; thus, being watchful of a large number of responses is a useful technique but is not sufficient for detecting all cases of threats to data integrity. Another technique, changing payment protocols (eg, intentionally not emphasizing participant payments in recruitment materials), was framed as a preventative measure rather than a definitive means of detecting fraud.

Finally, several techniques were deemed "no longer effective" by the authors: IP address and geolocation checks, reCAPTCHA, timing checks (ie, checking for unusually fast or unusually slow responses), open-ended questions, and honeypot questions (ie, questions not visible to the human eye but would be seen and answered by bots). Although many studies still report using these techniques, many authors also discussed how such methods can now be easily bypassed. For example, reCAPTCHA is not a challenge for most advanced bots, and certainly not for fraudulent participants. Likewise, proxy servers can help "fake" a local IP address. The invention of ChatGPT (OpenAI) and other artificial intelligence natural language processing chatbots makes short answer questions a less effective means of screening, as bots are often able to respond coherently to open-ended questions. Some bots can also be trained to complete surveys in a realistic timeframe and are also able to overlook honeypot questions. Therefore, although they are still somewhat useful for removing simpler bots and fraud attempts, these "no longer effective" techniques are unable to catch or detect more sophisticated attacks and should not be overly relied upon.

While discussing techniques to improve data integrity, many authors reflected on ethical considerations. For instance, as most "very effective" techniques require asking both personal and personally identifiable questions, additional procedures to protect participant privacy and rights may be necessary [37]. Another key ethical concern expressed by researchers was mistakenly excluding genuine participants. For example, fraud detection methods have the potential to introduce selection bias, such as when blocking responses from the same IP address deters residents of high-density housing developments [24]. Deterring genuine participants is also a major concern; for example, many techniques, such as requiring a video call at screening, can place additional burdens on participants and feel invasive [4]. Similarly, verification techniques that convey doubt to participants about their genuineness can compromise trust [38]. As a counter to these concerns, other authors discussed how techniques such as videoconferencing, when conducted with sensitivity, can foster and strengthen rapport and help researchers better understand their participant population [39]. Ultimately, thoughtful study design was encouraged, such as urging researchers to reflect on how to best balance their needs of (1) research integrity and quality; (2) feasibility and efficiency; and (3) safeguarding participants' rights, safety, and privacy [38]. Often, somewhat minor changes can help work toward this balance. For example, Singh and Sagar [37] encourage methods such as deidentifying data, using encryption processes or password-protected data storage, and using HIPAA (Health Insurance Portability and Accountability Act)-compliant online survey platforms. And Roehl and Harland [5] emphasized the importance of transparency during the consent process, so that participants are aware of what identifiable information will be requested from them and why.

Discussion

Principal Findings

Online research is expanding and holds great promise for innovative and impactful research. But as techniques to protect data integrity advance, so too do the methods of mendacious individuals providing false or unreliable responses for monetary gain. In this scoping review, we identified 90 articles published since 2020 that described methods for addressing online threats to data integrity [2-4,11,13,15,20-103]. We found that some of the most common techniques discussed were IP Address or Geolocation checks and reCAPTHCA. This is concerning, given that several articles detailed the reasons these techniques are no longer effective against sophisticated bots or fraudulent

participants. Overall, these findings reveal a crucial area for improvement in handling threats to online data integrity. Yet our review also discovered new and innovative techniques for addressing threats to online data integrity. Specifically, we found that authenticating personal information, posing background-related questions, and cross-checking inconsistent answers were deemed very effective techniques for addressing contemporary threats.

Recommendations

While there is no one foolproof way for researchers to prevent participant fraud, it is clear from this review that the field has moved beyond reCAPTCHA as a sufficient technique for ensuring data integrity. Bots are advancing and fraudulent participants are becoming more sophisticated, making reCAPTCHA ill-equipped to handle the scope of the current problem. We recommend that researchers engaging in online data collection develop a robust strategy for ensuring data integrity early in the design of their research protocol. For such designs, we recommend that researchers use multiple techniques (rather than relying on the soundness of just one technique), use the techniques described in the articles reviewed here, and draw the most from techniques deemed to be "very effective." Although the constraints of timelines, person power, and budgets can render some techniques unfeasible, many "very effective" techniques are efficient and low cost; for example, researchers can think creatively to develop background-related questions that, even with internet searching, would only be readily known and (or) accurately answered by their population. When relying on survey service platforms for access to online samples, researchers should be critical of the techniques used by the platforms to guarantee the quality of their panels. We also suggest that techniques for ensuring data integrity should be critically considered by journal editors, journal reviewers, and grant reviewers when evaluating the rigor of study methods.

The ethical concerns discussed by the articles in this review highlight the responsibilities of researchers to continue focusing on participant rights and privacy. Techniques for ensuring data integrity (eg, personal questions to authenticate identity) should be balanced against these responsibilities. More broadly, the comfort of participants and their rapport with the study team should be considered. The relative weight of these considerations will vary depending on many factors. For instance, more complex study designs (eg, longitudinal studies that rely on participant trust and investment for good retention) may require techniques that integrate authentication checks with rapport-building. The vulnerability of the sample and the sensitivity of the research topic must also be considered. Techniques for improving data integrity that increase participant burden, barriers, or privacy risk must be matched with greater participant accommodations—this could include greater compensation, clear explanations to participants during consent and enrollment about why certain questions are being asked, and enhanced data protection measures. Researchers attempting to exclude fraudulent participants should always be aware of their own biases and ensure that they are not excluding participants simply because they do not align with expected results.

Limitations and Future Directions

Our scoping review has some limitations. First, we only used one database for review (PubMed). This was done due to the large volume of articles on the subject and our focus on health research; however, it may have overlooked insights from other fields. Second, our search terms led to the studies reviewed primarily focusing on bots, fraudulent participants, and repeat participants, likely leading to an overestimation of these behaviors and an underestimation of the prevalence of participant satisficing and the tactics used to mitigate those behaviors. It is important to acknowledge that moving forward, some tactics that are currently in our "very effective" category may become less effective with the evolution of artificial intelligence, or as fraudulent participants become more familiar with current strategies. Going forward, more empirical studies should be conducted on research methods for addressing threats to data integrity, to quantitatively compare the effectiveness of various techniques used to address threats to data integrity in online research. Future work should also consider the participant perspective on these various techniques in order to improve their effectiveness and minimize negative consequences.

Conclusions

Threats to data integrity appear to be on the rise, particularly with online research, and numerous solutions and prevention strategies have been recommended. In order to aid researchers in developing robust online methodologies, this scoping review discusses the most common types of threats to data integrity, synthesizes the most common prevention methods, and discusses the ethical considerations raised about the techniques. Doubtless, new threats to data integrity will continue to emerge, and researchers should continue developing the most effective methods in response.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute and Food and Drug Administration's Center for Tobacco Products (CTP; grant U54CA287392). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Food and Drug Administration.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

IBS contributed to conceptualization (lead), data curation (lead), formal analysis (equal), investigation (lead), methodology (equal), and writing—original draft (lead). AKF was responsible for funding acquisition (equal), methodology (supporting),

and writing—review and editing (equal). APT, JGP, and NJKB contributed to methodology (supporting) and writing—review and editing (equal). JD participated in formal analysis (equal) and writing—review and editing (equal). MER contributed to conceptualization (supporting), funding acquisition (equal), formal analysis (equal), investigation (supporting), methodology (supporting), supervision (lead), and writing—review and editing (lead).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Study details described in the 90 manuscripts examined in this scoping review.
[DOCX File (Microsoft Word File), 57 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Techniques for addressing threats to data integrity in online research as mentioned in the 90 manuscripts examined in this scoping review.
[DOCX File (Microsoft Word File), 63 KB-Multimedia Appendix 2]

Checklist 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File (Microsoft Word File), 33 KB-Checklist 1]

References

1. Maheu C, Lemonde M, Mayo S, Galica J, Bally J. Moving research forward during COVID-19. Can Oncol Nurs J. 2021;31(4):490-492. [Medline: 34786468]
2. Myers KJ, Jaffe T, Kanda DA, et al. Reaching the "Hard-to-Reach" sexual and gender diverse communities for population-based research in cancer prevention and control: methods for online survey data collection and management. Front Oncol. 2022;12:841951. [doi: 10.3389/fonc.2022.841951] [Medline: 35756657]
3. Upadhyay UD, Jovel IJ, McCuaig KD, Cartwright AF. Using Google Ads to recruit and retain a cohort considering abortion in the United States. Contracept X. 2020;2:100017. [doi: 10.1016/j.conx.2019.100017] [Medline: 32550532]
4. Pellicano E, Adams D, Crane L, et al. Letter to the Editor: a possible threat to data integrity for online qualitative autism research. Autism. Mar 2024;28(3):786-792. [doi: 10.1177/13623613231174543] [Medline: 37212144]
5. Roehl JM, Harland DJ. Imposter participants: overcoming methodological challenges related to balancing participant privacy with data quality when using online recruitment and data collection. Qual Rep. 2022;27:2469-2485. [doi: 10.46743/2160-3715/2022.5475]
6. Krosnick JA, Alwin DF. An evaluation of a cognitive theory of response-order effects in survey measurement. Public Opin Q. 1987;51(2):201. [doi: 10.1086/269029]
7. Suskie LA. Survey research: what works for the institutional researcher. institutional research information series no. 1. North East Association for Institutional Research, C/O Larry Metzger, Ithaca College; 1988.
8. Bauermeister J, Pingel E, Zimmerman M, Couper M, Carballo-Diéguez A, Strecher VJ. Data quality in web-based HIV/AIDS research: handling invalid and suspicious data. Field Methods. Aug 1, 2012;24(3):272-291. [doi: 10.1177/1525822X12443097] [Medline: 23180978]
9. Teitcher JEF, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. J Law Med Ethics. 2015;43(1):116-133. [doi: 10.1111/jlme.12200] [Medline: 25846043]
10. Pullen Sansfaçon A, Gravel E, Gelly MA. Dealing with scam in online qualitative research: strategies and ethical considerations. Int J Qual Methods. Jan 2024;23. [doi: 10.1177/16094069231224610]
11. Lawrence PR, Osborne MC, Sharma D, Spratling R, Calamaro CJ. Methodological challenge: addressing bots in online research. J Pediatr Health Care. 2023;37(3):328-332. [doi: 10.1016/j.pedhc.2022.12.006] [Medline: 36717299]
12. Al-Fannah NM. Making defeating captchas harder for bots 2017. arXiv. Preprint posted online on Apr 10, 2017. URL: https://doi.org/10.48550/arXiv.1704.02803 [Accessed 2025-08-11]
13. Perkel JM. Mischief-making bots attacked my scientific survey. Nature New Biol. Mar 2020;579(7799):461-461. [doi: 10.1038/d41586-020-00768-0] [Medline: 32184487]
14. Storozuk A, Ashley M, Delage V, Maloney EA. Got bots? Practical recommendations to protect online survey data from bot attacks. Quant Method Psychol. 2020;16(5):472-481. [doi: 10.20982/tqmp.16.5.p472]
15. Chandler J, Sisso I, Shapiro D. Participant carelessness and fraud: consequences for clinical research and potential solutions. J Abnorm Psychol. Jan 2020;129(1):49-55. [doi: 10.1037/abn0000479] [Medline: 31868387]

16. Devine EG, Pingitore AM, Margiotta KN, et al. Frequency of concealment, fabrication and falsification of study data by deceptive subjects. Contemp Clin Trials Commun. Mar 2021;21:100713. [doi: 10.1016/j.conctc.2021.100713] [Medline: 33604482]
17. Siegel JT, Navarro MA, Thomson AL. The impact of overtly listing eligibility requirements on MTurk: an investigation involving organ donation, recruitment scripts, and feelings of elevation. Soc Sci Med. Oct 2015;142:256-260. [doi: 10.1016/j.socscimed.2015.08.020] [Medline: 26322721]
18. Siegel JT, Navarro M. A conceptual replication examining the risk of overtly listing eligibility criteria on Amazon's Mechanical Turk. J Applied Social Pyschol. Apr 2019;49(4):239-248. [doi: 10.1111/jasp.12580]
19. Sharpe Wessling K, Huber J, Netzer O. MTurk character misrepresentation: assessment and solutions. J Consum Res. Jun 1, 2017;44(1):211-230. [doi: 10.1093/jcr/ucx053]
20. Gordon JH, Fujinaga-Gordon K, Sherwin C. Fraudulent online survey respondents may disproportionately threaten validity of research in small target populations. Health Expect. Jun 2024;27(3):e14099. [doi: 10.1111/hex.14099] [Medline: 38845165]
21. Hardesty JJ, Crespi E, Nian Q, et al. The vaping and patterns of e-cigarette use research study: protocol for a web-based cohort study. JMIR Res Protoc. Mar 2, 2023;12:e38732. [doi: 10.2196/38732] [Medline: 36862467]
22. Davidson RD, Biddle K, Nassan M, Scammell TE, Zhou ES. The impact of narcolepsy on social relationships in young adults. J Clin Sleep Med. Dec 1, 2022;18(12):2751-2761. [doi: 10.5664/jcsm.10212] [Medline: 35946418]
23. Tannahill HS, Blais RK. Using military screening questions to anonymously recruit post-9/11 era service members and veterans using online survey methods. Mil Med. May 18, 2024;189(5-6):e1282-e1288. [doi: 10.1093/milmed/usad469] [Medline: 38140962]
24. Guest JL, Adam E, Lucas IL, et al. Methods for authenticating participants in fully web-based mobile app trials from the iReach Project: cross-sectional study. JMIR Mhealth Uhealth. Aug 31, 2021;9(8):e28232. [doi: 10.2196/28232] [Medline: 34463631]
25. Webb MA, Tangney JP. Too good to be true: bots and bad data from Mechanical Turk. Perspect Psychol Sci. Nov 2024;19(6):887-890. [doi: 10.1177/17456916221120027] [Medline: 36343213]
26. Mizerek E, Wolf L, Moon MD. Identifying and mitigating fraud when using social media for research recruitment. J Emerg Nurs. Jul 2023;49(4):530-533. [doi: 10.1016/j.jen.2023.04.002] [Medline: 37393079]
27. Chen AT, Komi M, Bessler S, Mikles SP, Zhang Y. Integrating statistical and visual analytic methods for bot identification of health-related survey data. J Biomed Inform. Aug 2023;144:104439. [doi: 10.1016/j.jbi.2023.104439] [Medline: 37419375]
28. Hohn KL, Braswell AA, DeVita JM. Preventing and protecting against internet research fraud in anonymous web-based research: protocol for the development and implementation of an anonymous web-based data integrity plan. JMIR Res Protoc. Sep 12, 2022;11(9):e38550. [doi: 10.2196/38550] [Medline: 36094806]
29. Bybee S, Cloyes K, Ellington L, Baucom B, Supiano K, Mooney K. Bots and nots: safeguarding online survey research with underrepresented and diverse populations. Psychol Sex. 2022;13(4):901-911. [doi: 10.1080/19419899.2021.1936617] [Medline: 36439051]
30. Salem M, Pollack L, Zepeda A, Tebb KP. Utilization of online systems to promote youth participation in research: a methodological study. World J Methodol. Sep 20, 2023;13(4):210-222. [doi: 10.5662/wjm.v13.i4.210] [Medline: 37771869]
31. Gratz KL, Tull MT, Richmond JR, Edmonds KA, Scamaldo KM, Rose JP. Thwarted belongingness and perceived burdensomeness explain the associations of COVID-19 social and economic consequences to suicide risk. Suicide Life Threat Behav. Dec 2020;50(6):1140-1148. [doi: 10.1111/sltb.12654] [Medline: 32589811]
32. Kumarasamy V, Goodfellow N, Ferron EM, Wright AL. Evaluating the problem of fraudulent participants in health care research: multimethod pilot study. JMIR Form Res. Jun 4, 2024;8:e51530. [doi: 10.2196/51530] [Medline: 38833292]
33. Bisdas S, Topriceanu CC, Zakrzewska Z, et al. Artificial intelligence in medicine: a multinational multi-center survey on the medical and dental students' perception. Front Public Health. 2021;9:795284. [doi: 10.3389/fpubh.2021.795284] [Medline: 35004598]
34. Burnette CB, Luzier JL, Bennett BL, et al. Concerns and recommendations for using Amazon MTurk for eating disorder research. Int J Eat Disord. Feb 2022;55(2):263-272. [doi: 10.1002/eat.23614] [Medline: 34562036]
35. Douglas BD, Ewell PJ, Brauer M. Data quality in online human-subjects research: comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. PLoS One. 2023;18(3):e0279720. [doi: 10.1371/journal.pone.0279720] [Medline: 36917576]
36. Bonett S, Lin W, Sexton Topper P, et al. Assessing and improving data integrity in web-based surveys: comparison of fraud detection systems in a COVID-19 study. JMIR Form Res. Jan 12, 2024;8:e47091. [doi: 10.2196/47091] [Medline: 38214962]

37. Singh S, Sagar R. A critical look at online survey or questionnaire-based research studies during COVID-19. Asian J Psychiatr. Nov 2021;65:102850. [doi: 10.1016/j.ajp.2021.102850] [Medline: 34534919]
38. Wright M, Matheson J, Watson TM, Sproule B, Le Foll B, Brands B. Participant fraud in virtual qualitative substance use research: recommendations and considerations for detection and prevention based on a case study. Subst Use Misuse. 2024;59(8):1261-1270. [doi: 10.1080/10826084.2024.2330892] [Medline: 38503716]
39. Wood NK, Bindler RJ. A videoconferencing verification method for enrollment of breastfeeding dyads to an online prospective mixed methods study during the COVID-19 pandemic. J Adv Nurs. Jul 2024;80(7):2970-2976. [doi: 10.1111/jan.15981] [Medline: 38012846]
40. Agley J, Xiao Y, Nolan R, Golzarri-Arroyo L. Quality control questions on Amazon's Mechanical Turk (MTurk): a randomized trial of impact on the USAUDIT, PHQ-9, and GAD-7. Behav Res Methods. Apr 2022;54(2):885-897. [doi: 10.3758/s13428-021-01665-8] [Medline: 34357539]
41. Bethel C, Rainbow JG, Dudding KM. Recruiting nurses via social media for survey studies. Nurs Res. 2021;70(3):231-235. [doi: 10.1097/NNR.0000000000000482] [Medline: 33060416]
42. Bush J, Blackwell CW. Social media as a recruitment strategy with transgender-identified individuals: using an ethical lens to direct methodology. J Transcult Nurs. Sep 2022;33(5):603-614. [doi: 10.1177/10436596221101928] [Medline: 35699438]
43. Campbell CK, Ndukwe S, Dubé K, Sauceda JA, Saberi P. Overcoming challenges of online research: measures to ensure enrollment of eligible participants. JJ Acquir Immune Defic Syndr. 2022;91(2):232-236. [doi: 10.1097/QAI.0000000000003035] [Medline: 36094490]
44. Davies MR, Monssen D, Sharpe H, et al. Management of fraudulent participants in online research: practical recommendations from a randomized controlled feasibility trial. Int J Eat Disord. Jun 2024;57(6):1311-1321. URL: https://onlinelibrary.wiley.com/toc/1098108x/57/6 [Accessed 2025-08-11] [doi: 10.1002/eat.24085] [Medline: 37921564]
45. Drysdale K, Wells N, Smith AKJ, Gunatillaka N, Sturgiss EA, Wark T. Beyond the challenge to research integrity: imposter participation in incentivised qualitative research and its impact on community engagement. Health Sociol Rev. Sep 2, 2023;32(3):372-380. [doi: 10.1080/14461242.2023.2261433] [Medline: 37786312]
46. Dulin P, Mertz R, Edwards A, King D. Contrasting a mobile app with a conversational chatbot for reducing alcohol consumption: randomized controlled pilot trial. JMIR Form Res. 2022;6(5):e33037. [doi: 10.2196/33037] [Medline: 35576569]
47. Dutra LM, Farrelly MC, Bradfield B, Ridenhour J, Guillory J. Modeling the Probability of Fraud in Social Media in a National Cannabis Survey. RTI Press; 2021.
48. Geiger G, Kiel L, Horiguchi M, et al. Latinas in medicine: evaluating and understanding the experience of Latinas in medical education: a cross sectional survey. BMC Med Educ. Jan 3, 2024;24(1):4. [doi: 10.1186/s12909-023-04982-y]
49. Glazer JV, MacDonnell K, Frederick C, Ingersoll K, Ritterband LM. Liar! Liar! Identifying eligibility fraud by applicants in digital health research. Internet Interv. Sep 2021;25:100401. [doi: 10.1016/j.invent.2021.100401]
50. Godinho A, Schell C, Cunningham JA. Out damn bot, out: recruiting real people into substance use studies on the internet. Subst Abus. Jan 2020;41(1):3-5. [doi: 10.1080/08897077.2019.1691131]
51. Gonzalez JM, Grover K, Leblanc TW, Reeve BB. Did a bot eat your homework? An assessment of the potential impact of bad actors in online administration of preference surveys. PLoS ONE. Oct 5, 2023;18(10):e0287766. [doi: 10.1371/journal.pone.0287766]
52. Gratz KL, Scamaldo KM, Vidaña AG, Richmond JR, Tull MT. Prospective interactive influence of financial strain and emotional nonacceptance on problematic alcohol use during the COVID-19 pandemic. Am J Drug Alcohol Abuse. Jan 2, 2021;47(1):107-116. [doi: 10.1080/00952990.2020.1849248]
53. Griffin M, Martino RJ, LoSchiavo C, et al. Ensuring survey research data integrity in the era of internet bots. Qual Quant. Aug 2022;56(4):2841-2852. [doi: 10.1007/s11135-021-01252-1]
54. Guastaferro K, Abuchaibe V, McCormick KV, et al. Adapting a selective parent-focused child sexual abuse prevention curriculum for a universal audience: a pilot study. PLoS One. May 16, 2024;19(5):e0302982. [doi: 10.1371/journal.pone.0302982]
55. Habib D, Jha N. AIM against survey fraud. JAMIA Open. Oct 8, 2021;4(4):ooab099. [doi: 10.1093/jamiaopen/ooab099]
56. Hartman R, Moss AJ, Rabinowitz I, et al. Do you know the Wooly Bully? Testing era-based knowledge to verify participant age online. Behav Res Methods. Oct 2023;55(7):3313-3325. [doi: 10.3758/s13428-022-01944-y]
57. Hauser DJ, Moss AJ, Rosenzweig C, Jaffe SN, Robinson J, Litman L. Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk. Behav Res Methods. ;55(8):3953-3964. [doi: 10.3758/s13428-022-01999-x]

58. Ikegami K, Yoshimoto Y, Baba H, Sekoguchi S, Ando H, Ogami A. Study protocol and preliminary results of the impact of occupational health workers' activities on their health: Nationwide Prospective Internet-Based Survey. JMIR Form Res. ;6(7):e35290. [doi: 10.2196/35290]
59. Ilagan MJ, Falk CF. Model-agnostic unsupervised detection of bots in a Likert-type questionnaire. Behav Res Methods. ;56(5):5068-5085. [doi: 10.3758/s13428-023-02246-7]
60. Ilagan MJ, Falk CF. Supervised classes, unsupervised mixing proportions: detection of bots in a Likert-type questionnaire. Educ Psychol Meas. Apr 2023;83(2):217-239. [doi: 10.1177/00131644221104220]
61. Karuchit S, Thiengtham P, Tanpradech S, et al. A web-based, respondent-driven sampling survey among men who have sex with men (Kai Noi): description of methods and characteristics. JMIR Form Res. 2024;8:e50812. [doi: 10.2196/50812]
62. Keith MG, McKay AS. Too anecdotal to be true? Mechanical Turk is not all bots and bad data: response to Webb and Tangney (2022). Perspect Psychol Sci. Nov 2024;19(6):900-907. [doi: 10.1177/17456916241234328]
63. Kim M, Lee H, Allison J. Challenges and lessons learned from a mobile health, web-based human papillomavirus intervention for female Korean American College Students: feasibility experimental study. JMIR Form Res. Jan 29, 2020;4(1):e14111. [doi: 10.2196/14111]
64. Kolc KL, Tan YXK, Lo AZY, Shvetcov A, Mitchell PB, Perkes IE. Measuring psychiatric symptoms online: a systematic review of the use of inventories on Amazon Mechanical Turk (mTurk). J Psychiatr Res. Jul 2023;163:118-126. [doi: 10.1016/j.jpsychires.2023.05.027]
65. Lepage S, Conway A, Goodson N, Wicks P, Flight L, Devane D. Online randomised trials with children: a scoping review. PLoS One. May 25, 2023;18(5):e0280965. [doi: 10.1371/journal.pone.0280965]
66. Levi R, Ridberg R, Akers M, Seligman H. Survey fraud and the integrity of web-based survey research. Am J Health Promot. Jan 2022;36(1):18-20. [doi: 10.1177/08901171211037531]
67. Loebenberg G, Oldham M, Brown J, et al. Bot or not? Detecting and managing participant deception when conducting digital research remotely: case study of a randomized controlled trial. J Med Internet Res. 2023;25:e46523. [doi: 10.2196/46523]
68. Lorenzo-Luaces L, Howard J, Edinger A, et al. Sociodemographics and transdiagnostic mental health symptoms in SOCIAL (Studies of Online Cohorts for Internalizing Symptoms and Language) I and II: cross-sectional survey and botometer analysis. JMIR Form Res. 2022;6(10):e39324. [doi: 10.2196/39324]
69. Mitchell JW, Chavanduka TMD, Sullivan S, Stephenson R. Recommendations from a descriptive evaluation to improve screening procedures for web-based studies with couples: cross-sectional study. JMIR Public Health Surveill. May 12, 2020;6(2):e15079. [doi: 10.2196/15079]
70. Mournet AM, Kleiman EM. Internet-based mental health survey research: navigating internet bots on reddit. Cyberpsychol Behav Soc Netw. Feb 1, 2023;26(2):73-79. [doi: 10.1089/cyber.2022.0173]
71. O'Donnell N, Satherley RM, Davey E, Bryan G. Fraudulent participants in qualitative child health research: identifying and reducing bot activity. Arch Dis Child. May 2023;108(5):415-416. [doi: 10.1136/archdischild-2022-325049] [Medline: 36669866]
72. Panesar P, Mayo SJ. "Taking out the trash": strategies for preventing and managing fraudulent data in web-surveys. Can Oncol Nurs J. 2023;33(2):283-284. [Medline: 37152829]
73. Parks AM, Duffecy J, McCabe JE, et al. Lessons learned recruiting and retaining pregnant and postpartum individuals in digital trials: viewpoint. JMIR Pediatr Parent. 2022;5(2):e35320. [doi: 10.2196/35320]
74. Pekarsky C, Skiffington J, Leijser LM, Slater D, Metcalfe A. Social media recruitment strategies to recruit pregnant women into a longitudinal observational cohort study: usability study. J Med Internet Res. 2022;24(12):e40298. [doi: 10.2196/40298]
75. Pozzar R, Hammer MJ, Underhill-Blazey M, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res. 2020;22(10):e23021. [doi: 10.2196/23021]
76. Pratt-Chapman M, Moses J, Arem H. Strategies for the identification and prevention of survey fraud: data analysis of a web-based survey. JMIR Cancer. Jul 16, 2021;7(3):e30730. [doi: 10.2196/30730] [Medline: 34269685]
77. Price M, Hidalgo JE, Kim JN, et al. The cyborg method: a method to identify fraudulent responses from crowdsourced data. Comput Human Behav. Aug 2024;157:108253. [doi: 10.1016/j.chb.2024.108253] [Medline: 38799787]
78. Reed ND, Bull S, Shrestha U, Sarche M, Kaufman CE. Combating fraudulent participation in Urban American Indian and Alaska Native Virtual Health Research: Protocol for Increasing Data Integrity in Online Research (PRIOR). JMIR Res Protoc. 2024;13:e52281. [doi: 10.2196/52281]

79. Ridge D, Bullock L, Causer H, et al. 'Imposter participants' in online qualitative research, a new and increasing threat to data integrity? Health Expect. Jun 2023;26(3):941-944. URL: https://onlinelibrary.wiley.com/toc/13697625/26/3 [Accessed 2025-08-11] [doi: 10.1111/hex.13724]
80. Rodriguez C, Oppenheimer DM. Creating a Bot-tleneck for malicious AI: psychological methods for bot detection. Behav Res Methods. Sep 2024;56(6):6258-6275. [doi: 10.3758/s13428-024-02357-9] [Medline: 38561551]
81. Rodriguez E, Peer K, Fruh V, et al. Digital global recruitment for women's health research: cross-sectional study. JMIR Form Res. 2022;6(9):e39046. [doi: 10.2196/39046]
82. Roman ZJ, Brandt H, Miller JM. Automated bot detection using Bayesian latent class models in online surveys. Front Psychol. 2022;13:789223. [doi: 10.3389/fpsyg.2022.789223]
83. Salinas MR. Are your participants real? Dealing with fraud in recruiting older adults online. West J Nurs Res. Jan 2023;45(1):93-99. [doi: 10.1177/01939459221098468] [Medline: 35587721]
84. Sefcik JS, Hathaway Z, DiMaria-Ghalili RA. When snowball sampling leads to an avalanche of fraudulent participants in qualitative research. Int J Older People Nurs. Nov 2023;18(6):e12572. URL: https://onlinelibrary.wiley.com/toc/17483743/18/6 [Accessed 2025-08-11] [doi: 10.1111/opn.12572] [Medline: 37632269]
85. Sharma P, McPhail SM, Kularatna S, Senanayake S, Abell B. Navigating the challenges of imposter participants in online qualitative research: lessons learned from a paediatric health services study. BMC Health Serv Res. 2020;24(1):724. [doi: 10.1186/s12913-024-11166-x]
86. Silva SSM, Meyer D, Jayawardana M. Detecting possible persons of interest in a physical activity program using step entries: Including a web-based application for outlier detection and decision-making. Biom J. Mar 2020;62(2):414-427. URL: https://onlinelibrary.wiley.com/toc/15214036/62/2 [Accessed 2025-08-11] [doi: 10.1002/bimj.201900008]
87. Simone M, Cascalheira CJ, Pierce BG. A quasi-experimental study examining the efficacy of multimodal bot screening tools and recommendations to preserve data integrity in online psychological research. Am Psychol. 2024;79(7):956-969. [doi: 10.1037/amp0001183]
88. Sosenko FL, Bramley G. Smartphone-based Respondent Driven Sampling (RDS): a methodological advance in surveying small or 'hard-to-reach' populations. PLoS One. 2022;17(7):e0270673. [doi: 10.1371/journal.pone.0270673]
89. Stephenson R, Chavanduka TMD, Sullivan S, Mitchell JW. Correlates of successful enrollment of same-sex male couples into a web-based HIV prevention research study: cross-sectional study. JMIR Public Health Surveill. 2020;6(1):e15078. [doi: 10.2196/15078]
90. Tran NK, Welles SL, Goldstein ND. Geolocation to identify online study-eligible gay, bisexual, and men who have sex with men in Philadelphia, Pennsylvania. Epidemiology. 2023;34(4):462-466. [doi: 10.1097/EDE.0000000000001607]
91. Veldheer S, Whitehead-Zimmers M, Bordner C, et al. Participant preferences for the development of a digitally delivered gardening intervention to improve diet, physical activity, and cardiovascular health: cross-sectional study. JMIR Form Res. 2023;7:e41498. [doi: 10.2196/41498]
92. Venugopal A, Marya A, Ludwig B, Alam MK. Satisficing and bots. Br Dent J. Jan 28, 2022;232(2):68-68. [doi: 10.1038/s41415-022-3891-9]
93. Vu M, Huynh VN, Bednarczyk RA, et al. Experience and lessons learned from multi-modal internet-based recruitment of U.S. Vietnamese into research. PLoS One. 2021;16(8):e0256074. [doi: 10.1371/journal.pone.0256074]
94. Walker LO, Murry N, Longoria KD. Improving data integrity and quality from online health surveys of women with infant children. Nurs Res. 2023;72(5):386-391. [doi: 10.1097/NNR.0000000000000671]
95. Wang J, Calderon G, Hager ER, et al. Identifying and preventing fraudulent responses in online public health surveys: lessons learned during the COVID-19 pandemic. PLOS Glob Public Health. 2023;3(8):e0001452. [doi: 10.1371/journal.pgph.0001452]
96. Wardropper CB, Dayer AA, Goebel MS, Martin VY. Conducting conservation social science surveys online. Conserv Biol. Oct 2021;35(5):1650-1658. URL: https://conbio.onlinelibrary.wiley.com/toc/15231739/35/5 [Accessed 2025-08-11] [doi: 10.1111/cobi.13747]
97. Willis TA, Wright-Hughes A, Skinner C, et al. The detection and management of attempted fraud during an online randomised trial. Trials. 2023;24(1):494. [doi: 10.1186/s13063-023-07517-4]
98. Woolfall K. Identifying and preventing fraudulent participation in qualitative research. Arch Dis Child. Jun 2023;108(6):421-422. [doi: 10.1136/archdischild-2023-325328]
99. Xu J, Tian G, He J, et al. The public's self-avoidance and other-reliance in the reporting of medical insurance fraud: a cross-sectional survey in China. Risk Manag Healthc Policy. 2023;16:2869-2881. [doi: 10.2147/RMHP.S438854] [Medline: 38149180]
100. Xu J, Zhang T, Zhang H, et al. What influences the public's willingness to report health insurance fraud in familiar or unfamiliar healthcare settings? a cross-sectional study of the young and middle-aged people in China. BMC Public Health. 2024;24(1):24. [doi: 10.1186/s12889-023-17581-9]

101. Young AM, Ballard AM, Cooper HLF. Novel recruitment methods for research among young adults in rural areas who use opioids: cookouts, coupons, and community-based staff. Public Health Rep. Nov 2020;135(6):746-755. [doi: 10.1177/0033354920954796]
102. Zhang H, Zhang T, Shi Q, et al. Who is more likely to report medical insurance fraud in the two scenarios of whether it results in a direct loss of individual benefit? A cross-sectional survey in China. Psychol Res Behav Manag. 2022;15:2331-2341. [doi: 10.2147/PRBM.S375823]
103. Zuniga C, Ragosta S, Thompson TA. Recruiting foreign-born individuals who have sought an abortion in the United States: lessons from a feasibility study. Front Glob Womens Health. 2024;4:1114820. [doi: 10.3389/fgwh.2023.1114820]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by Edward Mensah; peer-reviewed by Harsh Maheshwari, Odumbo Oluwole, Suraj Kath; submitted 06.01.2025; final revised version received 24.02.2025; accepted 10.04.2025; published 29.08.2025

Please cite as:
Strickland IB, Ferketich AK, Tackett AP, Patterson JG, Breitborde NJK, Davis J, Roberts M
Imposters, Bots, and Other Threats to Data Integrity in Online Research: Scoping Review of the Literature and Recommendations for Best Practices
Online J Public Health Inform 2025;17:e70926
URL: https://ojphi.jmir.org/2025/1/e70926
doi: 10.2196/70926