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

Isabella B. Strickland, Amy K. Ferketich, Alayna P. Tackett, Joanne G. Patterson, Nicholas J.K. Breitborde, Megan Roberts

Submitted to: Journal of Medical Internet Research
on: January 06, 2025

Table of Contents

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

Isabella B. Strickland[1]; Amy K. Ferketich[1]; Alayna P. Tackett[2]; Joanne G. Patterson[1]; Nicholas J.K. Breitborde[2]; Megan Roberts[1]

[1] The Ohio State University Columbus US
[2] The Ohio State University Wexner Medical Center Columbus US

Corresponding Author:
Megan Roberts

The Ohio State University
1841 Neil Ave
Columbus
US

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 bots, fraudulent participants, repeat participants, and satisficers, yet no review of this literature has been conducted.

Objective: To synthesize and evaluate the recent research published on methods for addressing threats to data integrity in online research.

Methods: We conducted a comprehensive review of the literature addressing threats to data integrity in online research. Ninety articles were ultimately reviewed and coded.

Results: Findings revealed that techniques to authenticate personal information (e.g., videoconferencing, mailing incentives to a physical address) were discussed by 47% of the articles and 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 reCAPTCHA 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.

(JMIR Preprints 06/01/2025:70926)
DOI: https://doi.org/10.2196/preprints.70926

Original Manuscript

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

Isabella B. Strickland,[1] Amy K. Ferketich,[1] Alayna P. Tackett,[2] Joanne G. Patterson,[1] Nicholas J.K. Breitborde,[2] & Megan E. Roberts[1]

[1]College of Public Health, The Ohio State University, Columbus, OH, United States
[2]College of Medicine, The Ohio State University, Columbus, OH, United States

Corresponding author: Megan E. Roberts, The Ohio State University College of Public Health, Columbus, Ohio 43210. Roberts.1558@osu.edu 614-292-4647

Declarations of Interest: None

Funding: Research reported in this publication was supported by NCI and FDA Center for Tobacco Products (CTP) under grant U54CA287392. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the Food and Drug Administration.

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 bots, fraudulent participants, repeat participants, and satisficers, yet no review of this literature has been conducted.
Objective: To synthesize and evaluate the recent research published on methods for addressing threats to data integrity in online research.
Methods: We conducted a comprehensive review of the literature addressing threats to data integrity in online research. Ninety articles were ultimately reviewed and coded.
Results: Findings revealed that techniques to authenticate personal information (e.g., videoconferencing, mailing incentives to a physical address) were discussed by 47% of the articles and 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 reCAPTCHA 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.

Keywords: Review; Fraud; Data Integrity; Bots; Online Data Collection

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* (a.k.a., 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* (a.k.a., 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.

Another growing threat to data integrity is *bots* (a.k.a., chatbots, AI 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, reCAPTHCA) 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 re-launches of research projects.[13,14]

Finally, there are *fraudulent participants* (a.k.a., imposter participants, scammers, 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 people with autism and/or parents of children with autism during online, qualitative interviews. In this particular example, several clues aroused researcher suspicion, such as keeping cameras off, inconsistent responses between prescreening and the interview, similarities in voices and mannerisms across interviews, and repeated enquiries 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 comprehensive review of the literature. This research gap makes it difficult and time consuming for researchers designing new online studies to decide best practices. The present review endeavors to close this gap through reviewing the most common types of threats to data integrity discussed in the literature, cataloging and evaluating the most common prevention methods, and discussing the ethical considerations raised about the techniques. Ultimately, this review aims to expand and centralize knowledge on 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, 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” and limited our results to articles published on or after 2020. This search yielding an initial 10,681 publications (Figure 1). 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: discussing an online study, using human participants, being written in English, and having at least one paragraph dedicated to discussing threats to data integrity. Ultimately, 90 articles met criteria for inclusion in this review.

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. One team member (IS) reviewed and coded each of the 90 articles; questions on coding were discussed and resolved with another team member (MR).

The following information was extracted and coded from the final 90 articles: article type (e.g. original research, commentary); type of data collection (e.g. qualitative, quantitative); methodology of the study (e.g. survey, qualitative interview); country where the study was conducted; recruitment methods (e.g. social media, survey service platform); type of suspected threat to data integrity (e.g. bots, fraudulent participants); the estimated prevalence of compromised data; proposed techniques to address threats to data integrity (e.g. authenticating personal information,

attention checks); and any ethical concerns raised by the researchers (e.g., risk to privacy).

Techniques to address threats to data integrity were additionally sorted in three 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 authors of the reviewed articles as being no longer sufficient in addressing threats to data integrity.

Results

The most common type of threat to data integrity documented by researchers was bots (n=50) followed by fraudulent participants (n=45), repeat participants (n=41), and satisficing participants (n=17). The most common recruitment method was social media advertising, followed by utilizing online survey service platforms, such as Mturk (see Supplemental Table 1). Several studies utilized more than one recruitment method.

Estimated prevalence of threats to data integrity ranged from ~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 and/or (2) identifying and removing poor-quality data if collected (see Supplemental Table 2). As illustrated in Table 1, a wide variety of techniques were represented. Techniques deemed to be “very effective,” included authenticating personal information, such as requesting to see participant’s 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 but that are not widely known by other groups. Additional “very effective” techniques concerned data checking: Cross-checking for inconsistent answers (e.g., between screening and a baseline survey) and checking for duplicate entries in the data (e.g., repeated email addresses).

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 (e.g., intentionally not emphasizing participant payments in recruitment materials), was framed as a preventative measure rather than definitive means of detecting fraud.

Finally, several techniques were deemed “no longer effective” by the authors: IP address and geolocation checks, reCAPTCHA, timing checks (i.e., checking for unusually fast or unusually slow responses), open-ended questions, and honeypot questions (i.e., 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 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 more simple 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 (Figure 2). One of the primary ethical concerns 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.[22] Many techniques, such as requiring a video call at screening, can also place additional burdens on participants and feel invasive.[4] 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

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. This review found that some of the most common preventative methods described in articles about threats to data integrity were IP Address/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 data integrity. Yet our review also discovered new and innovative techniques for addressing threats to data integrity, including authenticating personal information and posing background related questions.

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 researchers use multiple techniques (rather than relying on the soundness of just one technique), utilize the techniques described in the articles reviewed here, and draw the most from techniques deemed to be "very effective." When relying on survey service platforms for access to online samples, researchers should be critical of the techniques utilized 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 articles in this review highlight the responsibilities of researchers to continue focusing on participant rights and privacy. Techniques for ensuring data integrity (e.g., personal questions to authenticate identify) 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 also vary depending on the vulnerability of the sample and the sensitivity of the research topic. 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

Some limitations of this study are that we only used one database for review (PubMed). This was done due to the large volume of articles on the subject; however, it may have led to a disproportionate number of articles primarily centering around participant fraud in medical research studies due to the clinical focus of the PubMed database. In addition, 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 threats to data integrity, to quantitively compare the effectiveness of various prevention techniques.

References

1. Maheu C, Lemonde M, Mayo S, Galica J, Bally J. Moving research forward during COVID-19. Can Oncol Nurs J. 2021 Nov 1;31(4):490–492. PMCID: PMC8565433

2. Myers KJ, Jaffe T, Kanda DA, Pankratz VS, Tawfik B, Wu E, McClain ME, Mishra SI, Kano M, Madhivanan P, Adsul P. 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 [Internet]. Frontiers; 2022 Jun 8 [cited 2025 Jan 5];12. Available from: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2022.841951/full

3. Upadhyay UD, Jovel IJ, McCuaig KD, Cartwright AF. Using Google Ads to recruit and retain a cohort considering abortion in the United States. Contraception: X. 2020 Jan 1;2:100017.

4. Pellicano E, Adams D, Crane L, Hollingue C, Allen C, Almendinger K, Botha M, Haar T, Kapp SK, Wheeley E. Letter to the Editor: A possible threat to data integrity for online qualitative autism research. Autism. SAGE Publications Ltd; 2024 Mar 1;28(3):786–792.

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. The Qualitative Report. Fort Lauderdale, United States: The Qualitative Report; 2022 Nov;27(11):2469–2485.

6. Krosnick JA, Alwin DF. An evaluation of a cognitive theory of response-order effects in survey measurement. Public Opinion Quarterly. United Kingdom: Oxford University Press;

1987;51(2):201–219.

7. Suskie LA. Survey Research: What Works for the Institutional Researcher. Institutional Research Information Series No. 1 [Internet]. North East Association for Institutional Research, C/O Larry Metzger, Ithaca College, Ithaca, NY 14850 ($12; 1988. Available from: https://eric.ed.gov/?id=ED294497

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. 2012 Aug 1;24(3):272–291. PMCID: PMC3505140

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. PMCID: PMC4669957

10. Pullen Sansfaçon A, Gravel E, Gelly MA. Dealing With Scam in Online Qualitative Research: Strategies and Ethical Considerations. International Journal of Qualitative Methods. SAGE Publications Inc; 2024 Jan 1;23:16094069231224610.

11. Lawrence PR, Osborne MC, Sharma D, Spratling R, Calamaro CJ. Methodological Challenge: Addressing Bots in Online Research. Journal of Pediatric Health Care. Elsevier; 2023 May 1;37(3):328–332. PMID: 36717299

12. Al-Fannah NM. Making Defeating CAPTCHAs Harder for Bots [Internet]. arXiv; 2017 [cited 2025 Jan 6]. Available from: http://arxiv.org/abs/1704.02803

13. Perkel JM. Mischief-making bots attacked my scientific survey. Nature. 2020 Mar 17;579(7799):461–461.

14. Storozuk, A., Ashley, M., Delage, V., & Maloney, E. A. Got bots? Practical recommendations to protect online survey data from bot attacks. The Quantitative Methods for Psychology. 2020;16(5):472–481.

15. Chandler J, Sisso I, Shapiro D. Participant carelessness and fraud: Consequences for clinical research and potential solutions. Journal of Abnormal Psychology. US: American Psychological Association; 2020;129(1):49–55.

16. Devine EG, Pingitore AM, Margiotta KN, Hadaway NA, Reid K, Peebles K, Hyun JW. Frequency of concealment, fabrication and falsification of study data by deceptive subjects. Contemporary Clinical Trials Communications. 2021 Mar 1;21:100713.

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. 2015 Oct;142:256–260. PMID: 26322721

18. Siegel JT, Navarro M. A conceptual replication examining the risk of overtly listing eligibility criteria on Amazon’s Mechanical Turk. Journal of Applied Social Psychology. 2019;49(4):239–248.

19. Sharpe Wessling K, Huber J, Netzer O. MTurk Character Misrepresentation: Assessment and Solutions. Journal of Consumer Research. 2017 Jun 1;44(1):211–230.

20. Gordon JH, Fujinaga-Gordon K, Sherwin C. Fraudulent Online Survey Respondents May Disproportionately Threaten Validity of Research in Small Target Populations. Health Expectations. 2024;27(3):e14099.

21. Hardesty JJ, Crespi E, Nian Q, Sinamo JK, Breland AB, Eissenberg T, Welding K, Kennedy RD, Cohen JE. The Vaping and Patterns of e-Cigarette Use Research Study: Protocol for a Web-Based Cohort Study. JMIR Research Protocols. 2023 Mar 2;12(1):e38732.

22. Guest JL, Adam E, Lucas IL, Chandler CJ, Filipowicz R, Luisi N, Gravens L, Leung K, Chavanduka T, Bonar EE, Bauermeister JA, Stephenson R, Sullivan PS. Methods for Authenticating Participants in Fully Web-Based Mobile App Trials from the iReach Project: Cross-sectional Study. JMIR mHealth and uHealth. 2021 Aug 31;9(8):e28232.

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

Type of Threat	Other Terms	Definition
Satisficers	Cheaters, straightliners, speeders	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 renumeration
Bots	Chatbots, AI respondents	Computer algorithms deployed on studies in order to gain compensation without human effort for completion
Fraudulent Participants	Imposters, scammers, lying participants	Participants who lie about their identity or otherwise attempt to deceive researchers often with the intent of gaining study compensation

Table 2: Techniques used to address threats to data integrity in online research.

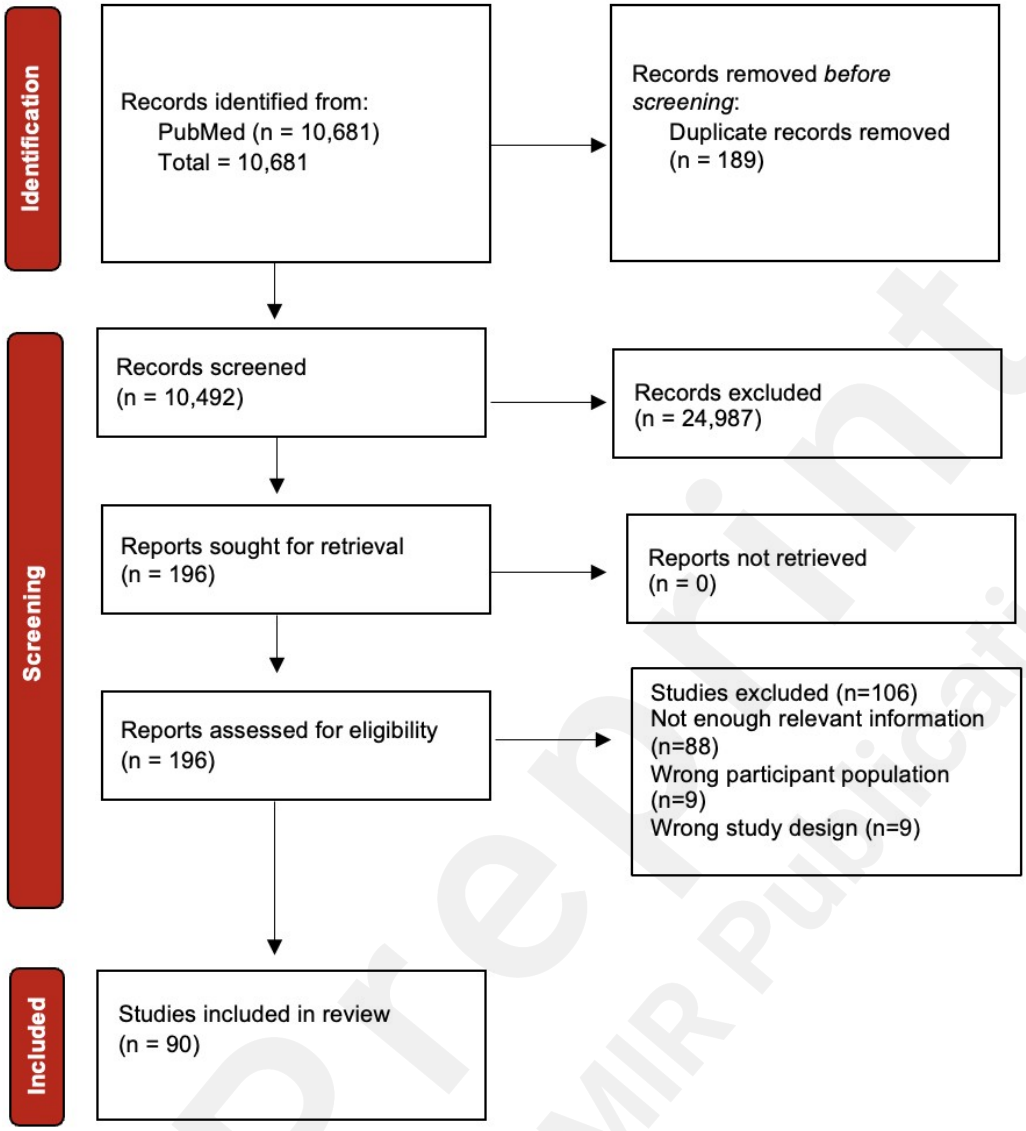
Rank	Technique	Description	Example	Freq.*
Very Effective	Authenticate Personal Information	Checking IDs, emails, addresses, zip codes, and phone numbers for authenticity. This includes using third party services to verify identities, requiring video calls at enrollment ("verification step"), or mailing 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: "*If you are keen to take part, we will arrange a preinterview chat, at a time that suits you. This will either be over the phone or online using Zoom, during which I will ask you some questions to make sure that you fit the study criteria. Your responses will be used to confirm your identity and to discourage scammer participants*" (Pellicano et al. 2024)	47.7 %
	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 four 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" (Tannahill and Blais 2024)	22.2 %
	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." (Guest et al 2021)	38.9 %
	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" (Guest et al 2021)	26.7 %

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: (a) embedded on the Grit scale—“Select ‘Somewhat like me’ for this statement,” (b) embedded on the Beck Depression Inventory—“1 – Select this option,” and (c) embedded on the Borderline Personality Inventory —“Select ‘Yes’ for this statement.’” (Webb and Tangney 2022)	27.8 %
	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” (Mizerek 2023).	7.8%
	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/constraint violations, and data inconsistencies” (Chen et al 2023)	23.3 %
	Watchful of a large number of responses	Checking the timestamps on survey submissions. A flurry of responses or sign-ups for a study 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 1-15–minute period” (Hohn et al 2022)	20.0 %
	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” (Bybee et al 2022)	6.6%
	Changing payment protocols	Intentionally not emphasizing and not automating participant payments. For example, not including payment amount in recruitment materials or using raffle payments.	“To maximize reach and limit fraud, gift cards could be manually distributed via text or email after each survey is verified” (Salem et al 2023)	20.0 %
	Not paying fraudulent respondents	Informing potential participant 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	11.1 %

			will not be approved or paid and your data will be discarded." (Gratz et al 2020).	
No Longer Effective	IP address/ 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" (Kumarasamy et al 2024).	57.8 %
	(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" (Bisdas et al 2021).	47.8 %
	Timing Checks	Checking for unusually fast or unusually 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, e.g., 4,220 min)" (Burnette et al 2022)	37.8 %
	Open-Ended Questions	Including open-ended questions and reviewing the responses. This can help assess attention as well as check for bots who may incoherently respond.	"Another indicator of data quality is suspicious responses to open-ended 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" (Douglas et al 2023)	23.3 %
	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" (Bonett et al 2024).	13.3 %

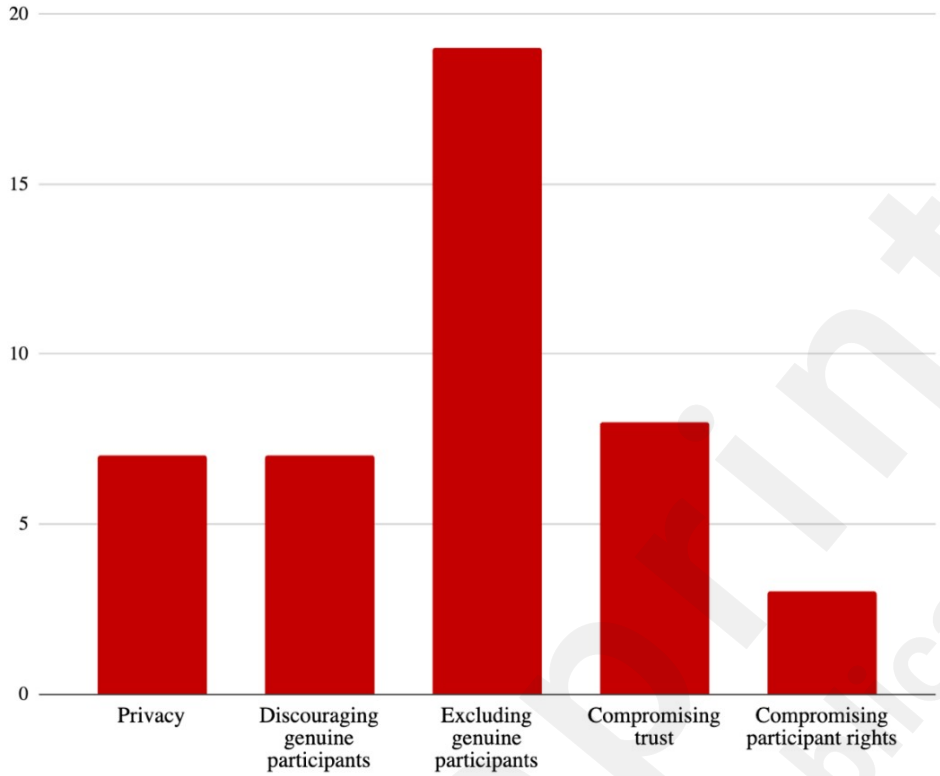
* Freq. = Frequency of articles mentioning this technique.

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



From: Page, MJ, McKenzie, JE, Bossuyt, PM, Boutron, I, Hoffman, TC, Mulrow, CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021; 372;n71.

Figure 2. Frequency of ethical considerations discussed in the reviewed articles.



Supplementary Files

Multimedia Appendixes

Supplemental table 1.
URL: http://asset.jmir.pub/assets/ac315e3b24fdd8f6f19c98bd603ccba0.docx

Supplemental table 2.
URL: http://asset.jmir.pub/assets/edb1b70e1f25bcb755efc0d60ed98ac1.docx