



LEEDS
BECKETT
UNIVERSITY

Citation:

Schneider, J and Ahuja, L and Dietch, JR and Folan, A-M and Coleman, J and Bogart, K (2024) Addressing fraudulent responses in quantitative and qualitative internet research: case studies from body image and appearance research. *Ethics & Behavior*. pp. 1-13. ISSN 1050-8422 DOI: <https://doi.org/10.1080/10508422.2024.2411400>

Link to Leeds Beckett Repository record:

<https://eprints.leedsbeckett.ac.uk/id/eprint/11423/>

Document Version:

Article (Accepted Version)

Creative Commons: Attribution 4.0

The aim of the Leeds Beckett Repository is to provide open access to our research, as required by funder policies and permitted by publishers and copyright law.

The Leeds Beckett repository holds a wide range of publications, each of which has been checked for copyright and the relevant embargo period has been applied by the Research Services team.

We operate on a standard take-down policy. If you are the author or publisher of an output and you would like it removed from the repository, please [contact us](#) and we will investigate on a case-by-case basis.

Each thesis in the repository has been cleared where necessary by the author for third party copyright. If you would like a thesis to be removed from the repository or believe there is an issue with copyright, please contact us on openaccess@leedsbeckett.ac.uk and we will investigate on a case-by-case basis.

Addressing Fraudulent Responses in Quantitative and Qualitative Internet Research: Case
Studies from Body Image and Appearance Research

Running Head: FRAUDULENT RESPONSES IN INTERNET RESEARCH

Jekaterina Schneider¹, Latika Ahuja¹, Jessica R. Dietch², Anne-Mairead Folan^{3,4}, Jillian
Coleman², Kathleen Bogart²

¹Centre for Appearance Research, School of Social Sciences, College of Health, Science and
Society, University of the West of England, Coldharbour Lane, Bristol, BS16 1QY, United
Kingdom

²School of Psychological Science, Oregon State University, 2950 SW Jefferson Way,
Corvallis, Oregon, United States

³Department of Psychology, School of Social Sciences, Frenchay Campus, University of the
West of England, Coldharbour Lane, Bristol, BS16 1QY, United Kingdom

⁴Portland Building, Department of Psychology, City Campus, Leeds Beckett University,
Leeds, LS1 3HE, United Kingdom

Author Note

Correspondence regarding this article should be directed to Jekaterina Schneider, Centre for
Appearance Research, School of Social Sciences, College of Health, Science and Society,
Frenchay Campus, University of the West of England, Coldharbour Lane, Bristol, BS16
1QY, United Kingdom. Email: kat.schneider@uwe.ac.uk.

ORCiDs

Jekaterina Schneider: <https://orcid.org/0000-0002-6069-4783>

Latika Ahuja: <https://orcid.org/0000-0001-5321-7906>

Jessica R. Dietch: <https://orcid.org/0000-0001-9434-0791>

Anne-Mairead Folan: <https://orcid.org/0000-0002-7567-0442>

Jillian Coleman: <https://orcid.org/0009-0005-9055-6778>

Kathleen Bogart: <https://orcid.org/0000-0002-6923-5700>

Word Count. Abstract: 100 words; Manuscript: 4,531 words.

Funding. Data collection for Case Study 1 was supported by the University of the West of England Vice Chancellor's Early Career Researcher (VC ECR) Development Award. Data collection for Case Study 3 was supported by the Moebius Syndrome Foundation. Data collection for Case Study 4 was supported by a research grant from the Dove Self-Esteem Project (Unilever; ref no. RHSS0180). Dr Dietch's work on this project was supported by a grant from the National Institutes of Health (K23HL157698).

Data Availability Statement. The data that support the findings of this study are available from the corresponding author, JS, upon reasonable request.

Authors' Contributions. **Jekaterina Schneider:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Supervision, Project Administration. **Latika Ahuja:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft. **Jessica R. Dietch:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft. **Anne-Mairead Folan:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft. **Jillian Coleman:** Conceptualization, Methodology, Investigation, Writing – Original Draft. **Kathleen Bogart:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Supervision.

Addressing Fraudulent Responses in Quantitative and Qualitative Internet Research: Case Studies from Body Image and Appearance Research

Abstract

The rise of online research methods has expanded the scope of research globally and has made research more inclusive. However, it has also led to a surge in fraudulent research participation, with individuals and bots infiltrating studies for personal gain or disruption. This is of particular concern in mental health studies, as fraudulent responses jeopardize interventions and care efforts. This paper addresses these challenges, presenting case studies from psychological research. The urgent need for a comprehensive understanding of fraudulent responses in both quantitative and qualitative online research is emphasized, urging the research community to confront and mitigate this issue effectively.

Keywords: Bots; Incentives; Fraudulent Research; Research Fraud; Strategies.

Introduction

Over the last few decades, online methods of data collection in research have become more prominent and widely used across multiple research fields, including psychology and mental health (Heiervang & Goodman, 2011; King et al., 2014). This surge in internet research can be attributed, in part, to remarkable technological advancements, which have facilitated more efficient data collection processes, thereby alleviating the burdens on both research participants and researchers (Caetano et al., 2018). Internet research has broadened the scale and scope of researchers to encompass a global audience (Woolfall, 2023), transcending geographical boundaries in data collection and enhancing data representation and quality, as well as reducing reliance on WEIRD (White, Educated, Industrialized, Rich, and Democratic) samples (Martinez et al., 2014; Sterzing et al., 2018; Van Selm & Jankowski, 2006). Additionally, its cost- and time-effective nature has promoted research environments to be more inclusive for participants with different social identities (e.g., sexuality, gender, ethnicity, socioeconomic background). Relatedly, social media is increasingly used as a source for promoting research studies and recruiting participants (Mizerek et al., 2023; Woolfall, 2023). Most recently, the COVID-19 pandemic has propelled an exponential growth in the utilization of online methods for recruiting participants and gathering both quantitative and qualitative research data (Renu, 2021; Woolfall, 2023; Yazici & Wang, 2023). These reasons underline the growing popularity of internet research.

While technological advancements have arguably made research activity across disciplines more efficient, evidence on the rapid growth in fraudulent activity in internet research is likewise increasing (Goodrich et al., 2023; Jackson et al., 2023). “Fake participants” or “fraudulent respondents” refer to individuals who participate in online research studies without genuine interest in contributing to the research process. Instead, they may sign up solely to receive incentives, compensation, or rewards offered to research

participants; for their own personal agenda; or to disrupt the research process itself. To participate in the research, fraudulent participants often deceitfully assert eligibility for a study (e.g., claiming specific conditions, claiming that one resides in a designated region, identifying with specific ethnicities; Hoerger, 2024) and/or fabricate responses within surveys. They may also fill out surveys themselves or use bots—automated software programs designed to simulate human actions and perform tasks on the internet—to submit one or multiple responses to a survey. It is often difficult to distinguish or identify whether a survey response was submitted by a human with authentic intentions, a fraudulent human responder, or a bot, which makes the nature of this problem complex.

Between 20–100% of survey responses have been found as fraudulent in previous research (Ballard et al., 2019; Pozzar et al., 2020). Fraudulent participation in online research introduces a litany of concerns and risks to the research process, including the compromise of data quality (e.g., by elevating the risk of Type I and Type II errors; Storozuk et al., 2020), the misallocation of resources (e.g., incentives, researcher time), and, most distressingly, the potential distortion of findings in studies addressing individual mental health and well-being. These concerns can contribute to health inequity by jeopardizing intervention and care endeavors, leading to the development of ineffective policies that fail to address the unique challenges faced by marginalized populations, and possibly cause further harm to the studied population. This can exacerbate health disparities by diverting attention and resources away from those who need them most.

Previous reviews (e.g., Drysdale et al., 2023; Godinho et al., 2020; Jones et al., 2021; Lawlor et al., 2021; Lawrence et al., 2023; O'Donnell et al., 2023; Ridge et al., 2023; Storozuk et al., 2020; Teitcher et al., 2015) provide numerous recommendations for preventing, identifying, and tackling fraudulent responses in online studies. The most common techniques include utilizing tools in survey software such as the reCAPTCHA

(Completely Automated Public Turing test to tell Computers and Humans Apart) and collection of metadata such as IP addresses, employing attention check or open-ended questions, and including screening questions at the beginning of surveys. However, due to technological advancements, bots are learning quickly and improving their ability to pass bot screeners, including the reCAPTCHA (Von Ahn et al., 2008), which questions the effectiveness of these detection strategies. Our assessment is that there has been a recent increase in the sophistication of fraudulent activities. While researchers may be using older techniques for data quality, including attention checks and the reCAPTCHA, this is no longer sufficient. Thus, we suspect researchers might have a false sense of security and not notice fraudulent activity in their findings. If they do, they may feel stigma or shame for not having anticipated and prevented fraud. As a result, researchers likely under-report this problem, which not only reduces the confidence in research findings but also underlines the adverse impact of fraudulent research activities on an individual's motivations for scientific inquiry.

Fraudulent activities have been detected across studies using both quantitative and qualitative methodologies (Drysdale et al., 2023; Godinho et al., 2020; Jones et al., 2021). However, despite emerging attention to the challenges posed by bots and fake participants in online surveys, limited consideration has been given to the phenomenon of fraudulent participants infiltrating online interviews and focus groups (Jones et al., 2021; O'Donnell et al., 2023; Ridge et al., 2023). The increasing prominence of internet-based qualitative and mixed-methods studies highlights the urgency of raising this concern within the research community to better understand the nature of fraudulent responders and activity (O'Donnell et al., 2023; Renu, 2021; Sefcik et al., 2023; Woolfall, 2023; Yazici & Wang, 2023). Therefore, the primary objective of this commentary paper is to provide an overview of the challenges frequently encountered when grappling with bots and fake participants in quantitative and qualitative online studies, respectively. To this end, we present four case

studies drawn from body image and appearance research, highlighting the obstacles we have confronted in dealing with fraudulent responses, and proffering potential strategies we intend to adopt moving forward. Furthermore, we delineate additional considerations imperative for effectively anticipating, mitigating, and confronting fraudulent responses in quantitative and qualitative internet research.

Methods

This research describes four case studies of fraudulent responses in body image and appearance research. The first two case studies highlight fraudulent activity in qualitative internet research, which utilized Zoom and/or Microsoft Teams to conduct one-to-one interviews with participants. The last two case studies highlight the same issue, but in the context of quantitative internet research. All studies were conducted online, provided incentives for study participation, and included targeted participant samples.

Results

Case Study #1 – A Qualitative Interview Study Exploring Positive Body Image among Men

Case Description

This project aimed to explore positive body image among men living in the United Kingdom (UK), with an intention to adapt existing body image interventions for men. The study recruitment was done via men's charities (e.g., Beyond Equality and the Blue Ribbon Foundation). Given that recruiting men is often challenging, particularly in psychology and body image research, a pragmatic decision was made to recruit via social media platforms, including Instagram, Twitter, and Facebook. Alternative recruitment strategies included identifying participants from the University Participant Pool and university student and staff networks. Study advertisement was done by circulating e-posters across different recruitment channels. Men were eligible to participate if they lived in the UK, were aged between 18–60

years, and experienced positive body image (assessed via a screening questionnaire). Eligible participants were invited to take part in a 60-minute interview via Zoom or Microsoft Teams. Recruitment was conducted between February and April 2021, and participants were offered £15 for their participation.

Commentary

The recruitment for the study started in February. Two weeks after the study went live, the screening survey began to receive a lot of traction and we received 40 responses in one day. This was perceived as odd, given the evidence that men are reluctant to participate in mental health research broadly and in appearance research specifically. The majority of these 40 responders scored 35 and above (on a scale of 40) on the Body Appreciation Scale, which was the screener used to identify eligible participants. Most of the responders had reported their location to be around London and identified themselves as Black British. At first, this was not perceived as suspicious. However, upon approaching the participants to set a time for the interview to take place, the research team received email replies that were either too short or not coherent (e.g., “Ok, can do”; “Works”). These emails often had nonsensical or generic subject lines (e.g., “participate in research”) and enquired about the incentives before the interviews were conducted. The first fraudulent responder was suspected during one of the interviews as he did not switch on the camera, gave short replies to the questions asked, and disconnected the Microsoft Teams call every five minutes. During the interview, it seemed he was sitting next to someone who was helping him with the responses. When asked about his occupation and location, his answer differed from the one provided on the demographic survey. This helped us identify the first fraudulent responder, after which all subsequent interviewees were screened before the actual interview started.

Case Study #2 – A Qualitative Interview Study Exploring Individuals’ Experiences of Cognitive Behavioral Therapy for Bulimia Nervosa

Case Description

This project aimed to explore individuals’ experiences of receiving cognitive behavioral therapy (CBT) to treat the eating disorder bulimia nervosa (BN). Twitter, Facebook, and Reddit were used to promote the study. When posting in Facebook groups or Reddit threads, these were either eating disorder/BN-specific, general mental health spaces, men’s groups, or spaces for marginalized groups, such as LGBTQ+ online spaces. The study was also advertised via the Call for Participants website and various eating disorder organizations, such as Beat and First Steps ED. Participants were eligible for inclusion if they were aged 18 years or above, had received CBT to treat BN, and had undergone this therapy in the UK. Eligible participants were invited to take part in a 30–60-minute interview via Zoom or Microsoft Teams. Recruitment commenced in March 2022 and is still ongoing at the time of writing. Participants were offered a £20 voucher for their participation.

Commentary

Predominantly after posting the study on Twitter (but also following the advertisement of the research on Call for Participants and similar websites), the study’s Qualtrics webpage, where participants were directed to in the Tweet/study poster, received dozens of sign-ups. The lead researcher would also receive numerous emails, and these would usually be written within seconds/minutes of one another. The language used within the emails was very similar, lacking any personalization, often simply ensuring the researcher that they were eligible for the study and wished to take part, and frequently contained grammatical errors. When the researcher asked for further information, this would usually result in the individual not responding, or providing a response that did not make sense (e.g., when prompted for more information about the nature of the CBT they received, numerous

people replied providing details of a therapy that was not CBT). Most people were screened out following this; however, a brief screening call was held with one individual due to the lack of clarity over whether they were a fraudulent responder. It became apparent very quickly into the screening call that they were not eligible, as they could not answer questions about where they received their treatment (instead repeating that they had one-to-one treatment) and appeared to be reading information online about BN while responding, after it became apparent during their first couple of sentences that they were unaware of what BN is. The individual used phrases such as “I displayed compensatory behaviors such as excessive exercise and purging following a binge”—language that was likely taken from the internet and that was inconsistent with phrases they had previously used.

Case Study #3 – A Quantitative Survey Study of Sleep Health in Moebius Syndrome

Case Description

This project explored sleep characteristics of adults and parents of children with Moebius syndrome, a rare congenital disorder. All individuals in any country who self-identified as having Moebius syndrome or being the parent of a child with Moebius syndrome were eligible. Recruitment was conducted by asking the Moebius Syndrome Foundation to share the public survey link within their private contact lists. Although the researchers did not share the survey link publicly, the Moebius Syndrome Foundation did share publicly on social media, including Facebook and Twitter. The researchers also shared a public survey link in closed Facebook groups and emailed participant lists from prior research with individuals known to have Moebius syndrome. The study was also advertised in-person at the Moebius Syndrome Foundation conference. Participants were offered a \$15 gift card as compensation.

Commentary

We initially opened the survey and shared it in early November 2022. Four days after launch, the survey began to receive hundreds of responses, which is inconsistent with the size of the population of interest. We temporarily closed the survey four days later, with over 3,000 responses collected, to reassess and strategize to prevent further fraudulent responses. Among the suspected fraudulent responses received, free text responses were nonsensical (e.g., “If you don’t sleep well, you can drink more milk before going to bed”) or clearly drawn from language on the internet (e.g., a list of all symptoms common to Moebius syndrome in response to a question about other comorbid disorders), and in many cases this text was identical across dozens of responses. Often, multiple responses came in within seconds of one another, and many responses were completed too quickly to be feasible given the length of the survey. Additionally, we noted nonsensical responses to closed questions (e.g., questions about how long it took to fall asleep or how long one was awake in the middle of the night exceeding 10 hours). We also noted inconsistencies in responses across similar questionnaires (e.g., two questionnaires about nightmares).

Case Study #4 – A Quantitative Randomized Controlled Trial of a Body Image

Intervention for Sports Coaches

Case Description

The aim of this pilot randomized controlled trial was to test a novel online body image intervention for sports coaches. The study consisted of completing an online questionnaire at baseline, taking part in five online modules over two weeks, and then completing a second online questionnaire at post-intervention. Initially, coaches were recruited through emails to sports organizations and known contacts of the core research team. However, recruitment proved challenging and, eventually, the study link including the baseline survey was posted across various social media platforms (i.e., Instagram, Facebook,

Twitter) to speed up recruitment. Participants were eligible for inclusion if they were current sports coaches of adolescent girls and residing in the United States (US). Participants were excluded if they coached only boys and/or men or had taken part in body image research in the past. Participant recruitment and data collection were conducted online between May and August 2022. Upon completion of the study (i.e., pre-intervention survey, intervention, and post-intervention survey), participants would receive an Amazon gift card worth \$25.

Commentary

Shortly after the survey was shared via social media, responses started coming in very rapidly, at almost one response per minute. This was accompanied by multiple emails in quick succession inquiring about incentives and when these will be paid out. Responses raised suspicion initially because names were spelled with lowercase letters, some names sounded unlikely or made up (e.g., joe biden), and the names did not match the email addresses provided. Upon examination of IP addresses, it later became clear that many responses were coming from the same IP address, and the location was outside of the US. This prompted us to close the survey and limit recruitment to known sources. A second survey was created to recruit via social media, but without disclosing the incentive associated with the study.

Commonalities Across Case Studies

The four case studies described above shared several commonalities with regards to the identification of fraudulent responses. First, communication patterns via email were a key source of suspicion. Across all case studies, researchers observed a sudden surge in email correspondence within a short timeframe. Emails from participants displayed a consistent style, often comprising just one or two sentences and exhibiting incoherence and repetitive phrasing (e.g., “I interested in research”; “Hello researcher I am eligible”; “I would love to share my experiences”). These emails lacked personalization, frequently contained

grammatical errors, and either lacked an email subject line or used a generic subject line like “research study”. Notably, the majority of these emails originated from Gmail or Yahoo accounts, many of which included random combinations of letters and numbers or followed a repetitive structure (e.g., “LastnameFirstname#####@gmail.com”). Participants’ responses to study-related emails were also noteworthy. Participants occasionally made immediate inquiries about incentives after interviews or pursued the research team the following day, emphasizing the urgency of reimbursement due to personal circumstances or family health issues. These incentive-related emails frequently contained repetitive questions and language across multiple participants.

Second, concerns emerged regarding the characteristics of survey responses. Researchers noticed a significant influx of survey submissions in a short time period, often accompanied by multiple responses in quick succession. Participants frequently wrote their names entirely in lowercase, which did not always align with the names embedded in their email addresses. Multiple responses originating from the same IP address, especially those received in close proximity, raised suspicion. Researchers also identified failed “attention check” item responses designed to capture careless responding (e.g., “for this item, please select ‘Strongly Disagree’”), although some identified fraudulent responders were able to successfully pass all of these items. Some responses were completed unrealistically quickly given the survey’s length, and many free-text responses appeared nonsensical or were copied verbatim from internet sources. Some participants used idiosyncratic phrases, such as “don’t hurt” in multiple responses. Improbable or impossible answers were observed in questions related to sleep patterns, as well as notable inconsistencies in responses to questions seeking similar information (e.g., age and date of birth). Furthermore, some participants exhibited unusually high scores on the screening measures employed for the studies.

Third, during the interview phase in studies employing qualitative methodology, various issues further raised concerns about participant authenticity. Some participants refused to activate their cameras, citing technical issues, while others spoke in a low voice and provided random or superficial answers, resulting in low-quality interviews. On several occasions, multiple users appeared on Microsoft Teams or Zoom calls designated for individual interviews. Participants frequently disconnected from these video calls, disrupting the interview process. Prolonged pauses before answering questions gave the impression that participants may have been conferring with someone before responding aloud. In one instance, excessive background noise interrupted the interview, and the call was repeatedly muted whenever a question was posed. Concerningly, inconsistencies were also present between responses provided in the pre-interview screening survey and responses given during the interview session.

Discussion

Given the exponentially growing prevalence of fraudulent responses in both quantitative and qualitative research studies, as well as advancements in artificial intelligence (AI) and bot technology, we are calling for a focused and global research agenda to combat this issue. As this is a continuously evolving issue, there needs to be a consistent and ongoing attempt at determining effective strategies for fraud detection. The strategies recommended in this and other reviews should therefore not be taken as a comprehensive and final strategy for fraud prevention; rather, a flexible agenda is required that can identify and respond to contemporary developments in this field. Documenting genuine participant experiences in a diverse field of research is crucial not only to advance our knowledge, but also to benefit the population we are studying without causing undue harm. Moreover, ensuring that only genuine participants are represented by our data allows us to draw accurate conclusions about the study findings, and prevents a waste of researcher, departmental, and university resources.

This paper builds and extends on the existing literature in this domain by presenting case studies and learnings from body image and appearance research. In our experience, Qualtrics spam/bot detection tools and the reCAPTCHA were not effective in flagging or preventing fraudulent respondents from accessing surveys. Analyzing responses in the aggregate enabled us to eliminate a substantial number of fraudulent entries. For instance, while an individual response might appear benign and trustworthy, the presence of numerous responses featuring identical free-text language made it far easier to identify and exclude fraudulent entries. Similarly, a large volume of similar responses submitted within seconds of one another (particularly at late hours in the respondents' apparent time zone) may help rule out a bank of respondents. Moreover, despite using standard methods like multiple-choice attention checks, limitations persist, with genuine participants failing and fraudsters passing. To enhance accuracy, we suggest employing intricate attention checks like fill-in-the-blank questions or other free-text responses with specific instructions. Making these free-text fields required can help prevent ambiguous blank answers from passing through. Although this approach demands individual assessment and may still screen out some authentic participants, it can substantially improve result confidence by effectively deterring fraudulent responses.

Importantly, no single strategy when employed in isolation can be entirely foolproof (Nur et al., 2023). Therefore, we advocate for the implementation of a variety of strategies across all phases of research, accompanied by vigilant monitoring of participant recruitment and data collection processes to ensure adaptability and responsiveness to emerging issues (see Table 1 for a summary of recommended strategies to consider pre-, during, and post-data collection). These strategies should consider factors such as their effectiveness, the burden they impose on researchers and participants, potential challenges raised for recruitment, implications for research design, and ethical considerations, among others. Ideally, a plan for

preventing fraudulent responses should be developed *before* data collection. This may include an exclusion criteria plan for suspected fraudulent data. For example, if a respondent has two or more of the “red flags” listed in Table 1, their data will be excluded from the study. It is essential to evaluate the suitability of these strategies on a case-by-case basis, and it is certain that strategies will need to be updated over time. To comprehensively address this issue, collaborative efforts within the scientific and research community are paramount. Moreover, we provide some additional recommendations and considerations below.

[TABLE 1 NEAR HERE]

Working with Vulnerable, Marginalized, or Underrepresented Populations

Researchers have increasingly turned to online research to understand the experiences of vulnerable, marginalized, or underrepresented populations, often due to their rarity (e.g., rare disorders or disabilities), geographical dispersion, or challenges related to accessibility or safety in in-person research settings. However, certain ethical and practical considerations must be taken into account when working with these populations. For instance, some individuals, such as transgender participants, may face unique challenges when it comes to providing proof of identity. Asking for identification confirmation may be inappropriate or unsafe in such cases. Similarly, in the context of research on topics like COVID-19, requesting participants to share their health status or medical documents may raise ethical concerns. Thus, researchers must navigate these complex issues while ensuring the ethical conduct of their studies (see below).

Working with Ethics Committees and Institutional Review Boards

We recommend that researchers engage with their institution’s ethical departments to identify training and support needs and gain approval for any additional wording to be included in information sheets and consent forms. This could include, for example, a notice in the information sheet or consent form to inform participants that if fraudulent activity is

suspected, their responses will be considered void and no incentive will be provided (note, however, that this could also put off authentic participants from taking part). Establishing systems and protocols to navigate such situations is crucial to protect researchers from potential ethical and legal dilemmas. Discussions with ethics boards should also explore the feasibility of collecting additional identifying information to aid in detecting fraudulent responses, with provisions for ensuring that the research process remains compliant to data protection policies when collecting and storing data and de-identifying responses after validation to maintain confidentiality.

Working with Other Academic Institutions

In addition to updating ethical guidelines, we propose updating reporting guidelines for internet studies to encompass items that confirm attempts to prevent fraudulent responses. Current guidelines (Eysenbach, 2004; Sharma et al., 2021) primarily focus on preventing repeated responses, which proves insufficient given the evolving landscape of fraudulent response patterns. With regards to the publication and dissemination of research findings, it is imperative that academic journals and journal reviewers incorporate a rigorous assessment of issues pertaining to fraudulent responses in internet research as an integral part of the manuscript review process. Journals should consider including guidelines and criteria for authors to transparently report on their strategies for detecting and mitigating fraudulent responses in their online studies, including the number of responses removed from datasets and the rationale for their removal. Reviewers, in turn, should be well-versed in evaluating the adequacy of these strategies, ensuring that the research findings are not compromised by data contamination.

Working with IT Services and Other Administrative Bodies

When built-in bot-prevention features of software are utilized, the research community should work with their IT services to provide data about the efficacy of the

features. IT departments often have a point of contact for research IT vendors and could assist with providing data about efficacy directly to the vendor. Researchers should document specific cases of known fraud to provide to their IT departments for feedback about the efficacy. Researchers could also partner with IT services to determine whether additional paid bot prevention features are available for software and could be added to the institution's contract. As participant fraud impacts the integrity of data, researchers should engage with stakeholders in IT services and research administration offices to discuss opportunities for developing university guidance, or engaging with software companies to improve the services they provide. Technological advancements also hold promise for countering research fraud while addressing previous challenges. For instance, software is being developed to integrate video questions and responses within online surveys (e.g., VideoAsk by Typeform Official; <https://www.videoask.com/>). These features, similar to identity verification tools in online banking and other domains, enable researchers to verify that actual human respondents are completing their surveys, reducing the likelihood of bot involvement. Finally, in research projects utilizing research agencies or survey software platforms (e.g., Qualtrics, Prolific, Amazon mTurk), strategies to prevent and detect potentially fraudulent responses should be initiated at the project's outset.

Working with Law Enforcement

Considering that fraudulent responses may constitute internet fraud, especially when incentives are involved, researchers should contemplate the value in notifying law enforcement agencies (e.g., the U.S. Federal Bureau of Investigation's Internet Crime Complaint Center, <https://www.ic3.gov>; the UK National Crime Agency, <https://www.nationalcrimeagency.gov.uk/who-we-are>). However, it should be noted that law enforcement bodies dedicated to this issue may not be available in all countries.

Conclusions

This commentary paper underscores the persistent challenge of fraudulent responses in internet-based quantitative and qualitative research and advocates for multifaceted strategies to enhance research integrity. Collaboration with research agencies, training, and updated reporting guidelines are recommended. These efforts, integrated into research ethics, can help researchers navigate fraudulent responses ethically. The recommendations discussed in this commentary are not fixed solutions; thus, we emphasize the need for ongoing vigilance and adaptability in the face of evolving technology. By collectively addressing this issue and working collaboratively with the broader research community, we can fortify the robustness of internet research and uphold the ethical principles that underpin our scientific pursuits.

Acknowledgements. We would like to thank Dr Julie Woodley for her constructive comments on our manuscript from the perspective of research ethics and Dr Melissa Stepney for sharing her insights on this topic in the early stages of the project.

Declaration of Interest. The authors report there are no competing interests to declare.

References

- Ballard, A. M., Cardwell, T., & Young, A. M. (2019). Fraud detection protocol for web-based research among men who have sex with men: Development and descriptive evaluation. *JMIR Public Health and Surveillance*, 5(1), e12344. <https://doi.org/10.2196/12344>
- Caetano, R., & Noel, J. (2018). Web-based research: Strengths, weaknesses, and JSAD's guidance for authors. *Journal of Studies on Alcohol and Drugs*, 79(6), 813–815. <https://doi.org/10.15288/jsad.2018.79.813>
- Drysdale, K., Wells, N., Smith, A. K., Gunatillaka, N., Sturgiss, E. A., & Wark, T. (2023). Beyond the challenge to research integrity: Imposter participation in incentivised qualitative research and its impact on community engagement. *Health Sociology Review*, 32(3), 372–380. <https://doi.org/10.1080/14461242.2023.2261433>
- Eysenbach, G. (2004). Improving the quality of Web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research*, 6(3), e34. <https://doi.org/10.2196/jmir.6.3.e34>
- Godinho, A., Schell, C., & Cunningham, J. A. (2020). Out damn bot, out: Recruiting real people into substance use studies on the internet. *Substance Abuse*, 41(1), 3–5. <https://doi.org/10.1080/08897077.2019.1691131>
- Goodrich, B., Fenton, M., Penn, J., Bovay, J., & Mountain, T. (2023). Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys. *Applied Economic Perspectives and Policy*, 45(2), 762–784. <https://doi.org/10.1002/aepp.13353>
- Heiervang, E., & Goodman, R. (2011). Advantages and limitations of web-based surveys: Evidence from a child mental health survey. *Social Psychiatry and Psychiatric Epidemiology*, 46(1), 69–76. <https://doi.org/10.1007/s00127-009-0171-9>

- Hoerger, M. (2024). Faking health vulnerabilities to meet eligibility criteria to participate in paid internet-mediated research during the COVID-19 pandemic: Three case reports. *Ethics & Behavior*, 1–6. <https://doi.org/10.1080/10508422.2024.2325411>
- Jackson, A. M., Woo, J., Olson, M., Dalisay, F., Pokhrel, P., Muller, C. J., & Okamoto, S. K. (2023). Methodological challenges in web-based qualitative research with medically underserved populations. *Journal of Medical Internet Research*, 25, e44086. <https://doi.org/10.2196/44086>
- Jones, A., Caes, L., Rugg, T., Noel, M., Bateman, S., & Jordan, A. (2021). Challenging issues of integrity and identity of participants in non-synchronous online qualitative methods. *Methods in Psychology*, 5, 100072. <https://doi.org/10.1016/j.metip.2021.100072>
- King, D. B., O'Rourke, N., & DeLongis, A. (2014). Social media recruitment and online data collection: A beginner's guide and best practices for accessing low-prevalence and hard-to-reach populations. *Canadian Psychology*, 55(4), 240–249. <https://doi.org/10.1037/a0038087>
- Lawrence, P. R., Osborne, M. C., Sharma, D., Spratling, R., & Calamaro, C. J. (2023). Methodological challenge: Addressing bots in online research. *Journal of Pediatric Health Care*, 37(3), 328–332. <https://doi.org/10.1016/j.pedhc.2022.12.006>
- Lawlor, J., Thomas, C., Guhin, A. T., Kenyon, K., Lerner, M. D., Ucas Consortium, & Drahota, A. (2021). Suspicious and fraudulent online survey participation: Introducing the REAL framework. *Methodological Innovations*, 14(3), 20597991211050467. <https://doi.org/10.1177/20597991211050467>
- Martinez, O., Wu, E., Shultz, A. Z., Capote, J., Rios, J. L., Sandfort, T., ... & Rhodes, S. D. (2014). Still a hard-to-reach population? Using social media to recruit Latino gay

- couples for an HIV intervention adaptation study. *Journal of Medical Internet Research*, 16(4), e3311. <https://doi.org/10.2196/jmir.3311>
- Mizerek, E., Wolf, L., & Moon, M. D. (2023). Identifying and mitigating fraud when using social media for research recruitment. *Journal of Emergency Nursing*, 49(4), 530–533. <https://doi.org/10.1016/j.jen.2023.04.002>
- Mournet, A. M., & Kleiman, E. M. (2023). Internet-based mental health survey research: Navigating internet bots on Reddit. *Cyberpsychology, Behavior, and Social Networking*, 26(2), 73–79. <https://doi.org/10.1089/cyber.2022.0173>
- Nur, A. A., Leibbrand, C., Curran, S. R., Votruba-Drzal, E., & Gibson-Davis, C. (2023). Managing and minimizing online survey questionnaire fraud: Lessons from the Triple C project. *International Journal of Social Research Methodology*, 1–7. <https://doi.org/10.1080/13645579.2023.2229651>
- O'Donnell, N., Satherley, R. M., Davey, E., & Bryan, G. (2023). Fraudulent participants in qualitative child health research: Identifying and reducing bot activity. *Archives of Disease in Childhood*, 108(5), 415–416. <https://doi.org/10.1136/archdischild-2022-325049>
- Pozzar, R., Hammer, M. J., Underhill-Blazey, M., Wright, A. A., Tulsy, J. A., Hong, F., Gundersen, D. A., & Berry, D. L. (2020). Threats of bots and other bad actors to data quality following research participant recruitment through social media: Cross-sectional questionnaire. *Journal of Medical Internet Research*, 22(10), e23021. <https://doi.org/10.2196/23021>
- Ridge, D., Bullock, L., Causer, H., Fisher, T., Hider, S., Kingstone, T., ... & Southam, J. (2023). 'Imposter participants' in online qualitative research, a new and increasing threat to data integrity?. *Health Expectations: An International Journal of Public*

Participation in Health Care and Health Policy, 26(3), 941–944.

<https://doi.org/10.1111/hex.13724>

Renu, N. (2021). Technological advancement in the era of COVID-19. *SAGE Open Medicine*, 9, 205031212110009. <https://doi.org/10.1177/20503121211000912>

Sefcik, J. S., Hathaway, Z., & DiMaria-Ghalili, R. A. (2023). When snowball sampling leads to an avalanche of fraudulent participants in qualitative research. *International Journal of Older People Nursing*, e12572. <https://doi.org/10.1111/opn.12572>

Sharma, A., Minh Duc, N. T., Luu Lam Thang, T., Nam, N. H., Ng, S. J., Abbas, K. S., ... & Karamouzian, M. (2021). A consensus-based checklist for reporting of survey studies (CROSS). *Journal of General Internal Medicine*, 36(10), 3179–3187.

<https://doi.org/10.1007/s11606-021-06737-1>

Sterzing, P. R., Gartner, R. E., & McGeough, B. L. (2018). Conducting anonymous, incentivized, online surveys with sexual and gender minority adolescents: Lessons learned from a national polyvictimization study. *Journal of Interpersonal Violence*, 33(5), 740–761. <https://doi.org/10.1177/0886260517744845>

Storozuk, A., Ashley, M., Delage, V., & Maloney, E. A. (2020). Got bots? Practical recommendations to protect online survey data from bot attacks. *The Quantitative Methods for Psychology*, 16(5), 472–481. <https://doi.org/10.20982/tqmp.16.5.p472>

Teitcher, J. E., Bockting, W. O., Bauermeister, J. A., Hofer, C. J., Miner, M. H., & Klitzman, R. L. (2015). Detecting, preventing, and responding to “fraudsters” in internet research: Ethics and tradeoffs. *Journal of Law, Medicine & Ethics*, 43(1), 116–133. <https://doi.org/10.1111/jlme.12200>

Van Selm, M., & Jankowski, N. W. (2006). Conducting online surveys. *Quality and Quantity*, 40(3), 435–456. <https://doi.org/10.1007/s11135-005-8081-8>

Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). reCAPTCHA: Human-based character recognition via web security measures. *Science*, *321*(5895),

1465–1468. <https://doi.org/10.1126/science.1160379>

Woolfall, K. (2023). Identifying and preventing fraudulent participation in qualitative research. *Archives of Disease in Childhood*, *108*(6), 421–422.

<https://doi.org/10.1136/archdischild-2023-325328>

Yazici, E., & Wang, Y. (2023). Attack the bot: Mode effects and the challenges of conducting a mixed-mode household survey during the Covid-19 pandemic.

International Journal of Social Research Methodology, 1–6.

<https://doi.org/10.1080/13645579.2023.2241797>

Tables

Table 1

Proposed Solutions for Deterring, Identifying, and Addressing Fraudulent Responses in Internet Research

Pre-Data Collection	During Data Collection	Post-Data Collection
<p>Incentive Restriction. Avoiding the inclusion of incentive-related information in advertisements can be an effective method to reduce fraudulent activity (Mournet & Kleiman, 2023). Similarly, offering incentives via post instead of electronically can be a deterrent to fraudulent research participation. However, this approach would have implications for anonymity and regulations around data collection and management, as additional identifiable data would have to be collected. This may deter people from responding or they may alter their answers if they know their responses may be attributed to them. If incentives will be restricted to only those participants who meet certain criteria (e.g., pass a certain number of attention check items, provide verification of identification), this should be clearly described in the informed consent form.</p>	<p>Free Text Responses. Incorporating free-text response fields with specific instructions can help filter out fraudulent responders and disengaged participants and assess participant comprehension and response quality. However, relying solely on researchers' judgment of response validity should be avoided, as language models can produce sophisticated responses. Therefore, researchers should be cautious of potential false positives flagged by such tools.</p>	<p>Response Time Monitoring. Monitoring the time taken to complete surveys can signal potential fraudulent responses. Both excessively short and long completion times may raise suspicion, prompting further investigation.</p>
<p>Recruitment through Trusted Channels. Advertising research and recruiting participants through known and trusted channels, such as healthcare sites and clinics, regulated university participant pools, organizational list servers, or closed social media groups can reduce the risk of bots or fraudulent participants infiltrating the research.</p>	<p>Camera Usage. Encouraging participants to keep their cameras on during interviews, or at least to turn them on during an initial check, can deter fraudulent responders. While not foolproof, this approach can prompt genuine participants to maintain video communication. However, this may introduce additional ethical implications if participants prefer to remain anonymous. An alternative approach may be to ask participants to present a form of ID (with or without an image) to verify that the name matches with that provided on the consent form.</p>	<p>IP Address and Location Checks. Regularly examining IP addresses and participant locations can help identify suspicious participation patterns. Participants could be asked to provide their zip code, city, county, or other location information, which could be checked against IP address and location data to assist with flagging potentially fraudulent activity. Additionally, websites are available that can check whether an IP address has been reported to be involved with fraudulent activity or is using a VPN (e.g., https://www.ipqualityscore.com/bot-management/bot-detection-check). Again, this approach has</p>

Screening Forms and Calls.

Implementing screening forms or calls before providing personalized survey links or before conducting interviews can be beneficial in verifying participant eligibility and authenticity. This step can discourage individuals attempting to gain incentives dishonestly, although it may also result in some participants not attending their screening calls and can increase participant burden and recruitment challenges.

Verified Survey Panels. Utilizing verified survey panels offers a secure method for reducing fraudulent responses, but it may not be suitable for research on rare conditions or specific populations (e.g., adolescents) due to recruitment limitations and potential cost constraints.

Survey Link Restriction. Closing compromised survey links and creating new ones can prevent fraudulent responses. Interested participants can then be directed to a screening survey or to contact the lead researcher for a legitimate survey link. Researchers could also create links specific to recruitment locations, so that if an influx of fraud is identified for one of the recruitment locations, that location's link could be closed, limiting disruption to recruitment. Making a link more challenging to access can dissuade fraudulent respondents, but it can also have the side effect of limiting genuine respondents and thus interfere with recruitment.

Daily Data Integrity Checks. Conducting meticulous daily checks on incoming data can expedite the identification and closure of compromised surveys, reducing the amount of corrupted data to manage. For example, checking email timestamps for accuracy can help confirm whether participants are located in the claimed region, especially when eligibility criteria involve specific geographic locations. Discrepancies in timestamps may indicate fraudulent responses.

ethical implications, given IP checks reduce participant anonymity. Further, it may not be appropriate if used with specific population groups (e.g., asylum seekers, perpetrators of potential crimes), where anonymity is often offered in exchange for more truthful or genuine responses for the purposes of research.

Data Screening and Cleaning.

Use automated and manual tools and researcher judgment to remove fraudulent responses prior to data analysis. For example, Qualtrics fraud detection tools can be used in combination with checking for inconsistencies in responses.

Manuscript Reporting. Include information on steps taken before, during, and after participant recruitment and data collection when writing up the manuscript. Be transparent about how many respondents were excluded and why.