OPEN ACCESS

Identifying and counteracting fraudulent responses in online recruitment for health research: a scoping review

Josielli Comachio ,[1] Adam Poulsen ,[2] Adeola Bamgboje-Ayodele ,[3,4] Aidan Tan ,[5,6] Julie Ayre ,[7] Rebecca Raeside ,[4] Rajshri Roy ,[8,9] Edel O'Hagan [10]

10.1136/bmjebm-2024-113170

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjebm-2024-113170).

For numbered affiliations see end of article.

Correspondence to:
Josielli Comachio; josielli.comachio@sydney.edu.au

To cite: Comachio J, Poulsen A, Bamgboje-Ayodele A, *et al*. *BMJ Evidence-Based Medicine* 2025;**30**:173–182.

Abstract

Objectives This study aimed to describe how health researchers identify and counteract fraudulent responses when recruiting participants online.

Design Scoping review.

Eligibility criteria Peer-reviewed studies published in English; studies that report on the online recruitment of participants for health research; and studies that specifically describe methodologies or strategies to detect and address fraudulent responses during the online recruitment of research participants.

Sources of evidence Nine databases, including Medline, Informit, AMED, CINAHL, Embase, Cochrane CENTRAL, IEEE Xplore, Scopus and Web of Science, were searched from inception to April 2024.

Charting methods Two authors independently screened and selected each study and performed data extraction, following the Joanna Briggs Institute's methodological guidance for scoping reviews and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews guidelines. A predefined framework guided the evaluation of fraud identification and mitigation strategies within the studies included. This framework, adapted from a participatory mapping study that identified indicators of fraudulent survey responses, allowed for systematic assessment and comparison of the effectiveness of various antifraud strategies across studies.

Results 23 studies were included. 18 studies (78%) reported encountering fraudulent responses. Among the studies reviewed, the proportion of participants excluded for fraudulent or suspicious responses ranged from as low as 3% to as high as 94%. Survey completion time was used in six studies (26%) to identify fraud, with completion times under 5 min flagged as suspicious. 12 studies (52%) focused on non-confirming responses, identifying implausible text patterns through specific questions, consistency checks and open-ended questions. Four studies examined temporal events, such as unusual survey completion times. Seven studies (30%) reported on geographical incongruity, using IP address verification and location screening. Incentives were reported in 17 studies (73%), with higher incentives often increasing fraudulent responses. Mitigation strategies included using in-built survey features like Completely Automated Public Turing test to tell Computers and Humans Apart (34%), manual verification (21%) and video checks (8%). Most studies recommended multiple detection methods to maintain data integrity.

Conclusion There is insufficient evaluation of strategies to mitigate fraud in online health

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Online recruitment for health research is increasingly popular but susceptible to fraudulent responses.

⇒ Fraudulent survey responses can compromise the integrity of health research data and result in incorrect study conclusions.

WHAT THIS STUDY ADDS

⇒ This scoping review revealed the various strategies used to identify and counteract fraudulent responses in online health research recruitment.

⇒ By synthesising these data, identifying strategies and exploring potential solutions, this review seeks to inform researchers about the extent of the issue and strategies to counteract fraudulent responses, thereby enhancing the quality and credibility of online research.

⇒ In the absence of a single effective strategy, researchers should use a combination of strategies to counteract fraudulent responses.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The findings underscore the urgent need for both comprehensive evaluation and subsequent enhancement of antifraud strategies in online health research.

research, which hinders the ability to offer evidence-based guidance to researchers on their effectiveness. Researchers should employ a combination of strategies to counteract fraudulent responses when recruiting online to optimise data integrity.

Introduction

In the digital age, the landscape of health research recruitment has been transformed by online platforms such as social media (eg, Facebook, X/Twitter, Instagram, Snapchat and TikTok).[1] These tools have revolutionised how researchers connect with potential study participants, offering unprecedented opportunities to reach diverse and widespread populations. Online recruitment methods streamline the participant enrolment process[2] and allow researchers to gather more representative samples[3 4] cost-effectively.[3]

The expansive reach and targeted capabilities of social media platforms also facilitate access to a wide range of individuals while providing a perceived sense of anonymity.[5] Using these platforms for recruitment, such as through paid advertisements and targeting of specific populations, enables researchers to connect with a broad audience without direct, in-person interaction. This sense of anonymity may help individuals respond more candidly.[6] However, it can also reduce accountability, potentially leading to lower engagement with the study materials, careless responses and even fraud.[6]

While online recruitment offers numerous advantages for health research, it is susceptible to various biases. Selection bias, whereby specific characteristics such as age, education or gender are over-represented or under-represented, is a persistent threat to data integrity.[7] Increasingly, the risk of fraudulent responses presents a significant challenge. Fraudulent responses occur when individuals or automated bots provide false information or misrepresent themselves to qualify for and participate in studies, including intentional duplicate entries and inaccurate data. These fraudulent activities can compromise the integrity and validity of the research data. If detected, researchers may need to increase sample sizes to exclude invalid responses[8]; however, if undetected, fraudulent responses can lead to incorrect study conclusions. To mitigate these risks, researchers must implement rigorous data validation techniques and employ advanced methods to identify and exclude fraudulent responses. Nonetheless, there is limited evidence guiding researchers on how to effectively identify and counteract these fraudulent responses.

This scoping review aims to describe how researchers identify fraudulent responses in health research when recruiting online and describe processes to counteract fraudulent responses.

Methods

Data sources and search

This scoping review adheres to the framework established by Arksey & O'Malley[9] and later recommendations by Levac *et al*.[10] Furthermore, this review follows the methodological framework set out by the Joanna Briggs Institute for scoping reviews and is conducted and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews.[11] The protocol was prospectively registered on the Open Science Framework (https://osf.io/fpdsb/).

Identification of the research question

The primary research question was 'How do health researchers identify and counteract fraudulent responses when recruiting research participants online?'.

Population, concept and context

The population of interest in this scoping review includes participants recruited for health research through online platforms. The concept under study is the prevalence and characteristics of fraudulent responses, which refer to instances where individuals provide false or misleading information during the recruitment process. By focusing on these elements, the review aims to provide valuable insights that can inform future strategies to identify and counteract these issues. Counteracting involves implementing measures to diminish or eliminate the impact of fraudulent responses, thereby enhancing the integrity and reliability of online health research recruitment.[12]

Search strategy

Nine research databases were systematically searched: Medline via Ovid, Informit, AMED (Allied and Complementary Medicine Database), CINAHL (Cumulative Index of Nursing and Allied Health Literature), Embase, Cochrane Central Register of Controlled Trials (CENTRAL), IEEE Xplore (Institute of Electric and Electronic Engineers), Scopus and Web of Science. Peer-reviewed articles, not limited to any design, published in English until April 2024, were eligible for inclusion. Keywords included "recruitment", "participant recruitment", "surveys and questionnaires", "online survey", "online platforms" and "social media". The full electronic search strategy is available in online supplemental file 1.

Inclusion and exclusion criteria

The inclusion criteria for this review were as follows: peer-reviewed studies published in English, irrespective of the study design or year of publication; studies that report on the online recruitment of participants for health research; and studies that specifically describe methodologies or strategies to detect and address fraudulent responses during the online recruitment of research participants. Health research included participants with specific health-relevant behaviours (eg, smoking) or populations with a health condition (eg, COVID-19). Given that qualitative studies necessitate participant interviews, which were presumed to inherently mitigate the risk of fraud, these studies were excluded from the analysis.

Fraudulent response in online surveys was defined as survey entries characterised by intentional deceit or invalid participation, such as submitting multiple entries to gain additional incentives, using automated systems (bots) to generate responses or providing false information to receive rewards.[13]

Study selection and data extraction

All the studies identified in the search strategy were exported into Covidence systematic review software. Duplicates were removed in Covidence.[14]

A group of seven investigators (AP, AB-A, AT, JA, RRa, RRo and EO'H) independently conducted the title and abstract screening to identify relevant studies. Subsequently, full texts were retrieved and assessed for eligibility. Pairs of investigators (JC, JA, RRa, AB-A or RRo) independently extracted data using a customised data extraction form developed specifically for this review. Descriptive summaries were provided for the general characteristics of the papers and the studies. Any disagreements on the extracted data were resolved through arbitration by a third investigator (EO'H).

The extracted data included information on (1) study details including the author(s), year of publication, study design (eg, randomised controlled trial, observational study), the specific

topic or research question being investigated, the sample size and any demographic characteristics of the sample (eg, age, gender, ethnicity and socioeconomic status). (2) Methods used for participant recruitment/advertising: We documented approaches employed by the studies to recruit or advertise to participants. This included the platforms used (ie, social media, online forums or dedicated research recruitment websites), the type of recruitment tools (eg, advertisements, emails and targeted posts) and any specific strategies or protocols followed to attract and engage potential participants (ie, incentives). (3) Strategies to identify fraudulent responses: methods and criteria used by the studies to detect fraudulent or suspicious participant responses. We recorded whether the studies used techniques such as completion time analysis, consistency checks, IP address tracking or other indicators to flag potentially fraudulent data. (4) Strategies to mitigate fraudulent responses: any strategies implemented to prevent or minimise the occurrence of fraudulent responses.

We obtained and verified data from investigators by contacting the corresponding authors of the included studies via email when required. If the authors did not respond, we sent a reminder email after 2 weeks to follow up on the request.

Data analysis

The sample was descriptively analysed to provide a comprehensive overview. The number of fraudulent and valid responses relative to the total responses recorded by the authors of each study was reported when available. Identifying and counteracting fraud are often intertwined, with the identity of fraud immediately leading to the exclusion of that participant and therefore counteracting fraud. In this review, strategies used to identify fraud were described using a predefined framework.[12] The framework was developed by researchers in response to their experience encountering fraud in two web participatory mapping experiments conducted in 2021. The study involved an online participatory mapping platform where participants were asked to map landscape values and perceived risks, alongside completing a traditional survey questionnaire. The authors compared results and from these developed five indicators to detect fraudulent survey responses, including completion time, non-conforming responses, temporal events, unexpected language and spatial incongruity. Using this framework facilitated the systematic assessment and comparison of the effectiveness of different strategies across studies.

Results

The search identified 2289 records. After duplicates were removed, 1826 records underwent title and abstract screening. Then, full-text article screening was performed on 34 articles retrieved. Finally, a total of 23 articles were included. The process of selecting studies is outlined in figure 1.

Of the 23 studies, 18 were quantitative and 5 were qualitative in design. Most studies were published within the last 5 years (22/23, 96%) and from the USA (19/23, 83%). When reported, most used the online survey platform Qualtrics (9/16, 56%) and recruited participants from social media (ie, Facebook, X/Twitter) (11/19, 57%).

The included articles most commonly investigated specific health behaviours or populations, including opioid users,[15] alcohol, tobacco, and other drugs,[16–19] cancer survivors,[20] COVID-19 patients,[20–23] and people living with HIV.[24–26]

All included studies used multiple methods for detecting fraud. 18 studies (78%) provided data on the number of participants excluded due to fraudulent or suspicious online survey participation relative to the total sample recruited. A summary of the characteristics of the included studies can be found in table 1. More details can be found in online supplemental file 3.

Identifying fraudulent responses

A summary of the fraud detection strategies employed and their evaluation is presented in table 2.

Detecting fraudulent responses

Survey completion time

This indicator was defined as inadequate survey completion time to complete the survey accurately and can be estimated by pilot testing.[12] This method was reported in six studies (6/23, 26%) included in this scoping review.[22 24 26–28] In their qualitative review, Singh and Sagar[23] also reported response time as a method of identifying fraud from online surveys. One study[22] reported that the same start and stop as an indicator of fraud, and one[24] reported using a response time of less than 5 min as part of a 10-point algorithm to detect fraud anticipated to require a maximum of 45 min to complete. Two studies (8%)[27 28] reported using the embedded date/time stamps to identify quick completions but did not state a specific time frame that would indicate fraud.

Non-confirming responses

This indicator was defined as text responses with unusual and repeated combinations that were perceived as implausible (eg, reporting an average daily smartphone use of more than 24 hours or exhibiting an 'all or none' response pattern). 12 studies (12/23, 52%) included in this scoping review reported on methods to identify and remove non-confirming responses.

Two studies[27 29] reported asking respondents specific open-ended questions, such as where they heard about the study or included directives (eg, select the third option below).[29] Other studies included pairs of questions[29] or reported collecting the same information at multiple points to detect inconsistencies in responses that might indicate fraudulent or careless participation.[19 22 27]

Four studies attempted to identify fraudulent responses by including questions that required insider knowledge.[27 29] Stuart and Moore[30] asked participants to self-disclose inclusion criteria. However, Levi *et al*[28] reported that false respondents were able to ascertain highly specific study eligibility requirements. Another strategy reported was to include an open-ended question to identify the possible use of bot automation,[29 30] where responses were reviewed for English language misuse, incoherent or nonsensical phrases, duplicate responses, and content copied verbatim from existing websites.

Glazer *et al*[31] reported using TransUnion's TLOxp (www.tlo.com) to verify publicly available information, including names, address history, date of birth, social security number, phone and email information, and possible relatives. These verification strategies were also highlighted in the qualitative results. For instance, Singh and Sagar[23] recommended checking for inconsistencies and implausible response patterns. Additionally, Pratt-Chapman *et al*[20] emphasised the value of specific questions (eg, where a respondent heard about the survey), open-ended questions, checking question pairs and directive questions.

Temporal events

This indicator was defined as survey time stamps outside the target population's expected norms. Four studies included in this scoping review reported on this indicator. Loebenberg *et al*,[18]

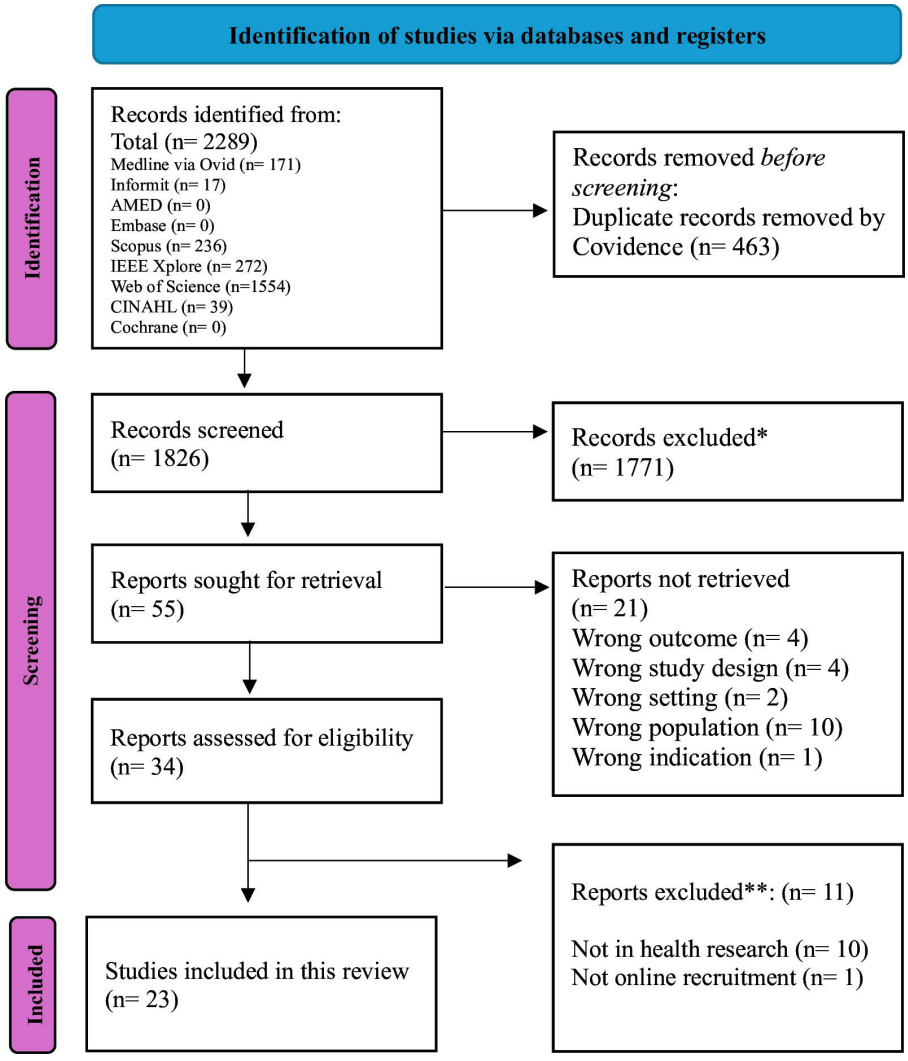

Figure 1 Flow diagram of the study selection process.

***Not online recruitment for health research and not related topics. **The list of excluded full-text articles assessed for eligibility can be found in online supplemental file 2.**

based in the UK, reported that bots were identifiable because they usually joined the study at odd night hours. Wang *et al*,[21] from the USA, used checking the time zone as a strategy to identify fraudulent responses. Similarly, Pozzar *et al*[29] included survey time stamp checks as part of various strategies used to identify fraudulent responses. Pratt-Chapman *et al*[20] also included this in their qualitative report on identifying fraudulent responses.

Geographical incongruity

This indicator was defined as geographic stamps that sit outside the expected norms for the target population. Seven studies (30%) reported on this indicator. Three verified internet protocol (IP) addresses,[24 27 29] whereas one included location screening.[32] Bonett *et al*[22] included four components to check for geographical incongruity, including (1) checking that the address matched the colloquial name of a neighbourhood, (2) verifying the ZIP code was inside the target location, (3) verifying that the IP address was not from outside USA, or that (4) identifying whether participants were using a virtual private network (VPN).[22] In addition, they did not permit more than three people to use the same residential address.[22] One study[29] reported using a virtual private server (VPS) with the Routing Information Protocol or Shiny tool from the R open-source programming language. This approach involves analysing the IP addresses of survey respondents and seeing if there are signs that they might be using a VPS.

Leobenberg *et al*[18] checked for geographical incongruity by checking postcodes against the first line of the street address. Levi *et al*[28] reported that they geo-located each participant using survey meta-data and cross-checked with partner databases. In addition, qualitative reports by Singh and Sagar,[23] Godinho *et al*[33] and Chandler *et al*[34] included IP checks as a recommendation for detecting fraud.

Counteracting fraudulent responses

Incentives

17 (73%) of the included studies reported offering incentives to participants. The incentives, in the form of gift cards, ranged in value from US$5 to US$200. In one study, researchers varied the rate of financial incentives to determine their influence on the rate of deception. They observed that increasing the incentive value heightened the likelihood of fraudulent responses.[35] In another study, the researchers deliberately did not advertise the amount or type of compensation,[27] while another study purposefully offered no incentive.[30] Another strategy to detect fraud was to lower the

Table 1 Summary of characteristics of the included studies

Author, year	Incentives offered for study completion	Total sample screened	Participants excluded for fraudulent or suspicious responses	Total valid entries	Types of fraudulent or suspicious data
1. Ballard *et al*[24] 2019	Yes	414	161 (38%)	253	Ineligible geolocation; invalid phone number; mismatching names; unusual email addresses.
2. Bonett[22], 2024	Yes	7950	4207 (52%)	3743	Ineligible geolocation; invalid residential address; repeated email address; invalid recruitment URL.
3. Campbell *et al*[25] 2022	Yes	739	413 (55%)	326	Geolocation data; repeated responses suggesting bots.
4. Chandler *et al*[34] 2020	NR	NR	NR	NR	Poor-quality responses, where participants may not pay sufficient attention to survey questions.
5. Fernandez-Lynch,[35] 2019	Yes	2275	500 (21%)	1775	Participants misreported their eligibility based on varying payment incentives and eligibility criteria for participation.
6. Glazer *et al*[31] 2021	Yes	1766	125 (7%)	1641	Potential or confirmed discrepancy in provided information on interest form.
7. Guest *et al*[26] 2021	Yes	13 931	493 (3%)	3253	Ineligible geolocation; incompletion of open-ended questions.
8. Godinho *et al*[33] 2020	Yes	860	276 (32%)	506	Incomplete addresses; attempted to enrol into the study more than once; fake addresses identified were created by one individual or group.
9. Habib and Jha[16] 2021	Yes	16	NR	NR	Multiple entries.
10. Hammond,[17] 2022	Yes	1503	324 (21%)	1179	Respondent replaced their face recording with a stock photo or video from the internet; used a private VPN to change the IP address.
11. Hohn *et al*[32] 2022	NR	45	0	45	Strategies such as multiple questions, several surveys were entered by respondents but not completed, survey time stamps and locations were employed.
12. Kramer *et al*[27] 2014	Yes	600	293 (48%)	307	Illogical responses to related paired items; shorter than expected time to complete the survey and inconsistently reported personal information such as names.
13. Levi *et al*[28] 2022	Yes	482	182 (37%)	300	Geolocation did not match the self-reported address; geolocation was located outside of the recruiting counties; and the name could not be verified in our partner databases; same IP address.
14. Loebenberg *et al*[18] 2023*	Yes	5562	1161 (20%)	5602	Invalid contact numbers where participants were not known; postcode did not match the first line of the address; the same landline number provided by multiple respondents with different geographical postal addresses.
15. Pozzar *et al*[29] 2020	Yes	271	256 (94%)	15	Exhibited evidence of bot automation; provided a duplicate or unusual response to one or more open-ended items and exhibited evidence of inattention.
16. Pratt-Chapman *et al*[20] 2021	Yes	1977	1081 (54%)	569	Open-ended suggestion responses are duplicates; Respondents indicate a survey source prior to dissemination of the survey from that source; email/address was suspicious or duplicate.
17. Singh and Sagar[23] 2021	NA	NA	NA	NA	Inconsistent (eg, age and date of birth responses not consistent) or implausible response patterns (eg, average daily smartphone use of greater than 24 hours, 'all or none' response pattern).
18. Stuart and Moore[30] 2021	NR	704	168 (23%)	536	Inconsistency in the open-ended questions responses; incomplete entries.
19. Wang *et al*[21] 2023†	Yes	950	NR	NR	Survey time stamps and locations; unusual email format; response to optional open-ended question text identical to another respondent; Email/address/telephone number already reported by another respondent.
20. Wilkerson *et al*[19] 2015	Yes	543	107 (19%)	427	Duplicates among the IP addresses-mail addresses, and key demographic characteristics.

Continued

Table 1 Continued

Author, year	Incentives offered for study completion	Total sample screened	Participants excluded for fraudulent or suspicious responses	Total valid entries	Types of fraudulent or suspicious data
21. Willis *et al*[38] 2023	Yes	1241	268 (21%)	NR	Duplicates emails; short time completion (›5 min).
22. Wood and Bindler[36] 2023	Yes	109	43 (39%)	NR	Inconsistencies in the responses provided.
23. Young *et al*[15] 2020	Yes	410	67 (16%)	234	Ineligible geolocation; incompletion of open-ended questions.

*863 study entries were identified as bots during data screening in the first 2 months of recruitment. Out of the 5956 participants (excluding bots) who enrolled, 298 (5%) were found to be false participants.

†After Facebook advertisements for an hour and reaching approximately 125 individuals, 2578 screeners were attempted and 950 survey responses were completed, however, it is not clear how many indicated possible fraud. Hammond, 2022: Total sample 276 from the community and 1227 from the panel; valid entries:114 from the community and 1065 from the panel.

IP, internet protocol; NA, not applicable; NR, not reported; URL, uniform resource locator; VPN, virtual private network.

value of the remuneration and only send it to a physical address.[29] Habib and Jha[16] reported using the anonymous incentive method to randomly select participants for a reward. In their qualitative review, Pratt-Chapman *et al*[20] recommended that researchers explicitly state that fraudulent responses will not be compensated.

In-built survey features

Eight studies (34%) reported using in-built survey options to prevent fraud. Ballard *et al*[24] reported that they ensured that the 'prevent ballot box stuffing' option in Qualtrics was activated. This option prevents participants from taking a survey more than once by placing a cookie on their browser when they submit a response. Other features are activated in Qualtrics to prevent indexing, which blocks search engines from finding the survey and presenting it in search results.[24]

Eight studies (34%)[20–22 25 27 29 32 36] reported using a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) feature. The Q_RecaptchaScore is a specific type of CAPTCHA developed by Google. Both serve as an indicator in reports to determine the likelihood that a response was generated by a bot rather than a human.[37] Bonett *et al*[22] set cut-off levels for fraud, including (1) Q_RecaptchaScore >0.5 (score ranging from 0 to 1, greater than or equal to 0.5 means the respondent is likely a human) and (2) Q_RelevantIDFraudScore >30 (score ranging from 0 to 100, with higher values indicating a greater likelihood of fraud). Others reported using specific survey features to identify fraudulent responses, such as speed bump questions (ie, difficult answer questions designed to screen out bots and automatic responders)[21] or honey pot questions (ie, hidden questions only detectable by bots).[20–22]

In addition, two studies (8%)[22 27] checked the uniform resource locator (URL) to ensure that it matched the URL distributed by the study team. In contrast, Campbell *et al*[25] reported that failure to verify via a unique web link was an indication of fraud.

Manual verification

Manual verification refers to when a researcher contacts a participant to verify their identity or manually checks survey responses for inconsistencies. Manual verification in the included studies was considered an additional step in the recruitment process and data-quality check despite being more laborious compared with automated strategies.[33 34]

Five studies (21%)[25–27 29 32] reported manual verification to identify fraud, and three studies (13%)[23 33 34] included manual checking in their literature review as a recommended strategy to detect fraud. Campbell *et al*[25] described a manual review of screening results to identify potential fraudulent responses, with suspected cases followed up by phone. Guest *et al*[26] reported manual checks of age, IP address, completion scores and time stamps for data inconsistencies. Hohn *et al*[32] advocated for a data integrity plan outlining manual checks to prevent fraudulent activity. Kramer *et al*[27] suggested dual-method contact (phone and email) for respondent verification. Pozzar *et al*[29] described quality assurance checks by the research team for completed surveys. One study[36] employed videoconferencing (eg, Zoom) scheduled at the participant's convenience to verify their identity.

Automated facial response analysis

One study[17] embedded public service announcements of 30–40 s within a survey and a computer webcam recorded their facial expressions. The records were analysed using automated facial coding engines.

Response when fraud was suspected

For most studies (78%), the evaluation of the fraudulent responses was intertwined with detection. If fraud was detected, then the participant's responses were removed. Some authors recommended utilising a fraud detection algorithm[16 20 24 27 29] and advising participants that fraudulent responses would be monitored. Others conducted monthly data checks,[17 18 26] and one asked participants to use their professional email to verify participant identity.[38]

Discussion

This scoping review identified 23 studies focusing on strategies to detect and counteract fraudulent responses in online health research recruitment. While no single method prevented invalid or fraudulent responses combining multiple strategies can enhance data integrity. Most of these studies (83%) were conducted in the USA, published within the last 5 years (2019–2024) and primarily used Qualtrics as an online survey platform and a mix of recruitment channels (eg, Facebook, Instagram and web-based platforms). This review uncovered various methods for detecting fraud in online recruitment, but their evaluation was inconsistent, failing to outline how researchers could counteract the risk.

Although online recruitment offers many opportunities for health research,[2 39] it also carries the risk that research participants may exaggerate or distort their experience. Without initial direct contact, detecting dishonest responses becomes more challenging

Table 2 Overview of fraud detection and mitigation methods and effectiveness across studies

Strategy	Specific features	Evidence for effect
Detecting fraud		
Survey completion Time	► Short completion time identifies suspicious responses (eg, completion time less than 5 min)	Not evaluated
Non-confirming responses	► Incorrect response to an open-ended question; including specific directives (eg, in 5 words describe what this survey is about) identifies suspicious responses ► Inconsistent responses to pairs of questions (when the same question is asked twice) identify suspicious responses ► Incorrect response to a question that requires specific knowledge identifies suspicious responses ► Inconsistent responses when compared with publicly available information identify suspicious responses ► Anonymous email format (eg, 12458@hotmail.com) identifies suspicious responses	These measures may reduce automated attacks and reduce time spent manually checking responses but do not prevent invalid responses[27]
Temporal events (odds night hours responses)	► Survey time response outside of daytime hours when compared with current time zone identifies suspicious responses	Not evaluated
Geographical incongruity	► Geographical verification: the address provided by the participant does not exist in the real-world, does not match the expected local neighbourhood's name or the ZIP code is outside the target location identifies suspicious responses ► The IP address is outside the expected country, or the participants are using a VPN identifies suspicious responses	Not evaluated
Counteracting fraud		
Incentives	► Decrease incentive value or deliver incentive to physical addresses only. ► Do not advertise the amount or type of compensation. ► Do not provide incentives or clearly state that fraudulent responses will not be compensated.	Higher incentives led to participants providing fraudulent responses.[18 35]
In-built survey features	► Activate the 'prevent ballot box stuffing' option in survey platforms like Qualtrics to ensure that participants cannot take the survey more than once. This feature works by placing a cookie on the participant's browser when they submit a response. ► Search engine blocking: To prevent unauthorised access, features that block search engines from indexing the survey can be activated, ensuring the survey does not appear in search results	Not evaluated
CAPTCHA feature	► Add CAPTCHA field (eg, text-based, 3D CAPTCHA, image) can block automated bots, it cannot stop fraudulent responses from actual people who may be incentivised to manipulate the survey and provide false information.	CAPTCHA, attention checks, can minimise bots, however, alone is unlikely to prevent fraudulent responses in internet-based research promoted on social media[18 20]
Manual verification	► Perform regular checks of survey responses including data audits, IP addresses and time stamps ► Apply dual-method contact, for example, phone and email for respondent verification. ► Review responses to optional open-ended question text to ensure that it is not identical to that of another respondent ► Include video checks (eg, video conferencing with the participant) or add a public service announcements and computer webcam to record their facial expressions to verify identify.	Recommended but not evaluated[17 36]

CAPTCHA, Completely Automated Public Turing test to tell Computers and Humans Apart; VPN, virtual private network.

for researchers.[40] A multilayered approach is likely necessary given the variety and evolving nature of fraudulent activities, ranging from simple to complex schemes involving sophisticated techniques. Relying on a single method may fail to detect certain types of fraud. Consequently, included studies have reported a combination of approaches to counteract fraudulent responses in the context of increasing online recruitment demand.[12 17 19 21 25 35]

In this scoping review, manual verification was used by 34% of included studies, and monthly data checks, along with the use of professional emails, are recommended.[41] However, these strategies have undergone limited evaluation, leading to ambiguity in their effectiveness. For example, about temporal events, there is no consensus on what constitutes 'late' survey entries. Furthermore, higher monetary compensation appears to correlate with an increased risk of fraudulent responses.[20 32] While fraud prevention measures like telephone calls, repeated questions or reduced incentives may mitigate this risk, they can also increase participant burden, reduce accessibility and potentially compromise the generalisability of the sample.[13 42]

There are several automated features to detect fraud. In-built survey features were common, with 34% of studies using options like the cookies option and reCAPTCHA in Qualtrics to prevent fraud.[20–22 25 27 29 32 36] Specific tactics included speed bump questions, honey pot questions and URL verification to ensure survey integrity. CAPTCHA schemes (eg, text-based, image-based, audio-based, video-based or cognitive-based CAPTCHAs) offer several

advantages: they are easy to implement, consistently applied across studies and often more sophisticated than other manual strategies.[43] However, they can still be vulnerable to automated and human-assisted fraud.[44] While CAPTCHA schemes offer certain benefits, their overall effectiveness remains poorly documented, and the lack of transparency in their evaluation highlights a critical need for continuous assessment and refinement. Currently, it is challenging to provide evidence-based recommendations to researchers regarding their effectiveness.

Fraud mitigation strategies have also been explored in non-health research, and the approaches to preventing fraudulent responses differ significantly between health and non-health surveys, reflecting the unique challenges each domain faces.[45] Some researchers consider transitioning to prevetted participants through third-party companies as a solution to these issues.[46] However, while prevetting can reduce the likelihood of fraudulent responses, it introduces new challenges, such as higher costs and potential biases that may limit the generalisability of findings. The use of pre-vetted participants can skew the sample towards individuals recruited multiple times for different studies, thus possibly affecting the diversity and representativeness of the research population.[47]

Health surveys are known to require more rigorous methods due to the sensitive nature of data, ethical considerations and the vulnerability of participants, often compounded by incentives like monetary compensation or potential medical treatment. Despite this need for rigour, challenges such as ethical concerns about privacy, reduced participation due to intrusive methods and balancing data integrity with participant burden persist.[28 46] Non-health surveys, on the other hand, could typically employ simpler fraud detection methods like CAPTCHA or IP address verification, which are less intrusive but also less robust. This contrast underscores the importance of context-specific approaches, where health surveys must carefully balance rigorous fraud detection with maintaining participant trust and engagement.

Our results mirror those of Ziyi *et al*[48] who systematically evaluated the effectiveness of tests in preventing and detecting fraudulent responses across two different online surveys. Those authors tested 28 strategies, including reCAPTCHA, VPN detection, checking for consistency (by asking participants for their occupation twice with the question written differently both times) and checking IP address. Their results show that inconsistent knowledge about the subject was the most useful indicator of fraud, although reported that tests related to consistency and attention were also useful for detecting fraud. They concluded that there is no single perfect strategy to prevent and detect invalid responses in online surveys.[48]

Lawlor *et al*[49] proposed the REAL framework (Reflect, Expect, Analyse, Label) to guide fraud prevention in online surveys. While studies included in this review generally adopted analysis strategies, the framework's emphasis on reflection, expectation setting, and labelling of fraudulent responses was less evident in the included studies. Lawlor *et al*[49] also highlight the potential for fraud to occur postrecruitment, which was not addressed by the studies included in this review. This broader perspective, which advocates for reflection, expectation setting, and labelling as critical components of fraud prevention, remains underused. The limited adoption of these practices suggests that further refinement and application of such comprehensive strategies are necessary to more effectively address the persistent challenge of survey fraud.

Strengths and limitations of the study

This scoping review has several strengths. Among its strengths, addressing the strategies to identify and counteract fraudulent responses collected through social media is timely and relevant, given the increasing use of these platforms in health research. This study focused on quantitative studies in health research and compiled strategies including studies from comprehensive databases. Additionally, the review highlights the critical need for more rigorous evaluation and development of fraud detection strategies in the context of social media recruitment, drawing attention to an underexplored area in the existing literature. By addressing this gap, the review emphasises the importance of providing clear guidance for researchers to improve data integrity, ensuring that online studies can yield reliable and valid results.

However, the study also has some limitations. Initially, our inclusion criteria were limited to English-language studies. However, given the increasing accessibility of language translation tools and the potential for valuable insights from international research, we acknowledge this as a limitation. While no non-English studies were identified in the current review, future research may benefit from a more inclusive approach. Additionally, we did not report on other serious biases such as selection bias, as it was not explicitly addressed in the included studies.

Implications for future research and policy

In the absence of evidence for one effective strategy, researchers should implement multiple fraud detection strategies, conduct pilot tests to establish benchmarks, and continuously monitor and update their recruitment methods. Collaboration between researchers, platform providers and policy-makers is necessary to develop robust fraud detection and prevention frameworks, guidelines and best practices. Future research should explore under-researched indicators, such as unexpected language settings and conduct comparative studies to evaluate the effectiveness of fraud detection methods. Additionally, emerging technologies like machine learning and artificial intelligence could enhance fraud detection capabilities, leading to more effective online health research recruitment solutions.

Conclusion

Effective strategies for detecting online research fraud remain elusive. This review outlines commonly employed methods and highlights the need for rigorous evaluation. Given the absence of a gold standard, researchers should implement a multifaceted approach to safeguard data integrity.

Author affiliations
[1]Sydney Musculoskeletal Health, School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
[2]Brain and Mind Centre, The University of Sydney, Sydney, New South Wales, Australia
[3]Biomedical Informatics and Digital Health, School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
[4]Susan Wakil School of Nursing and Midwifery, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
[5]NHMRC Clinical Trials Centre, Camperdown, New South Wales, Australia
[6]Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
[7]Sydney Health Literacy Lab, Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
[8]Discipline of Nutrition and Dietetics, Susan Wakil School of Nursing and Midwifery, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia

[9]Charles Perkins Centre, The University of Sydney, Sydney, New South Wales, Australia
[10]Westmead Applied Research Centre, School of Medicine, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia

X Josielli Comachio @josiicomachioo, Adam Poulsen @AdamPoulsen and Edel O'Hagan @EdelOH

Acknowledgements The authors acknowledge the Digital Health and Informatics Network's Early Career Researcher community, within the University of Sydney, for providing the opportunity for the coauthors to collaborate on this project. We also acknowledge the leadership of the Digital Health and Informatics Network for their support of this collaborative work. In addition, we acknowledge Kanchana Ekanayake, the Academic Liaison Librarian, for her invaluable assistance in crafting the comprehensive search strategy.

Contributors This work represents a collaborative effort by a subgroup of EMCR researchers from the Digital Health Informatics Network's early and mid-career researchers (EMCR) committee with the University of Sydney. In recognition of their equal contributions to the project, with an exception for the first and senior authors, authorship is listed alphabetically. All authors approved the final version of the article. All authors had access to all the data in the study and could take responsibility for the integrity of the data. E'OH is guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Only a secondary analysis of published literature was performed. Therefore, this study does not involve human or animal participants and does not require ethical approval or participant consent.

Provenance and peer review Not commissioned; internally peer reviewed.

Data availability statement No data are available. Extracted data and analyses are available in supplementary files. Any further data are available from the corresponding author on reasonable request.

ORCID iDs
Josielli Comachio http://orcid.org/0000-0002-3286-6998
Adam Poulsen http://orcid.org/0000-0002-0001-3894
Adeola Bamgboje-Ayodele http://orcid.org/0000-0002-5629-1236
Aidan Tan http://orcid.org/0000-0003-0354-4006
Julie Ayre http://orcid.org/0000-0002-5279-5189
Rebecca Raeside http://orcid.org/0000-0003-2016-6393
Rajshri Roy http://orcid.org/0000-0002-7939-1790
Edel O'Hagan http://orcid.org/0000-0002-1914-5918

References

1 **Goldman N**, Willem T, Buyx A, *et al*. Practical Benefits, Challenges, and Recommendations on Social Media Recruitment: Multi-Stakeholder Interview Study. *J Med Internet Res* 2023;25:e44587.
2 **Whitaker C**, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. *J Med Internet Res* 2017;19:e290.
3 **Hudnut-Beumler J**, Po'e E, Barkin S. The Use of Social Media for Health Promotion in Hispanic Populations: A Scoping Systematic Review. *JMIR Public Health Surveill* 2016;2:e32.
4 **Guillory J**, Wiant KF, Farrelly M, *et al*. Recruiting Hard-to-Reach Populations for Survey Research: Using Facebook and Instagram Advertisements and In-Person Intercept in LGBT Bars and Nightclubs to Recruit LGBT Young Adults. *J Med Internet Res* 2018;20:e197.
5 **Ayers JW**, Poliak A, Dredze M, *et al*. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med* 2023;183:589–96.
6 **Speed E**, Davison C, Gunnell C. The anonymity paradox in patient engagement: reputation, risk and web-based public feedback. *Med Humanit* 2016;42:135–40.
7 **Haddad C**, Sacre H, Zeenny RM, *et al*. Should samples be weighted to decrease selection bias in online surveys during the COVID-19 pandemic? Data from seven datasets. *BMC Med Res Methodol* 2022;22:63.
8 **Bauermeister J**, Pingel E, Zimmerman M, *et al*. Data Quality in web-based HIV/AIDS research: Handling Invalid and Suspicious Data. *Field methods* 2012;24:272–91.
9 **Arksey H**, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005;8:19–32.
10 **Levac D**, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci* 2010;5:69.
11 **Page MJ**, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
12 **Johnson MS**, Adams VM, Byrne J. Addressing fraudulent responses in online surveys: Insights from a web-based participatory mapping study. *P and N* 2024;6:147–64.
13 **Teitcher JEF**, Bockting WO, Bauermeister JA, *et al*. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. *J Law Med Ethics* 2015;43:116–33.
14 **Covidence**. Covidence systematic review software: veritas health innovation. Melbourne, Australia,
15 **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* 2020;135:746–55.
16 **Habib D**, Jha N. AIM against survey fraud. *JAMIA Open* 2021;4:ooab099.

17 **Hammond RW**, Parvanta C, Zemen R. Caught in the Act: Detecting Respondent Deceit and Disinterest in On-Line Surveys. A Case Study Using Facial Expression Analysis. *Soc Mar Q* 2022;28:57–77.
18 **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.
19 **Wilkerson JM**, Shenk JE, Grey JA, *et al*. Recruitment Strategies of Methamphetamine-Using Men Who Have Sex with Men into an Online Survey. *J Subst Use* 2015;20:33–7.
20 **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* 2021;7:e30730.
21 **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:e0001452.
22 **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* 2024;8:e47091.
23 **Singh S**, Sagar R. A critical look at online survey or questionnaire-based research studies during COVID-19. *Asian J Psychiatr* 2021;65:102850.
24 **Ballard AM**, Cardwell T, Young AM. Fraud Detection Protocol for Web-Based Research Among Men Who Have Sex With Men: Development and Descriptive Evaluation. *JMIR Public Health Surveill* 2019;5:e12344.
25 **Campbell CK**, Ndukwe S, Dubé K, *et al*. Overcoming Challenges of Online Research: Measures to Ensure Enrollment of Eligible Participants. *J Acquir Immune Defic Syndr* 2022;91:232–6.
26 **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* 2021;9:e28232.
27 **Kramer J**, Rubin A, Coster W, *et al*. Strategies to address participant misrepresentation for eligibility in Web-based research. *Int J Methods Psychiatr Res* 2014;23:120–9.
28 **Levi R**, Ridberg R, Akers M, *et al*. Survey Fraud and the Integrity of Web-Based Survey Research. *Am J Health Promot* 2022;36:18–20.
29 **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:e23021.
30 **Stuart WP**, Moore B. Evidence-Based Facebook Recruitment of Study Participants. *Comput Inform Nurs* 2021;39:355–61.
31 **Glazer JV**, MacDonnell K, Frederick C, *et al*. Liar! Liar! Identifying eligibility fraud by applicants in digital health research. *Internet Interv* 2021;25:100401.
32 **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* 2022;11:e38550.
33 **Godinho A**, Schell C, Cunningham JA. Out damn bot, out: Recruiting real people into substance use studies on the internet. *Subst Abus* 2020;41:3–5.
34 **Chandler J**, Sisso I, Shapiro D. Participant carelessness and fraud: Consequences for clinical research and potential solutions. *J Abnorm Psychol* 2020;129:49–55.
35 **Fernandez Lynch H**, Joffe S, Thirumurthy H, *et al*. Association Between Financial Incentives and Participant Deception About Study Eligibility. *JAMA Netw Open* 2019;2:e187355.
36 **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* 2024;80:2970–6.
37 **Fraud detection**. Qualtrics. 2022. Available: https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/ 2022
38 **Willis TA**, Wright-Hughes A, Skinner C, *et al*. The detection and management of attempted fraud during an online randomised trial. *Trials* 2023;24:494.
39 **Batterham PJ**. Recruitment of mental health survey participants using Internet advertising: content, characteristics and cost effectiveness. *Int J Methods Psychiatr Res* 2014;23:184–91.
40 **Atherton H**, Sawmynaden P, Sheikh A, *et al*. Email for clinical communication between patients/caregivers and healthcare professionals. *Cochrane Database Syst Rev* 2012;11:CD007978.
41 **Chen AT**, Komi M, Bessler S, *et al*. Integrating statistical and visual analytic methods for bot identification of health-related survey data. *J Biomed Inform* 2023;144:104439.
42 **Rankin KM**, Rauscher GH, McCarthy B, *et al*. Comparing the reliability of responses to telephone-administered versus self-administered Web-based surveys in a case-control study of adult malignant brain cancer. *Cancer Epidemiol Biomarkers Prev* 2008;17:2639–46.
43 **Dinh NT**, Hoang VT. Recent advances of Captcha security analysis: a short literature review. *Procedia Comput Sci* 2023;218:2550–62.
44 **Gutub A**, Kheshaifaty N. Practicality analysis of utilizing text-based CAPTCHA vs. graphic-based CAPTCHA authentication. *Multimed Tools Appl* 2023;2023:1–33.
45 **Goodrich B**, Fenton M, Penn J, *et al*. Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys. *Applied Eco Perspectives Pol* 2023;45:762–84.
46 **Kuehn BM**. Companies' Use of Web to Recruit Patients for Studies Brings Opportunities, Risks. *JAMA* 2008;299:2733.
47 **Qureshi N**, Edelen M, Hilton L, *et al*. Comparing Data Collected on Amazon's Mechanical Turk to National Surveys. *Am J Health Behav* 2022;46:497–502.
48 **Ziyi Z**, Shuofei Z, Jaron M, *et al*. Beyond bot detection: combating fraudulent online survey takers, Lyon, France. Proceedings of the ACM Web Conference 2022; Lyon, France, 2022
49 **Lawlor J**, Thomas C, Guhin AT, *et al*. Suspicious and fraudulent online survey participation: Introducing the REAL framework. *Methodol Innov* 2021;14:20597991211050467.