Effective recruitment or bot attack? The challenge of online research surveys and recommendations to reduce risk and improve robustness: a case study commentary

Liesje Donkin, Nathan Henry, Amy Kercher, Mangor Pedersen, Holly Wilson, Amy Hai Yan Chan

Submitted to: Interactive Journal of Medical Research
on: May 15, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Effective recruitment or bot attack? The challenge of online research surveys and recommendations to reduce risk and improve robustness: a case study commentary

Liesje Donkin[1]; Nathan Henry[1]; Amy Kercher[1]; Mangor Pedersen[1]; Holly Wilson[2] BSocSci(Hons), MSc; Amy Hai Yan Chan[2] BPharm(Hons), PhD

[1]Auckland University of Technology Auckland NZ
[2]University of Auckland Auckland NZ

Corresponding Author:
Liesje Donkin
Auckland University of Technology
Private Bag 92006
Auckland
NZ

Abstract

Online research has exploded in popularity in recent years, enabling researchers to offer both investigations and interventions to broader participant populations than ever before. However, challenges associated with online research have also increased – notably difficulties verifying participant data and deliberate data manipulation by bot and spam responses. To encourage researchers to reflect on the impact of bot attacks on research and how to manage this.* This article presents two case studies where online research was affected by bot and spam attacks, targeting the offer of compensation for research participants and recommendations based on these experiences are made. Screening and verification processes utilised are presented. Based on our experience, security and screening within online research platforms are partly effective, but no solution is available to protect researchers completely against bot attacks Implications for future research and advice for health researchers are discussed.

(JMIR Preprints 15/05/2024:60548)
DOI: https://doi.org/10.2196/preprints.60548

Original Manuscript

Effective recruitment or bot attack? The challenge of online research surveys and recommendations to reduce risk and improve robustness.

Liesje Donkin[1], Nathan Henry[2], Amy Kercher[1], Mangor Pedersen[1], Holly Wilson[3] and Amy Hai Yan Chan[3]

1) Department of Psychology and Neuroscience, School of Clinical Sciences, Auckland University of Technology, Auckland, New Zealand
2) Auckland University of Technology, Auckland, New Zealand
3) School of Pharmacy, The University of Auckland, Auckland, New Zealand

Abstract

Online research has exploded in popularity in recent years, enabling researchers to offer both investigations and interventions to broader participant populations than ever before. However, challenges associated with online research have also increased – notably difficulties verifying participant data and deliberate data manipulation by bot and spam responses. This article presents two case studies where online research was affected by bot and spam attacks, targeting the offer of compensation for research participants and recommendations based on these experiences are made. Screening and verification processes utilised are presented. Based on our experience, security and screening within online research platforms are partly effective, but no solution is available to protect researchers completely against bot attacks. Implications for future research and advice for health researchers are discussed.

Keywords: online research; research methodology; surveys; data integrity; bot attacks; technology

Corresponding Author

Associate Professor Liesje Donkin

Liesje.Donkin@aut.ac.nz

Auckland University of Technology (AUT)

Private Bag 92006

Auckland 1142

New Zealand

Introduction

The internet and digital technologies have irretrievably changed the conduct of research. Whether it be participant engagement, digital intervention delivery, data collection, or distribution of findings, technology has allowed researchers to engage in wider-reaching recruitment. Technology has also enabled people to access and participate in research that may not have been previously possible. In particular, health researchers have embraced online recruitment to invite participation from varied population groups easily and to collect data from large samples.

There has been a significant increase in sample sizes in the last decades, intending to increase statistical power, replicability, and generalizability of research findings [1]. The adoption of open science practices and the emergence of global collaborative efforts, such as the Human Connectome Project [2] and the UK biobank study [3], including tens of thousands of participants, has facilitated the pooling of resources and increased the overall sample sizes in research studies. However, with the increase in public accessibility of online studies comes a potential increase in poor quality or false participants, ranging from careless respondents who respond with insufficient effort [4] to malicious responses such as automated attacks in the form of computer-programmed bots [5, 6].

Bots can be fully automated to target online research via automated malware or malicious algorithms [7], or they may be ‘hybrid’ where there is an element of human control. An example of a hybrid bot is when the human may complete the initial survey screening questionnaires and then allow an automated bot or algorithm to complete the remainder of the survey. Bot attacks have existed since the early 2000s and are becoming increasingly common. This increase may be linked to the rise of online paid research panels and crowdworker platforms such as Amazon Mechanical Turk, which bots can leverage. Whether the bots are automated malware or human respondents who are completing surveys for financial gain without meeting study inclusion criteria [8], it is irrefutable that caution must be exercised when conducting an online or remote study [9, 10].

Bot attacks must be identified as they impact the integrity of datasets [5, 11, 12], creating a situation where ineffective interventions may appear effective, and conversely, effective interventions may seem ineffective. Similarly, one of the significant risks of undetected bot attacks is the potential to misrepresent populations [5, 13] and influence decision-making based on erroneous data. Misrepresentation is particularly problematic for vulnerable [14] and at-risk populations such as indigenous communities, who are often poorly serviced by many existing interventions [15] or populations where new effective interventions are desperately needed [13, 16]. Thus, misrepresentations of bots as members of this population could have dire consequences.

This paper discusses strategies researchers can use to help reduce the impact of bots on research. Like several recent papers [5, 13, 17-19], the researchers were impacted by bot attacks on two independent projects hosted on different platforms at two institutions. We present case studies of the two research projects (see Multimedia Appendix I) that were affected by bot attacks and demonstrate the impact that these would have had on the demographics of the final dataset. We share the steps the research teams took to identify potential bot participants, outline strategies to mitigate the bots’ effects and reduce the chances of bot attacks in future projects, and provide recommendations for other researchers.

Ways to identify bot responses

Several signs in both case studies indicated likely bot attacks – both in terms of researcher review of the dataset and the survey metadata and through software flagging functions. We also attempted to develop an algorithm to predict the likelihood of the participant being a bot to help determine authentic participants from bots. These strategies are discussed further below.

Researcher strategies to identify bots

In both studies, recruitment appeared far more efficient than expected, particularly as there was no targeted advertising and the recruitment methods used were broad. Both studies had high numbers of participant registrations in a brief period, with registrations frequently being close together (often less than minutes apart).

Secondly, study participant demographics were disproportionate to the sample pool, and populations often hard to recruit were overrepresented. As this was more subtle than the high number of registrations, this may not have been detected until data analysis was underway if unusual recruitment patterns were not present. Overrepresentation was particularly obvious for CS2, where we had planned to purposefully recruit based on ethnicity and monitored to ensure we were successful in this recruitment but had yet to start targeted recruitment. It is noted that although the proportion remained high in the corrected dataset, this was overinflated due to the small sample size and likely would not have remained proportionally high once higher numbers or participants were recruited.

Thirdly, a review of participant email addresses indicated that most suspected bot attacks had emails that followed similar patterns (e.g., a combination of letters and numbers followed by the exact email domains) or were non-sensical. Whilst not all email users choose not to use their name in their emails, and the desire to protect against identity theft means that some people do not have emails that make sense, it was the similarity between the email addresses that highlighted the differences between suspected bots and the participants that we believed to be genuine.

Finally, unusual responses in the online surveys also highlighted potential bot responses. Specifically, random answering for survey questions as evidenced by inconsistent or incongruent responses, a lack of answers to qualitative or text responses, and patterned survey responses indicated likely bots.

Use of survey platform data

In terms of utilising the functionality of the survey platforms, several additional strategies can be employed to determine if a bot attack has occurred. Some survey platforms can flag surveys that are likely bots, so this should be utilised where possible. Although responses are flagged, researchers still need to review the responses as above to identify atypical patterns of responses. Whilst this requires researchers to review survey responses, the flagging system can potentially reduce the time required to identify bot attacks.

Potential statistical prediction to identify bots

For CS2, there were two parts to the study: 1) participants were invited to complete the ‘research’ questionnaire, which consisted of eight forms, and 2) to use the online intervention component of the study accessed through a second external website using their contact details. By cross-validating the information from the intervention platform with the initial research survey data (N=503), we obtained confirmation of true participants (n=27). However, we could not be sure whether the high number of remaining participants were bots or actual participants who did not continue to the next part of the study, although we assumed they were bots.

Yet this scenario presented us with a unique opportunity to attempt to create a bot detection algorithm to distinguish bots from genuine participants in CS2 better using strategies previously recommended in the literature [17, 20]. To achieve this, we assigned suspicion scores on a scale of 0-10 to responses that we evaluated as being potentially anomalous based on the following criteria:

- *Survey completion time:* We tracked the completion rates, completion times of individual forms within the survey and the time difference between initial and final form completion. A weak bimodal distribution appeared for completion times in CS2, allowing us to apply a higher suspicion score to completion times of less than five minutes.

- *Email address analysis:* We scrutinised patterns in email addresses, such as random strings of letters and digits, along with unusual domain names, to identify suspicious email patterns typically associated with bots. Higher suspicion scores were applied to unusual email address combinations.
- *Conflicting response analysis:* We evaluated responses to specific correlated survey questions to detect inconsistencies that might indicate automated responses. For instance, we sought contradictory answers to two questions concerning the frequency of experiencing ‘not being able to stop or control worrying’ and ‘worrying too much about different things’ over the past two weeks. Greater negative correlations between paired questions were assigned higher suspicion scores.

To construct the bot detection model, we utilized the XGBoost algorithm with a logistic regression base model for binary identification of bots and participants. XGBoost is a scalable tree-boosting system that has shown excellent performance in classification tasks due to its proficiency in managing high-dimensional data and capturing intricate patterns [21, 22]. We partitioned the dataset, allocating 80% for model training and 20% for testing model performance. Hyperparameters were fine-tuned through experimentation and cross-validation to attain optimal performance. We used receiver operating characteristic (ROC) curve results to optimize accuracy, sensitivity, and specificity by adjusting the classification threshold. Post-optimization, the model achieved an overall accuracy of 0.937, with a sensitivity of 0.967 (reflecting a high rate of bot detection), but a specificity of 0.400 (reflecting a poor rate of actual participant detection), for a classification threshold of 0.25. Hence, the optimised model was heavily biased towards detecting potential bots and unable to identify true participants consistently. The model’s balanced accuracy was moderately low at 0.683, reflecting the challenge of simultaneously identifying true positives and true negatives.

Based on this model, the most important features in order of predictive utility were the time difference between the first and last form completions within the baseline survey, the time difference between adjacent forms from “unique” participants, the total completion time of the overall survey, the email suspicion score, and conflicting responses for correlated questions. We also ran univariate logistic regressions for individual features to validate the XGBoost model results. Heavy zero-inflation (i.e., high numbers of participants with suspicion scores of 0) and poor discriminability were present in each feature, making it impossible to identify bots accurately even when features were combined.

To perform supervised learning this way, we assumed a binary classification of participants versus bots. In reality, the data was imperfectly labelled. Instead of labelling participants as ‘not bot’ or ‘bot’, the best we could achieve was labelling them as ‘not bot’ or ‘maybe bot’. Because the study comprised the online research questions and the online intervention, we used the login and intervention usage data to confirm some true participants (as the intervention website required interaction with diary components and activities, thus confirming a true participant). Nevertheless, there remained a high degree of uncertainty regarding the identity of the remaining participants who only completed the first baseline research questionnaire but did not proceed to log into the intervention. Thus, the bot detector's accuracy is probably exaggerated.

Ways to manage a bot attack

Researchers should regularly monitor recruitment and data collection to aid in the early identification of bot attacks. Close monitoring is critical soon after studies are advertised online as the risk of a bot attack, based on these two studies, seems to occur soon after study advertisement. Whilst frequent data monitoring will not stop a bot attack, it may reduce the breadth of the attack by being able to intervene early. Researchers can monitor data manually by regularly checking recruitment or by setting up alerts for new enrolments, which will signal if a high volume of enrolments occurs over a short time.

Close or pause the survey

Once a bot attack is detected, it is essential to close or pause the survey until appropriate steps can be taken. In both case studies, new surveys were created and circulated with additional security measures in place. Closing or pausing the survey ensures that no new bot enrolments can occur (although settings may allow incomplete surveys to be completed), and extra security measures can be implemented.

Seek institutional support

After identifying the bot attack, several actions were taken to help the research teams consider what to do next. Actions included notifying the respective ethics committees about the attack and seeking consultation and advice about responding. Despite this surge in internet use for health research, there remains limited guidance from health research ethics committees on managing bot attacks or ensuring the validity of online research studies. The emergence of large-scale digital research and artificial intelligence have presented unique challenges in research ethics and safety. AI and 'big data', particularly digital data, present unique ethical considerations, which institutional research ethics boards may not be adequately equipped to deal with [23]. For example, the issue of transparency is unique to contemporary AI models, as they are often considered 'black boxes' where the user is unaware of how the model has reached its conclusion [24]. Accountability remains a complex issue as it is unclear who should be held responsible when AI is involved [25]. Who is accountable if an algorithm provides a clinical recommendation based on digital data? The developer of the algorithm, the user of the algorithm, or a clinician who implements the clinical decision? In this context, we believe retaining autonomy and 'human control' of digital data is paramount. AI can also blur the concept of free and informed consent, and managing privacy becomes more challenging with AI's ability to identify individuals even after data deidentification [26]. Data bias is also a concern, as AI may not always detect or could potentially generate biased results, including harmful gender and racial biases [27]. It is crucial to continue addressing these ethical challenges to ensure responsible and safe implementation of AI technologies.

In both our case studies, several ethical dilemmas arose. First was the issue of reimbursement. As there is often a set budget for participant reimbursement, an ethical dilemma arises when the number of claims exceeds those budgeted. Specifically, in studies where actual enrolments exceed planned enrolments, there could potentially be hundreds of people who cannot be reimbursed if the study protocol outlines reimbursement of all respondents regardless of data quality. Second was the inability to differentiate between bots and actual participants in a way that could 100% confirm which responses were 'real' and which were not. Failure to accurately identify between participants and bots could lead to a risk of reimbursing bots and erroneously excluding reimbursement of actual respondents due to budgeting constraints.

Lastly, there were concerns about potential reputational risks for the researchers and their academic institutions if the research team reimbursed ‘fake’ responses or missed payment for ‘real’ participants. The respective university ethical and legal teams were consulted about researcher obligations to compensate based on completed answers, even if suspected bot attacks generated these. The standard advice was that the bot attack was a misrepresentative response and, therefore, there was no obligation to compensate, given that the terms of research participation had not been met.

In CS1, where participants were to be sent vouchers for compensation, careful screening of voucher claims revealed only four valid registrations (out of over 1200 claims), which were subsequently paid. Where participant contact details were available in CS2, we gained ethical approval to re-contact participants to indicate that there had been an attack on the study. We asked participants to complete their baseline data collection again to obtain their compensation. The re-completion rate was approximately 55%, and no complaints were received about this by the researchers or the ethics committee.

Remove data

Where there was a high probability of being a bot participant, this data was removed from the dataset, and a record was made of this decision. This decision was based on the above identification parameters. Where it was difficult to tell if the participant was a bot-participant or a legitimate participant, this was noted to allow further scrutiny during the longitudinal follow-up and for reporting in publications as appropriate. It is noted that in both case studies, two researchers were involved in each review of data and the decision to remove participants based on being a bot. It was also determined that it was important to do this before data analysis occurred, document all research decisions, and inform the respective ethics committees.

Ways to counteract bot attacks

Plan ahead

We believe that all researchers involved in online research should have a bot-management protocol in their data management plan. This could be based on protocols such as Ballard et al. (2019) and include the frequency of monitoring enrolments for bot attacks, steps to take when the survey is active if a bot attack is suspected, and a data handling and analysis plan for the study and that includes planning for bot-related and suspicious data. There should be a clear justification for removing data, and the dataset should be reviewed independently by two reviewers, including any decision made to remove data. Decisions should be documented, and publications related to the study should disclose the level of data removal related to the bot attack. Similarly, ethics committees who have approved the study should be consulted and informed about bot attacks, given the impact on research integrity.

Study advertisements

The researchers in this study and others [10, 17, 28]believe that studies that are advertised on social media, particularly media where bots are highly active and reach are widespread (such as X or Facebook), are more likely to experience bot attacks. The placement of study advertisements should be specific to the study population, where possible, and highlight the benefits of participating in the research other than compensation. Settings on survey platforms can also limit access to pages via routes other than where they were initially posted; for example, in CS1, settings were changed so that only participants who accessed the survey via the Facebook pages where the links were posted were included in the dataset. In addition, only those who accessed the voucher registration page via the survey could enter their details. Using previously verified distribution lists, such as those associated with professional or patient organisations, may reduce the likelihood of bot attacks as the advertising is more targeted.

Participant Information Sheets and Consent Forms

Participant information sheets and consent forms should highlight the steps researchers will take if they believe a participant has misrepresented themselves. Currently, participant information sheets often outline the inclusion and exclusion criteria and may ask that the participant indicate that they meet these criteria; however, there is no explicit statement about data management for suspected bot attacks. Upon reviewing the participant information sheet for CS2, the ethics committee felt that the wording and context did sufficiently imply that reimbursement was given only if the questionnaires were completed as part of the intervention study, so this provided adequate grounds for researchers to omit reimbursement to participants who did not complete the questionnaires as part of the fuller study. Nevertheless, an ethics amendment was made where the wording was changed to stipulate that reimbursement more clearly was only given if the questionnaires were completed *and* the study intervention was used as part of participation.

For future studies, we would recommend the inclusion of text such as “Any data that is thought to be generated by a bot or non-human means, is a duplication beyond what is required by the study, or is misrepresentative may be removed by the research team. If this occurs, study compensation for participation, as outlined in this participant information sheet, will not be offered. If you believe your data has been removed unfairly, you may discuss this with the lead researcher, who may ask you for proof of the legitimacy of your claim. If there are issues with data collection, you may be contacted to complete parts of the research again, which you may choose to decline”. Consent forms should also include reference to this, such as “I understand that if my data is believed to be misrepresented it may be removed from the study and I may not receive compensation for participation”. Before including this in external documents, we recommend consulting with your institutional legal team and ethics committee.

Compensation of participants

Offering financial compensation for research participation is a common strategy when engaging in health-related research, encouraging a breadth of responses [29] and helping to reduce the impact of self-selection bias [30]. Compensating participants is considered an ethical approach to ensuring widespread health research engagement and facilitates participation by financially disadvantaged participants [31, 32]. Compensation also recognises the time participants have contributed to the research project and the value of their contribution. However, compensation means that research can also attract participants who may not be interested in the outcome of the research and are participating for other reasons.

Advertising compensation in study advertisements may increase the likelihood of studies being targeted by bots [5, 6, 17]. Likewise, direct compensation for online data collection increases the risk of a bot attack, which may be mitigated by offering a prize draw instead. However, a prize draw may be less effective for recruitment [33], may not fairly compensate participants for their time, and may be less likely to recruit or may greater disadvantage more some individuals [34]. Alternatively, offering compensation contingent on completing specific tasks may reduce the likelihood of fraudulent participation, particularly if compensation payment is only made at the end of the study.

Ethics committees recommend offering compensation relative to the degree of involvement of the participant. For example, a five-minute survey may only be eligible for a prize draw, whereas a study with an in-person three-hour visit that requires blood tests and medical scans will warrant a higher compensation. However, there is limited guidance on what compensation is expected for different activities, and research teams arbitrarily determine compensation. Thus, studies where researchers offer a higher level of compensation for less effort (such as a short online survey), might be more appealing to bot attacks. Given this, we would advocate for more explicit guidance on research compensation, particularly for online studies.

Clarifying participant identity to claim compensation may also be another barrier for financially motivated bot attacks. Examples of this could be participants being asked to provide identification at the end of the survey to claim compensation and to ensure the participant is a person or the research teams sending vouchers to a postal address rather than an email address. Another potential way to determine authenticity may be using electronic bank transfers to a confirmed bank account rather than digital vouchers. Alternatively, cryptocurrency could also be effective, as the process of claiming blockchain-based rewards could involve complex cryptographic tasks or multi-factor authentication methods that are easy for humans but difficult for bots. Cryptocurrency transactions are also publicly verifiable by default on the blockchain, providing a transparent record of each payment (note that cryptocurrency has not yet achieved sufficiently widespread adoption to enable this method, and its use could potentially be seen as contentious.) These strategies could reduce the potential for bots or hybrid bot attacks to be paid and could prevent participants from claiming multiple compensations for multiple completions of the same survey.

Enhanced security settings

Researchers should familiarise themselves with the security options within their survey platforms to reduce bot attacks. Security options include activating settings to reduce multiple responses from the same participant, bot detection features, and CAPTCHA settings. For CAPTCHA, Qualtrics generates scores for each participant on a 0 to 1 scale based on several variables. Scores less than 0.5 are likely bots. It was noted that in the case studies above, this was not enough to deter bots but was a method of helping to identify potential bots. RelevantID settings also use code to determine if participants are taking the survey multiple times, allowing researchers to remove duplicate participants.

Crafty survey development

Two survey design steps may help make it more difficult for bots to access surveys. The first requires registration with an email address, and then potential participants are required to validate their email address to log on to the online survey. This registration requires more actions to log in, which makes it harder for less sophisticated bots to access the survey. However, as noted in CS2, bots or hybrid bots may be able to overcome this. Additionally, initial screening or inclusion criteria could be set up as individual questions that are randomized in order, and the correct answers need to be selected for inclusion in the final dataset. The answering patterns of a bot would likely mean that they would self-exclude from the final data set.

Surveys can also include questions that are only likely to be completed by bots, not participants. These are called “honey pot” questions, and each survey platform has a different way of creating them. For example, a field may be created that is invisible to humans, but can be detected by a bot. It is also noted that creating these questions in some coding software may be easier for a bot to detect, and recent online threads have recommended that JavaScript be the preferred coding platform for creating these fields.

The inclusion of text-based questions that need answering may also help with the identification of bot participants. Examples of this could include a series of questions requiring a typed answer about the ’country’s capital, what year it is, the city the person lives in, and basic mathematical equations such as “What is one plus three?”. Question format can also increase the chance of detecting a bot by directing participants only to select a specific number of items from a list where all items could be answered.

Kay and Saucier have created an online database of 660 items that can be used to detect careless or insufficient-effort responders in survey data. The authors propose that the best way to identify participants who respond in a misleading or invalid way is to add frequency and infrequency items.

Frequency items should be endorsed almost by every respondent (e.g. it should be illegal to kill an innocent person), and infrequency item statements should be endorsed by almost no respondents (e.g. when someone tells a funny joke, I feel angry) [35]. Thus, failure to endorse a frequency item or an infrequency item would likely indicate a bot.

The final recommendation is randomisation of survey items or questions that make it difficult for bots to "predict" the question and answer "appropriately", thus making it easier to identify non-sensical surveys. Similarly, including qualitative questions may also help to identify bots versus actual participants, as bots are less likely to complete these meaningfully or may replicate answers. Thus, reviewing qualitative answers for quality or content duplication may help identify a bot. Additionally, including qualitative answers may make bot responding more challenging and provide more insights for future research not captured in closed-text answers [14]. It is noted that these strategies individually will not necessarily prevent hybrid people-bot attacks but will increase the difficulty of survey completion. If multiple strategies are used, it will likely be more deterring for the bot and make it easier to identify bot-responding.

Meta-data

The use of IP address tracking may be helpful to identify actual participants compared to bots. Coupled with location filters, IP address tracking can help determine when multiple attempts are made of the same survey and when the participant is not within the geographical catchment of the survey. However, it is noted that virtual private networks (VPNs) can circumnavigate this, and IP address tracking alone is not foolproof. Some ethics committees consider IP addresses to be identifiable information that must be disclosed during the informed consent process if collected. Using this strategy should be approved by the overseeing ethics committee. For further details on utilising IP addresses to mitigate 'bots' effects, see White and Brodhead (2023) [8].

Challenges

There is likely no foolproof way to detect bots and determined bot users could overcome many of these strategies through a human answering some of the screening questions and then running the bot script. Thus, bot-attacks in online research are an ongoing problem to manage. There are also several ethical issues when identifying potential bots.

Decreased anonymity

Requiring participants to provide a form of identification prevents researchers from collecting anonymous responses. This may be particularly problematic for studies where participants may be reluctant to participate due to fear, stigma or where the behaviour being studied is illegal [36]. There are strategies to reduce the impact of the potential to be identified, such as the software randomly allocating codes that are not tied to answers or a researcher who does not have access to the survey data (or even the topic) being the contact person for voucher claims. Researchers working in areas where stigma and shame are common may be reluctant to design surveys where participants may need to be identified; however, there is research showing that larger incentives and greater privacy do not necessarily equate to disclosure of more sensitive data [37] and greater anonymity may impact the accuracy of online surveys [38]. Thus, protecting participant identity through de-identification may be the preferred route to prevent bots and increase survey accuracy.

Removal of data

Removing bot data is important for research integrity, but this should be done cautiously so as not to remove data belonging to valid participants. This is particularly valid when considering removing data with incomplete answers (particularly qualitative answers), as this may result in removing data from people who may struggle with literacy but also misrepresent attrition and engagement – both critical issues in digital interventions [39-41].

Data removal should be done by two independent, blinded researchers and disclosed in publications to ensure transparency and rigour and decrease biases. Similarly, sharing the processes around managing bot-suspected material will aid the research community in developing a consistent approach to bot-data management. Specifically, we believe that researchers should report the methods used to determine bot data, the percentage of data removed due to the data being bot-generated, and the percentage of any suspicious remaining data. As such, we need consistent community guidelines for handling and reporting bot-related data and research impact.

Abusive Responses/Researcher Protection

Soon after the early closure of one of the studies, the researchers began to receive requests for participation compensation. When “participants” were told that there would be no compensation due to suspected fraudulent behaviour, researchers began to receive abusive messages. Whilst this is often not at the level of abuse that has recently been disclosed with the rise of the internet [42, 43], receiving such abuse is a risk to the wellbeing of researchers [44, 45]. Given this, a plan is needed to determine if and how to respond to such communications and ensure the ’researcher’s wellbeing [44, 45].

Discussion

As technology evolves, so too will the sophistication of bot attacks. As such, bots are here to stay. Therefore, researchers will need multiple, regularly updated strategies to combat and manage ’bots’ effects, including institutional support and ethics committees knowledgeable in this space. It is likely that health research will need to consider multiple bot prevention strategies, use multimodal recruitment and data collection, and develop clear guidelines to help researchers manage bot-related data in online research.

Authentic participant data is valuable, and therefore, researchers need algorithms and a decisional process to detect bot data. A method that is overly sensitive to bots may remove actual participants unnecessarily, thus reducing the size of the data set and may result in the study being underpowered. On the other hand, an algorithm or decisional process with low specificity may allow an excessive number of bot data to be included, reducing data quality and increasing study costs through compensation paid to non-authentic participants. Further, bots are likely to provide either extremely noisy or biased answers that negatively impact data quality. When creating a bot detection algorithm, it is up to the researcher to decide whether they wish to prioritise sensitivity (the ability to detect participants) or specificity (the ability to detect bots).

When attempting to develop an appropriate algorithm to detect bots, we faced severe issues with zero-inflated data where most participants – whether bots or not – registered a suspicion score of 0 for multiple variables. This indicates a lack of information about the methods used to create the bot. It highlights the possibility that the attack was performed by human actors responding manually to each survey, making it seem more authentic. Yet even with perfect knowledge of our artificial bot dataset, any bot detection algorithm trained on this dataset would suffer from severe overfitting issues, as the attack strategies used will likely differ from those used for another project. Consequently, our detection algorithm probably wouldn’t generalize well to real-world attacks, except for similarly designed REDCap surveys. To make this even more challenging, each of the survey tools currently available to researchers (such as Qualtrics, REDCap, or SurveyMonkey) have different programming interfaces and database structures that would each require a unique version of the bot detection algorithm that is configured to their setup. Hence, we remain unaware of any tool that provides sufficient accuracy for bot detection across the wide range of currently used survey tools. Detection is made more difficult by imperfect participant labelling. For example, researchers can rarely be completely certain that an email address belongs to a bot, even in the presence of ‘obvious’ signs such as random strings of letters or digits. Genuine participants wishing to preserve their anonymity online may create a dummy email address for responding to surveys or use ‘hide my email address’ systems, which can also create unusual email handles. The best way to counter fake email addresses is by screening potential participants beforehand and sending a private, individual survey link to pre-screened addresses rather than making the link publicly available [17]. However, this may not be practical for all projects as it can be labour-intensive and as highlighted in CS2, is not foolproof.

Implications

Health research often engages vulnerable populations, and special care must be taken to protect privacy and ensure appropriate risk management is available where needed – including the detection of bots that may skew findings. One vulnerable group are the Indigenous populations. Health research must be responsive and capture the needs of indigenous populations who often experience poorer health outcomes. In CS1, bot responses identified as Māori (the indigenous population of New Zealand) made up 30% of the sample, and after bot measures were in place, the percentage identifying as Māori was 10%. In CS2, ‘bot’ responses inflated the participants that identified as Māori to almost four times what we expected in recruitment based on the population composition. Had we not identified the bot, we would have been quite confident in the analysis we made relating to Māori participants, given the high number, despite these results being unrelated to Māori at all. Therefore, there is a risk that failure to detect and adapt to bot technology could lead to misinformation and contribute to poorer outcomes for indigenous populations based on misinformed conclusions.

Given this, we encourage researchers to consider whether online surveys are the best way to obtain a representative sample with high data accuracy. Online recruitment may preclude some people, and although we increasingly see people being connected online, regular data connection, access to technology, and technology literacy may not yet be everyone’s privilege, including vulnerable communities. This is particularly pertinent for those where the digital world may be at odds with core and cultural values [46] or where surveys may not accurately capture experiences [47]. Instead, the use of multimodal surveys may reach a broader group, including those who may not be aware of online surveys, may be more responsive to the needs of communities, and may be able better to manage the impact of bots [14].

Recommendations

From our experiences and literature review, it would seem prudent to develop clear guidelines for conducting online research to reduce the risk of bot attacks and increase the robustness of online research. There are currently few guidelines that exist to guide practice in this area. The Association of Internet Researchers released their latest ethical guidelines for Internet research in 2019 [48]. The guidelines provide useful information about data management and security, and consider some of the critical issues that we discuss in our paper, including how to protect the researcher where the researcher’s public identity is known; specific ethical topics such as accountability, trust and transparency, which have different considerations for online research; and issues related to the accuracy of data including in-built biases from algorithms used for collection and how to use metadata. The EQUATOR network refers to three reporting guidelines relating to digital health research –the CHERRIES checklist – a checklist for reporting results of internet e-surveys [49]; the CONSORT-EHEALTH to standardise the reporting of evaluations of web and mobile interventions [50], and the iCHECK-DH which is a guideline and checklist on how to report digital health implementations [51, 52]. However, these EQUATOR guidelines refer more to the reporting of the online study than the conduct and do not refer to bot attacks or management [53]. We advocate for developing standardised guidance for researchers conducting online research that describes key considerations and a standardised approach to ensuring data accuracy, validity and protection against bot attacks or other avenues leading to misrepresenting the population of interest.

Future directions

Bot attacks will likely become more widespread and more difficult for researchers to detect as botware and attack algorithms become more sophisticated. The experiences from our research group and others [5, 13, 17-19] highlight a critical, growing problem that deserves more focused researcher attention. Nevertheless, online research has many advantages, such as the ability to reach large numbers of participants and the ability to complete the research remotely. Our experiences should not deter researchers from conducting online research studies. Instead, our paper is a call to action to raise awareness and encourage researchers to consider the risks and benefits of online research. We recommend developing guidelines around detecting and managing bot data in online surveys to help raise awareness of these issues, provide guidance around survey design and data management, and encourage transparency in reporting data that bots may have impacted.

Researchers should also consider the recent exponential development of large language models (LLMs) such as ChatGPT [54]. Unlike manually programmed bots, LLM-assisted bots can interpret and respond to surveys more coherently, making LLM responses more challenging to detect. This also allows attackers to automate responses to qualitative questions, a task previously reliant on human guidance. We believe it is only a matter of time before LLM-assisted bots become sophisticated enough to respond to any survey design. Consequently, it may become impossible to trust any results obtained from public survey links. Therefore, we recommend that researchers begin implementing more effective security strategies if they have not already done so.

A question that arises at this point, and a potential focus for further research, is whether bot attacks are generalisable across different parts of the world, as different countries have varying protocols for paying research participants [55]. Our experience also suggests that all current research platforms may be vulnerable to, illegal bot attacks, and the associated responsibility of software developers to ensure the security and privacy of people is paramount. Thus, bot attacks are likely to be a global issue.

Conclusion

Whilst online research studies increase the ease of participant recruitment and accessibility to a diverse range of respondents, the rise of sophisticated bot programmers and algorithms to automate survey responses risks invalidating online research. Careful planning of online research study designs and incorporating measures to minimise bot responses, such as using a mix of closed, open-ended, and randomised questions, is necessary to protect online studies from bot attacks. However, these measures should be weighed against the risk of inadvertently disqualifying or turning away real, genuine participants. There is an urgent need for standardised practices and guidelines to be developed to provide researchers with clear guidance on safeguarding against bot attacks and actions to take if a bot attack is suspected. As bot attacks are here to stay, this paper aims to raise researchers' awareness and create a call to action before the problem becomes more widespread and challenging to manage.

References

1. Mills, M.C. and C. Rahal, *A scientometric review of genome-wide association studies.* Communications Biology, 2019. **2**(1): p. 9.
2. Van Essen, D.C., et al., *The WU-Minn Human Connectome Project: an overview.* Neuroimage, 2013. **80**: p. 62-79.
3. Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.* PLoS Med, 2015. **12**(3): p. e1001779.
4. Ward, M. and A.W. Meade, *Dealing with careless responding in survey data: Prevention, identification, and recommended best practices.* Annual Review of Psychology, 2023. **74**: p. 577-596.
5. Griffin, M., et al., *Ensuring survey research data integrity in the era of internet bots.* Quality & Quantity, 2022. **56**(4): p. 2841-2852.
6. Lawlor, J., et al., *Suspicious and fraudulent online survey participation: Introducing the REAL framework.* Methodological Innovations, 2021. **14**(3): p. 20597991211050467.
7. Dittrich, D., F. Leder, and T. Werner. *A case study in ethical decision making regarding remote mitigation of botnets*. in *International Conference on Financial Cryptography and Data Security*. 2010. Springer.
8. White, A.N. and M.T. Brodhead, *Detecting Fraudulent Responses in Online Survey Research.* 2023.
9. Smyk, M., J. Tyrowicz, and L. Van der Velde, *A cautionary note on the reliability of the online survey data: The case of wage indicator.* Sociological Methods & Research, 2021. **50**(1): p. 429-464.
10. Pozzar, R., et al., *Threats of Bots and Other Bad Actors to Data Quality Following Research Participant Recruitment Through Social Media: Cross-Sectional Questionnaire.* J Med Internet Res, 2020. **22**(10): p. e23021.
11. Dupuis, M., E. Meier, and F. Cuneo, *Detecting computer-generated random responding in questionnaire-based data: A comparison of seven indices.* Behavior Research Methods, 2019. **51**(5): p. 2228-2237.
12. Ballard, A.M., T. Cardwell, and A.M. Young, *Fraud detection protocol for web-based research among men who have sex with men: development and descriptive evaluation.* JMIR public health and surveillance, 2019. **5**(1): p. e12344.
13. Bybee, S., et al., *Bots and nots: safeguarding online survey research with underrepresented and diverse populations.* Psychology & Sexuality, 2022. **13**(4): p. 901-911.
14. Yazici, E. and Y. Wang, *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, 2023: p. 1-6.
15. Barrett, N.M., et al., *Creating an environment to inform, build, and sustain a Māori health research workforce.* Journal of the Royal Society of New Zealand: p. 1-15.
16. Corsini-Munt, S., et al., *Vulvodynia: a consideration of clinical and methodological research challenges and recommended solutions.* Journal of Pain Research, 2017: p. 2425-2436.
17. Storozuk, A., et al., *Got bots? Practical recommendations to protect online survey data from bot attacks.* The Quantitative Methods for Psychology, 2020. **16**(5): p. 472-481.
18. Loebenberg, G., et al., *Bot or not? detecting and managing participant deception when conducting digital research remotely: case study of a randomized controlled trial.* Journal of Medical Internet Research, 2023. **25**: p. e46523.

19. Goodrich, B., et al., *Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys*. Applied Economic Perspectives and Policy, 2023. **45**(2): p. 762-784.
20. Xu, Y., et al., *Threats to online surveys: Recognizing, detecting, and preventing survey bots*. Social Work Research, 2022. **46**(4): p. 343-350.
21. Chen, T. and C. Guestrin. *Xgboost: A scalable tree boosting system*. in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.
22. Bentéjac, C., A. Csörgő, and G. Martínez-Muñoz, *A comparative analysis of gradient boosting algorithms*. Artificial Intelligence Review, 2021. **54**: p. 1937-1967.
23. Bouhouita-Guermech, S., P. Gogognon, and J.C. Bélisle-Pipon, *Specific challenges posed by artificial intelligence in research ethics*. Front Artif Intell, 2023. **6**: p. 1149082.
24. Brożek, B., et al., *The black box problem revisited. Real and imaginary challenges for automated legal decision making*. Artificial Intelligence and Law, 2023.
25. Sand, M., J.M. Durán, and K.R. Jongsma, *Responsibility beyond design: Physicians' requirements for ethical medical AI*. Bioethics, 2022. **36**(2): p. 162-169.
26. Erlich, Y., et al., *Identity inference of genomic data using long-range familial searches*. Science, 2018. **362**(6415): p. 690-694.
27. Jain, A., et al., *Awareness of Racial and Ethnic Bias and Potential Solutions to Address Bias With Use of Health Care Algorithms*. JAMA Health Forum, 2023. **4**(6): p. e231197.
28. Orabi, M., et al., *Detection of Bots in Social Media: A Systematic Review*. Information Processing & Management, 2020. **57**(4): p. 102250.
29. Abdelazeem, B., et al., *The effectiveness of incentives for research participation: A systematic review and meta-analysis of randomized controlled trials*. PLoS One, 2022. **17**(4): p. e0267534.
30. Kaźmierczak, I., et al., *Self-selection biases in psychological studies: Personality and affective disorders are prevalent among participants*. Plos one, 2023. **18**(3): p. e0281046.
31. Permuth-Wey, J. and A.R. Borenstein, *Financial remuneration for clinical and behavioral research participation: ethical and practical considerations*. Annals of epidemiology, 2009. **19**(4): p. 280-285.
32. Jennings, C.G., et al., *Does offering an incentive payment improve recruitment to clinical trials and increase the proportion of socially deprived and elderly participants?* Trials, 2015. **16**: p. 1-9.
33. Dykema, J., et al., *Guaranteed Incentives and Prize Drawings: Effects on Participation, Data Quality, and Costs in a Web Survey of College Students on Sensitive Topics*. Social Science Computer Review, 2023: p. 08944393231189853.
34. Bierer, B.E., et al., *Fair payment and just benefits to enhance diversity in clinical research*. J Clin Transl Sci, 2021. **5**(1): p. e159.
35. Kay, C.S. and G. Saucier, *The Comprehensive Infrequency/Frequency Item Repository (CIFR): An online database of items for detecting careless/insufficient-effort responders in survey data*. Personality and Individual Differences, 2023. **205**: p. 112073.
36. Ellard-Gray, A., et al., *Finding the hidden participant: Solutions for recruiting hidden, hard-to-reach, and vulnerable populations*. International journal of qualitative methods, 2015. **14**(5): p. 1609406915621420.
37. Murdoch, M., et al., *Impact of different privacy conditions and incentives on survey response rate, participant representativeness, and disclosure of sensitive information: a randomized controlled trial*. BMC medical research methodology, 2014. **14**: p. 1-11.
38. Lelkes, Y., et al., *Complete anonymity compromises the accuracy of self-reports*. Journal of

Experimental Social Psychology, 2012. **48**(6): p. 1291-1299.

39. Meyerowitz-Katz, G., et al., *Rates of Attrition and Dropout in App-Based Interventions for Chronic Disease: Systematic Review and Meta-Analysis.* J Med Internet Res, 2020. **22**(9): p. e20283.
40. Lipschitz, J.M., et al., *The engagement problem: A review of engagement with digital mental health interventions and recommendations for a path forward.* Current treatment options in psychiatry, 2023. **10**(3): p. 119-135.
41. Borghouts, J., et al., *Barriers to and facilitators of user engagement with digital mental health interventions: systematic review.* Journal of medical Internet research, 2021. **23**(3): p. e24387.
42. Doerfler, P., et al., *" I'm a Professor, which isn't usually a dangerous job": Internet-facilitated Harassment and Its Impact on Researchers.* Proceedings of the ACM on Human-Computer Interaction, 2021. **5**(CSCW2): p. 1-32.
43. Bisaillon, J., et al., *Cyberbullying of professors: what measures are in place in universities and what solutions are proposed by victims?* Studies in Higher Education, 2023. **48**(11): p. 1639-1650.
44. Mattheis, A.A. and A. Kingdon, *Does the institution have a plan for that? Researcher safety and the ethics of institutional responsibility.* Researching cybercrimes: Methodologies, ethics, and critical approaches, 2021: p. 457-472.
45. Gelms, B., *Social media research and the methodological problem of harassment: Foregrounding researcher safety.* Computers and Composition, 2021. **59**: p. 102626.
46. Donkin, L., et al., *An exploration of the goodness of fit of web-based tools for Māori: a qualitative study using interviews and focus groups.* JMIR Formative Research, 2024. **9**: p. e.
47. Holcombe-James, I. and E. Rennie, *Survey-based research in remote Indigenous communities: considerations for methods*, in *Field guide to intercultural research*. 2021, Edward Elgar Publishing. p. 54-67.
48. Heise, A.H.H., et al., *Internet Research: Ethical Guidelines 3.0*. 2019.
49. Eysenbach, G., *Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES)*. 2004, Gunther Eysenbach Centre for Global eHealth Innovation, Toronto, Canada. p. e34.
50. Eysenbach, G. and C.-E. Group, *CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions.* Journal of medical Internet research, 2011. **13**(4): p. e1923.
51. Perrin Franck, C., et al., *iCHECK-DH: guidelines and checklist for the reporting on digital health implementations*. Journal of Medical Internet Research, 2023. **25**: p. e46694.
52. Perrin Franck, C., et al., *Correction: iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations*. Journal of Medical Internet Research, 2023. **25**: p. e49027.
53. Altman, D.G., et al., *EQUATOR: reporting guidelines for health research.* The Lancet, 2008. **371**(9619): p. 1149-1150.
54. Kumar, S., et al. *A Comprehensive Review of the Latest Advancements in Large Generative AI Models*. in *International Conference on Advanced Communication and Intelligent Systems*. 2023. Springer.
55. Grady, C., et al., *An analysis of U.S. practices of paying research participants.* Contemporary Clinical Trials, 2005. **26**(3): p. 365-375.

Supplementary Files

Multimedia Appendixes

Case studies referred to in the manuscript show casing two bot attacks.
URL: http://asset.jmir.pub/assets/229df903b264f57e9f7db32cdf9a7daa.docx