JMIR Publications
The Leading eHealth Publisher

Issues with online studies, an institutional example of a widespread challenge

angela shone, Camilla Babbage, Katherine Bird, Lauren Marsh, Mirabel Pelton, Shireen Patel, Sarah Cassidy, Stefan Rennick-Egglestone

Submitted to: JMIR Mental Health
on: March 15, 2024

Preprint
JMIR Publications

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

Preprint
JMIR Publications

Issues with online studies, an institutional example of a widespread challenge

angela shone[1] PhD; Camilla Babbage[2]; Katherine Bird[2]; Lauren Marsh[2]; Mirabel Pelton[2]; Shireen Patel[2]; Sarah Cassidy[2]; Stefan Rennick-Egglestone[2]

[1]university of nottingham nottingham GB
[2]University of nottingham nottingham GB

Corresponding Author:
angela shone PhD
university of nottingham
east atrium Jubilee conference center
nottingham
GB

Abstract

This paper reports on the growing issues experienced when conducting internet-based research. Non-genuine participants, repeat responders, and misrepresentation are common issues in health research posing significant challenges to data integrity. A summary of existing data on the topic and the different impacts on studies is presented. Seven case studies experienced by different teams within our institutions are then reported, primarily focused on mental health research. Finally, strategies to combat these challenges are presented, including protocol development, transparent recruitment practices, and continuous data monitoring. These strategies and challenges impact the entire research cycle and need to be considered prior to, during and post data collection. With a lack of current clear guidelines on this topic, this report attempts to highlight considerations to be taken to minimise the impact of such challenges on researchers, studies and wider research. Researchers conducting online research must put mitigating strategies in place, and reporting on mitigation efforts should be mandatory in grant applications and publications to uphold the credibility of online research.

(JMIR Preprints 15/03/2024:58432)
DOI: https://doi.org/10.2196/preprints.58432

Preprint JMIR Publications

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Preprint
JMIR Publications

Issues with online studies, an institutional example of a widespread challenge

Blandine French[1-2]*, Camilla Babbage[2-3], Katherine Bird[1], Lauren Marsh[1], Shireen Patel[2-4], Mirabel Pelton[1-5], Sarah Cassidy[1], Stefan Rennick-Egglestone[6-7]

*Corresponding author

[1]School of Psychology, University of Nottingham, Nottingham, UK

[2]School of Medicine, Academic Unit of Mental Health & Clinical Neurosciences, Institute of Mental Health, University of Nottingham, Nottingham, UK

[3]NIHR MindTech MedTech Co-operative, Institute of Mental Health, School of Medicine, The University of Nottingham, Nottingham, UK

[4]NIHR Applied Research Collaboration East Midlands, University of Nottingham, Nottingham, UK

[5]Autism Research Centre, University of Cambridge

[6]School of Health Sciences, Institute of Mental Health, University of Nottingham, Nottingham, UK [primary affiliation]

[7]NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust, Nottingham, UK [secondary affiliation]

Blandine French; Blandine.french@nottingham.ac.uk; PhD; 0000-0002-9395-5919

Camilla Babbage; Camilla.babbage@nottingham.ac.uk; PhD; 0000-0002-6337-3352

Katherine Bird; katherine.bird@nottingham.ac.uk; MSc; 0000-0002-1749-1673

Lauren Marsh; lauren.marsh@nottingham.ac.uk; PhD; 0000-0001-6021-4383

Mirabel Pelton; mp2162@cam.ac.uk ; PhD; 0000-0002-3378-8944

Shireen Patel; shireen.patel@nottingham.ac.uk; MSc; 0000-0002-5537-0717

Sarah Cassidy; sarah.cassidy@nottingham.ac.uk; PhD;

Stefan Rennick-Egglestone; stefan.egglestone@nottingham.ac.uk; PhD; 0000-0003-4187-011X

Preprint JMIR Publications

Abstract

This paper reports on the growing issues experienced when conducting internet-based research. Non-genuine participants, repeat responders, and misrepresentation are common issues in health research

posing significant challenges to data integrity. A summary of existing data on the topic and the different impacts on studies is presented. Seven case studies experienced by different teams within our institutions are then reported, primarily focused on mental health research. Finally, strategies to combat these challenges are presented, including protocol development, transparent recruitment practices, and continuous data monitoring. These strategies and challenges impact the entire research cycle and need to be considered prior to, during and post data collection. With a lack of current clear guidelines on this topic, this report attempts to highlight considerations to be taken to minimise the impact of such challenges on researchers, studies and wider research. Researchers conducting online research must put mitigating strategies in place, and reporting on mitigation efforts should be mandatory in grant applications and publications to uphold the credibility of online research.

Keywords

Internet research, online studies, qualitative studies, surveys, mental health

Introduction

Internet-based research, such as online surveys or qualitative interviews via video conferencing platforms, has grown in popularity and usability in the last two decades but more specifically since 2020 and the COVID-19 global pandemic [1–3]. Online studies have enabled researchers to continue conducting research studies when in-person testing was not possible [4]. Online studies also facilitate the recruitment of large samples, specific samples or populations excluded from research (described as underserved communities, or previously “hard to reach [5]”) more easily [6]. It also offers many other benefits such as minimising recruitment costs [7] or allowing for anonymity, often favoured for specific populations [8]. While online anonymity is attractive for certain groups as the researcher doesn’t meet the participants, and may even allow participation from individuals who might otherwise feel excluded from research, this lack of in-person validation poses a distinct threat to validity, quality and integrity of the data [9].

Whilst internet-based research has been heavily utilised by researchers due to its many advantages, this increased use has also led to researchers questioning whether their online participant pool is genuine. Internet-based research describes “any research involving the remote acquisition of data from or about human participants using the internet and its associated technologies” [10] including both quantitative (e.g. surveys or questionnaires) and qualitative methodologies (e.g. focus groups, interviews). The following five categories of behaviour, by participants and others, can significantly impact data quality and integrity in online studies:

- Non-genuine participants - Participants faking their lived experience.
- Repeat responders - Participants taking part more than once.
- Misrepresentation – Exaggeration of specific details.
- Lack of engagement – Participants answering quickly, not paying attention and not reading the questions fully.
- Bots - Automated software application that performs repetitive tasks over a network.

These categories encompass different methodologies (qualitative and quantitative) but also different samples (general population and specific samples). While historically online studies were used by behavioural scientists (primarily aimed at the general population), health research (aimed at specific groups) has also increasingly transitioned online [11]. Behaviour science might encounter most categories presented above but primarily experience issues with repeat responders, lack of engagement and bots [12,13]. To remedy and minimise these issues many secure online behavioural platforms have been developed, such as Amazon Mechanical Turk (MTurk) [14] or Prolific Academic [15], however, their reliability is often contested [12,16]. These issues have a significant impact on data integrity with lack of engagement for instance as high as 40% [17].

While behavioural research often has limited inclusion criteria, health research tends to focus on recruiting groups with specific conditions or inclusion criteria. Of the above challenges faced by researchers when working with specific health participant groups, non-genuine participants, repeat responders and misrepresentation are more often experienced. Many studies indicate that a subset of participants in online surveys tend to falsely assert eligibility to obtain entry into the study, particularly when the incentive for participation is monetarised [18,19]. Hydock (2017) reported that a “small but nontrivial portion of participants in online survey studies misrepresented their identity for the chance of financial gain” [20].

All the categories above are problematic in online research; however, this publication focuses solely on health internet-based research and its most common barriers, namely: non-genuine participants, repeat responders and misrepresentation. Challenges and strategies to overcome these difficulties are presented, most relevant to researchers in this field of work, which requires necessary caution given the complex health conditions and patients being regarded.

The motivations for repeat responders or non-genuine participants are varied and while monetary incentive is often the main drive [20], participants can also be motivated by wanting to access a special intervention or treatment [21] or be politically motivated. Indeed, one study experienced purposeful interference and registration in their study from specific groups with vested interests [22]. Finally, repeat registration can also be due solely to human error, for example, not realising

they had already signed up, though insincere behaviours could be to receive more of the already mentioned motivations.

Many studies have reported the significant impacts of these issues on their studies. In a series of three studies, Chandler & Paolacci [19] found that between 3-40% of participants misrepresented themselves, for example saying they were residents of the United States when they were not. Bowen and colleagues [23] recorded the numbers of repeated responses (627/1900, 33%) and demonstrated the wide extent by classifying them into four categories, infrequent (2-5 responses from same person), persistent (6-10), repeater (11-30) and hackers (45-67). These issues have also been reported in qualitative online interviews [24–26], for example, Roehl and Harland [24] reported 5 out of 14 participants as inauthentic. Finally, many studies report great difficulties after advertising on social media. Salinas [27] found that only 50/512 (9.8%) respondents from Facebook adverts were genuine while Pozzar and colleagues [28] reported that after seven hours of advertising their study on Twitter, 94.5% were inauthentic, 16.2% of them showing evidence of bot automation.

The impact of these issues on data integrity is significant. Chandler and Sisso [29] demonstrated that inauthentic data has a significant impact on experiments – by increasing between group variance, on individual differences – by providing false information and on the association between variables – by suppressing, inflating or reversing correlation. For example, Ysindron and colleagues [16] found that non-genuine participants in a study on diabetes (49% of the sample) reported significantly poorer physical and mental health issues compared to the clinical group, suggesting substantial exaggeration of adverse health. This leads to incorrect conclusions, potentially creating inappropriate recommendations. The impact of these issues also greatly increases research costs both in terms of financial cost and time [30].

After having experienced many of these issues in their health research studies, the research team collated their individual experiences. Strategies and tips were exchanged, leading to institution-wide presentation and awareness to inform any future studies and initiating institutionally recognised strategies in place. This publication reports a sample of case studies experienced within our UK-based institution at the University of Nottingham and reflections on the methodological challenges and strategies extracted from the high-quality publications in the area.

Approach

Case studies

We present seven case studies experienced within our institution, all relating to mental health

research. This list is not exhaustive of all issues experienced (as it wouldn't be possible to summarise it all in one paper) in our institution but rather describes some examples of the breadth of this issue and how it impacted each research team. Table 1 represents the different case studies included in this manuscript.

Table 1- University of Nottingham case studies

Case study	**Methodology used**	**Category of issues experienced**	**Number/percentage of participants impacted**	**Impact on study**
1	Qualitative- Interviews	Non-genuine participants Repeat responders	8/12 were nongenuine data	Loss of funds, inability to finish project or publish, time-waisted by PI and researcher to identify data
2	Randomised controlled trial	Repeat responders	100 (8.9% of 1123) accounts were suspended due to repeat registration.	Human resources to develop and administer repeat registration protocol and suspend accounts suspected of repeat registration. Additional burden and distress for participants who had repeatedly registered in error.
3	Randomised Controlled trial	Non-genuine participants Repeat responders	Approximately 482 were suspected as non-genuine participants.	Impact on study researchers in terms of increased workload but it did not impact on recruitment figures.
4	Quantitative - survey	Non-genuine participants Repeat responders		The process of cleaning the data was difficult in decision-making, time-consuming and led to delays in completing the project. Inability to widely

				advertise afterwards.
5	Qualitative-interviews	Non-genuine participants	- Approx. 50 were suspected as nongenuine - No genuine participants were recruited - 7 were identified as nongenuine	Loss of time had to widen recruitment to the general population to deliver the project
6	Qualitative - Focus Groups	Non-genuine participants Repeat responders	Approx. 115 suspected non-genuine participants applied. 83% of selected participants were non-genuine.	Loss of funds, loss of time, loss of data integrity.
7	Quantitative-Administration of online task	Non-genuine participants	20 out of 31 non-autistic participants consented to the study were non-genuine	Loss of comparison group, unable to fully deliver funder expectations, time, stress

Case Study 1- Online interviews on CBT and ADHD

What happened?

The study aimed to establish the experience of Cognitive Behaviour Therapy (CBT) in adults with attention deficit hyperactive disorder (ADHD) through semi-structured interviews. The research study invited adults with ADHD from the PI's research database and received multiple and prompt expressions of interest. She interviewed 12 participants in a couple of weeks but after finishing the data collection, she felt that there was "something wrong" with some interviews.

How did you find out?

Out of 12 interviews, eight were suspicious. The interviewer felt that the responses were very brief and not in-depth, some stories were very similar if not identical and the participants followed a similar pattern. All eight had similar format of Gmail email addresses, refused to put their camera on,

related similar stories, had identical foreign accents and were actively asking about payments. Additionally, from our decade-long experience in this topic, we know that most adults with ADHD have very difficult experiences with CBT, and this group all reported very positive experiences of CBT which is very unusual. As only participants from the database had been invited, all participants (six) who were not on the database were excluded automatically after asking how they heard about the study. The remaining two narratives were too similar to some of the excluded ones and were therefore also excluded.

How did it impact your research (financial and time costs)?

The whole student project was compromised by this issue. As participants were paid £20 per interview, no extra budget was available to conduct additional interviews as well as no extra time within the constraint of the master's dissertation submission. The analysis had to be conducted on four interviews alone which was acceptable for the dissertation purposes (after explaining the issue to the head of teaching) but not for publication. The student was very disappointed by the amount of time that had been wasted on these and the impact it had on her opportunity to get her project published.

Case study 2 – Randomised Control Trial, NEON

What happened?

The Narrative Experiences Online (NEON) study conducted two definitive pragmatic online trials of a web-based digital health intervention, which provided access to a collection of mental health recovery narratives. To participate in the trial, UK residents completed an online informed consent form and registered an online account using a personal email address. Formal identity verification was not required by our regulator, and we chose not to use it, to avoid contributing to paranoia for a trial population who had a personal experience of psychosis (for the NEON Trial) [31]. We were prepared for the possibility of repeat registration of accounts due to published accounts of previous online trials [21], and so our protocol included an outline procedure to suspend accounts suspected of repeat registrations [31]. As recruitment to our trials progressed, we observed a range of instances of suspected repeat registrations. We decided to formalise our decision-making procedures on account suspension.

How did you find out?

The web application that delivered our trials included an administrative control panel providing access to information about all registered accounts. Through regular monitoring, we saw instances of sequential registrations, using email addresses that appeared related, with little time between each

registration, or from the same Internet Protocol (IP) address. We communicated with some participants and learnt that some instances were accidental repeat registrations due to confusion with trial procedures, and others were deliberate (including to obtain payment vouchers). Many participants did not respond to our messages. In some cases, there were legitimate reasons for these indicators. These included registration by more than one person from the same household, often on the same device, where trial information had been shared amongst the household.

How did it impact your research?

We spent a substantial effort on a protocol for decision-making about account suspensions (online supplement 1), which was necessary given ambiguous decision-making, e.g. where some participants did not respond to contact. Our protocol developed by the research team was amended and authorised by a Trial Management Group, and subsequently authorised by a Programme Steering Committee. We produced our protocol for the important purposes of 1) supporting trial integrity, and 2) enabling ethical conduct in communication with participants. Enacting the protocol required a substantial amount of time on behalf of administrators, who collected information on possible repeat registrations in "Repeat Registration Reports" (Online Supplement 2). It also required time from the decision-makers who reviewed those reports. For analytical integrity, further effort was spent on articulating a modified Intention To Treat (ITT) principle which excluded accounts suspended due to repeat registrations [32], and on developing a modified CONSORT diagram to report on suspensions [32,33].

Case Study 3 – Randomised Control Trial, Mindful Life-Well at Work

What happened?

The Mindful Life-Well at Work study is an RCT assessing whether Mindfulness-Based Cognitive Therapy- for Life (MBCT-L) is more effective than Stress Reduction Psychoeducation (SRP). Participants are recruited from healthcare, social care and teaching sectors. To participate in the trial, clinical staff who receive the study flyer circulated by their participating NHS trust sites complete an online eligibility screening and informed consent form. All forms and questionnaires are completed independently by participants using the online REDCap platform. During the second recruitment wave we received an influx of emails; 483 emails were received over five days. The emails were sent out at strange times (for example 3 am) and a large number of emails were received within a few minutes of each other. The number received was considerably larger than in earlier recruitment periods.

How did you find out?

Initially, the study researchers did not suspect any dubious activity. However, on the second day when a study researcher was recording the data for the non-eligible participants, they noticed that all inclusion/exclusion criteria had been selected for most cases. This was peculiar, as data from previous participants implied that usually one or two criteria were selected. Upon closer scrutiny, they noticed that the email account names did not match the names provided. The study team also observed that signatures on the consent forms indicated that the same person was completing them. The first batch of emails was sent out from Gmail and the second batch was from Hotmail. The data on REDCap suggested that many tabs had been opened at the same and forms completed one at a time – i.e. submitted each consent form approximately 5 seconds after each other.

How did it impact your research (financial and time costs)?

Despite the study team feeling quite convinced that these were not emails from actual participants, they investigated further to ensure that actual participants were not being missed. The senior database manager was consulted for advice. He informed the study team that as IP addresses were not collected because they are identifiable, all cases would need to be treated as actual participants until it was verified that they were not genuine. All those not deemed to be eligible were emailed and provided with alternative sources of support. All those who had consented were emailed to ask if they would be happy to be contacted when programme dates became available. Of the 483 emails sent out only one reply was received. In terms of resource implications, this required additional time from study researchers as they had to go through numerous emails, ascertain whether it may be a fake participant, record this on a separate log and reply to all emails. We were pleased that this had been identified early on before any non-genuine participants were randomised to the trial.

Case study 4 – Online survey to assess the acceptability of SPARX

What happened?

This project assessed whether any changes needed to be made to SPARX [34], a game developed in New Zealand that delivers cognitive behavioural therapy to young people with low mood and anxiety, ahead of a pilot and feasibility randomised controlled trial (RCT) in the UK [35]. To do this, a JISC online survey was promoted via relevant networks and organisations, including social media. Our target was to reach 100-200, 11–19-year-olds. The survey was designed so that consent (for those over 16) or consent (from a parent or legal carer) and assent (from those under 16) were initially obtained, which would direct young people to complete the study survey. After consent, on the final page of the survey, young people were directed to another online survey via a link where they could register for a prize draw. Questions were open-ended, asking the age of the responder,

what they liked about playing computer games (question 8) and what they thought the advantages of e-therapy were (question 14). In total 415 young people gave consent; 391 participants completed the online survey and 372 registered for the prize draw.

How did you find out?

The first suspicious activity noted was the speed of recruitment, receiving 299 responses to the survey in one day. On collating responses, it became clear that many participants (n=184, 47%) were not eligible (i.e. being outside of the age 11-19). Responses to question 8 were exact, case-sensitive duplicates of other responders, with 132 (63.8%) answers being replicas and not answering the question. From this, only 59 of 207 eligible (28.5%) responses were not noted as suspicious. A further 17 responses were noted as suspicious as when reviewing question 14, many responses seen in question 8 were also made here and continued to not make sense in response to the question asked. For example, the words ‘the Forest’ appeared throughout the survey responses. From 391 responses to the survey, we were left with 42 (10.7%) responses believed to be genuine.

How did it impact your research?

The process of cleaning the survey data was time-consuming and led to delays in completing the project. In addition, we had wanted to recruit many more young people but due to being unsure where the suspicious activity was coming from, we were reluctant to continue pushing for recruitment. Furthermore, given the subjective manner we had to employ to clean the data, it has been difficult to decide whether we should publish this data. It is also interesting to note the feelings I experienced as a researcher in losing trust in the data I collected. Perhaps most importantly, given the reliance on this research in informing the upcoming RCT, we had to be very cautious about how we applied this survey data to that study design.

Case study 5 – Online Interviews on self-harm in transgender people

What Happened?

We recruited transgender people for an online interview study regarding factors for self-harm. The study offered a £10 voucher for transgender views on an existing research tool and was advertised on social media (i.e., Reddit), alongside other recruitment pathways. Once the study was advertised online, we received over 40 responses expressing interest in 24 hours. The sheer number of respondents raised suspicions as respondent numbers far exceeded previous recruitment drives. At this stage, it was noted that all emails, without exception, were gmail.com accounts and followed the same pattern: first name, last name, then a series of numbers. Also, some participants used terminology which, in our experience, is not used by the transgender community (i.e., “I am a

transgender”). These all raised suspicions, but we did not want initial suspicions to result in genuine participants being missed so follow-up emails were sent with study and consent details to all.

How did we find out?

The nature of the study required potential participants to complete a well-being plan alongside standard informed consent. This entailed providing contact details of their GP and a trusted person, and the address from which they will take part. People who regularly take part in self-harm research are familiar with ethical requirements and complete necessary forms in an appropriate manner. However, here respondents either did not complete the well-being plan or omitted the necessary information. Some did complete wellbeing plans, but these raised further suspicions that these were fake participants as GP details included a mobile phone and a Gmail address, neither of which are typical for UK GPs. Because of this, we performed cursory Google searches of both the GP and home addresses provided in the well-being plan to ensure they were legitimate addresses. In all cases, it transpired that the addresses provided were either entirely false (i.e., did not exist at all), or were commercial addresses. In one instance, for example, the ‘home address’ was a hotel in central London, and another was an industrial property. While these addresses may be temporary residential properties or used during work hours to take part, however, this did not feel authentic. The false addresses, false or suspicious GP details, poorly completed well-being plans, suspicious terminology, and all respondents using Gmail accounts with similar email addresses left the researcher believing these respondents were inauthentic, and they were not recruited to the study.

How did it impact your research (financial and time costs)?

The process was disheartening, and the researcher spent hours responding to and identifying non-genuine responses, but early suspicions meant the researcher avoided wasting significant time and resources interviewing and having discarded data. However, no genuine participants were recruited for this study, and we had to widen the study to the wider general population.

Case study 6- Online Focus group on inclusion at work for autistic adults

What happened?

We advertised for autistic participants to take part in an online focus group. The advert was posted by a project partner on Twitter. Approximately 150 potential participants responded within 24 hours, which was more than anticipated and more than we could invite to take part. The advert was removed, and the research team were suspicious that there were some non-genuine respondents. To tackle this, we asked respondents for demographic and diagnostic information, and their motivation for being involved. We received 114 responses, the majority of which we believe were inauthentic.

We invited 6 respondents to the focus group who we thought were most likely genuine participants.

We ran the online focus group with 6 participants for one hour. All participants turned on their cameras and provided verbal answers to the discussion questions. However, responses from 5 out of the 6 participants were general and provided little depth or personal perspective. When prompted, they did not expand their points or give specific examples. All three researchers present strongly suspected that 5/6 participants did not have lived experience of autism although this was impossible to prove.

How did you find out?

The large volume of volunteers in the short time window was the first indication that the initial recruitment phase had attracted non-genuine participants. This was confirmed by the request for further information as many motivation statements were duplicated between respondents, or were superficial, textbook-like responses. Like previous case studies, most respondents had Gmail addresses in the form of first name, last name, and random number. Despite carefully selecting participants based on the quality of their responses and removing responses where inconsistent or suspicious information was provided, we were still not confident that our included participants were genuine.

How did it impact your research (financial and time costs)?

Participants received a £30 voucher, regardless of their contribution. We experienced an ethical dilemma about whether to include the data. Given we had no proof of inauthentic activity, we chose to analyse all the data, but also collect and include additional data to identify converging themes from a wider range of people and experiences. Aside from potentially compromising our data, this experience was financially costly (paying additional participants), time-consuming (dealing with many emails, filtering and identifying genuine responses, running and analysing data from inauthentic focus groups), and had an emotional impact on the researchers.

Case study 7 – Exploring patterns of self-harm in autistic and non-autistic adults.

What happened?

We recruited autistic and non-autistic adults for an online interview study comparing proximal and distal risk markers for self-harm between the two groups, using the Card Sort Task for Self-Harm (CaTS). Participants were offered a £10 voucher for participating. We advertised the study through

social media, charities, and volunteer research databases. Although many genuine participants signed up for our study, we became suspicious that a small group of participants recruited from one platform were not genuine participants.

How did we find out?

Our study required participants to have an initial meeting with the researcher to complete an online demographics questionnaire to check eligibility, create a well-being plan and obtain informed consent to take part. The well-being plan (identical to the one in case study 6) entailed providing multiple contact details and addresses. Several indicators raised our suspicions that some participants were not genuine. First, several indicators raised our suspicions that some participants were not interested in the research process with initial emails focussing on payment. Participants appeared to lack knowledge about the subject and could not provide details of their experiences after being probed. Second, there was evidence of participants attempting to register and take part in the study more than once such as refusing to turn their camera on. Despite this, it was clear to the researchers that the person had already taken part in the study. Further, it was also clear that identical demographic data had been entered more than once, provided neighbouring or fake addresses, fake phone numbers, and incomplete information (e.g. no details of healthcare provider or trusted person to contact). Together, these indicated a small group of participants who were attempting to register and take part in the study more than once, presumably for financial gain.

How did it impact your research (financial and time costs)?

The small group of participants who attempted to take part in the study more than once, and provided unreliable data, had to be excluded from the dataset. It took a significant amount of time to be confident that our study had been infiltrated by non-genuine participants. By the time we had identified an issue and taken action, we had run out of time and funding to complete data collection for the non-autistic group. This meant that we had to revise the aims of the research to explore patterns of self-harm in the autistic group only - we were not able to make comparisons with non-autistic adults as we had planned. We had to submit a report explaining this to the funder on completion of the study.

Methodological/Research challenges and strategies

To minimise the impact of issues with online studies (qualitative or quantitative), it is essential to have a systematic process for “determining the level of suspicion required to remove potentially unreliable data” [36]. This includes many strategies that have been summarised in publications

[4,26,36–38] such as a framework outlining the importance to Reflect, Expect, Analyse and Label [36].

To think about the different challenges, it is important to think about these issues in relation to the whole research cycle and ask different questions at different stages of the research process [25]:

- During funding application – What problems are likely to occur? What resources do I need to include to manage these problems?
- Before recruitment - what is likely to happen? Can I develop a protocol that thinks about as many potential issues as possible?
- During recruitment - can I verify that the participant met the inclusion criteria? How confident am I in this information?
- During data collection - was the participant hesitant or flustered when asked probing questions for additional detail? What were my first impressions of honesty in my reflexive journal? Did I note any nonverbal cues that might be a clue to participant dishonesty?
- During data analysis - did I find places where the participants contradicted themselves? Were a participant's answers detailed enough that the participant seemed knowledgeable about the topic?

There are many noteworthy strategies put forward in the current literature [4,26,36–38] and we bring together these challenges and strategies with our case studies, to support applications for how one might implement them within their online research. It is important to note that some of the strategies presented might not be accepted by certain ethics committees and/or lived-experience groups and liaising with them is essential before recruitment.

Pre data collection

Pre-data collection refers to two different stages of the research cycle, ethics application and recruitment.

Challenges:

The challenges experienced at the ethics application stage involve thinking in advance about what the potential threats to data integrity are. This will apply to threats throughout the research project and how the research team plans on dealing with these. In terms of recruitment, these threats could be multiple sign-ups from the same participants or non-genuine participants signing up (pretending to have specific health conditions or eligibility criteria). Some of the patterns that have been observed in detecting those include: a lot of entries very quickly, a lot of consecutive entries with the same email format (name.surname999@gmail.com), entries from countries outside the recruitment

area, similar IP addresses, refusal to provide a phone number, fast response to communication or inconsistencies.

Strategies:

We present below some of the strategies our research team has put in place. It is important to note that some strategies might not be accepted by certain ethic committees and/or lived-experience groups and liaising with them is essential prior to recruitment.

Some useful strategies that we have put in place and found useful include:

- Protocol. Creating a protocol before recruitment mitigating these issues as much as possible and where possible, co-produced with a lived-experience group to ensure the acceptability of the chosen strategies. To develop this protocol, it is useful to think about the following questions: How will you deal with inauthentic data? And how will I establish inauthentic data?
- Transparency. Include all steps to be taken for non-genuine/multiple registration in the participant information sheet.
- Social media. Close recruitment from particular social media platforms. Once your advert is on public platforms, you can no longer control how far and where it is distributed with many people sharing it, leading to non-genuine registrations. If advertising on social media, don't use terms such as "payment" or "gift cards" in adverts.
- Incentives. One strategy is to not offer any incentives which will not be attractive to any non-genuine participants. Alternatively, offering "in-kind" incentives that would only be attractive to your target group can be helpful. For example, in a sleep study, participants were offered a sleep training course upon completion.
- Technical considerations. Monitor responses from the same IP addresses. Installing cookies detector and CAPTCHA. Creating single-use links for each response. Having experts e.g., computer science, social media, and database experts as part of the team

During data collection

Challenges:

It is also important to check these issues once participants are recruited onto the study. Our team has experienced many challenges at this stage despite inputting strategies before data collection. Some of the challenges included people providing contradicting/ inconsistent responses (between eligibility criteria and consent form), providing similar stories/responses to the study multiple times, vague answers or those who cannot elaborate when prompted, and shorter than average time in responses.

In qualitative studies, refusing to put cameras on video calls, poor quality or technical problems. In quantitative studies, giving straight-line answers, high levels of non-response, answers that do not make sense and empty free boxes.

Strategies:

Following recruitment but before data collection, a few strategies can be put in place.

- Interviews. Asking participants to do a brief phone interview to check eligibility criteria before taking part. Providing a valid phone number as well as asking to briefly turn on the camera will often rebut non-genuine participants.
- Identity. For participants with specific conditions, insider knowledge is useful - asking them to describe their lived experiences could help identify non-genuine participants. Additionally, asking for a valid ID or documentation on the condition (for those who have received a diagnosis) can help minimise multiple and non-genuine responses.

It is also important to have strategies in place as data is collected.

- Data. Keep checking data as it comes through. Check for survey or interview duration, duplicate responses or look for inconsistencies in responses. Keep an eye out for responses that don't "feel right". Roll out recruitment gradually to have time to check and stop any potential issues with the data collection process.
- Implement data screening. Follow the data screening plan and the protocol on reimbursement and data inclusion.
- Reflexive notes. Keep track of your decision-making process and any challenges occurred at this stage for full transparency.

Post data collection

Challenges:

Once data collection is undertaken and multiple checks have been performed, it is also important to think about what will become of any data that you establish as non-genuine. For example, it is possible that participants sign up multiple times or do not complete certain questions, but these might be due to legitimate reasons. Issues around transparency in reporting the extent of data removed can arise in terms of how much to disclose in limited word count or potential concerns from reviewers.

Strategies:

- Checking. Contacting participants who responded multiple times or whose responses seem inconsistent to give them a chance to explain any personal difficulties with the study. Conducting regular debriefs with the research team to make decisions as a group and support

the researchers' wellbeing.

- Incentives. Aside from offering no incentives or in-kind incentives, check data before giving incentives, avoiding automated payments.
- Transparency. Notify ethics or funder of issues with non-genuine participants and their impact on the study. Be transparent in reports/papers, establishing the ratio of this impact is important and should be disclosed in further publications.

It is important to note that while these strategies are often very useful, they also often offer a lot of drawbacks in terms of inclusivity, accessibility, or engagement. There are no fool-proof strategies and researchers need to assess the pros and cons each provides.

Discussion

This publication reflects the significant impact and presence of issues with online studies by disclosing some of our experiences and the challenges faced with the strategies implemented. Primarily focused on mental health research, our experiences report a range of difficulties with repeated responders and non- genuine participants. Careful considerations and strategies are presented to help mitigate the threats these experiences can have on data integrity. These experiences and suggestions are given alongside those put forward in the literature, along the research cycle to give clear suggestions for consideration in future online research. We have shown that the issues around online data collection are broad and widely experienced, and our case studies reflect the increasing threat to data integrity. While we were not able to capture all the case studies in this publication, we know of at least twenty individual cases experienced.

It is important to note that a lot of the strategies presented offer drawbacks. For example, tracking IP addresses can be easily changed by the use of a VPN which prevents the researcher from tracking genuine country of origin, but also it is impossible to track through Gmail accounts [39]. Additionally, it restricts members of the same households from joining, potentially excluding genuine participants. Other strategies can also be perceived as unethical or counterproductive. For instance, not paying participants for their time and effort [40] or assessing eligibility either in person or on the phone [41] might not be acceptable for certain participant groups and does not capitalize on the full benefits and appeal of online research. Some strategies are also not friendly to all groups. For example, while some conditions can be easily "confirmed" with a diagnostic letter, many mental health conditions can be trickier to demonstrate or are stigmatised conditions that may prevent people from getting involved (e.g., anxiety, self-harm). Questioning the participants' lived experience through some form of justification might be extremely insensitive and unacceptable.

Many considerations must be taken when thinking about strategies through the research cycle. Some strategies can become quickly obsolete for example, as technology evolves, new ways of bypassing existing strategies are always emerging. For example, the recently created platform of ChatGTP can easily replicate a believable experience of living with certain conditions. Additionally, experienced users have been known to exchange tips on social media platforms on how to bypass study criteria and maximise rewards over engagement (ProlificAC on Reddit).

In terms of the research cycle, it is important to remember that it impacts all areas of the research cycle (from grant application to writing up) and to think about all the different threats to data integrity as early as possible. Most strategies also require resources (time and money). Hence, appropriate resources must be requested in grants, or provided directly by institutions from central funding. Institutions might consider putting in place structural support for researchers doing online studies, such as through continuous professional development of staff that support studies, e.g., research librarians, and research data managers.

Additionally, while many strategies can be implemented, most will need approval at the ethics application stage, and while these strategies aim to reduce non-genuine participation, it is not a guarantee and to date, we are aware of no studies that have reported an optimal protocol that addresses all these issues successfully. In a study on fraud detection, Ballard et al. [8] implemented nine different strategies to detect and deter non-genuine responses and their final results demonstrated that only 161/414 (38.9%) responses were genuine.

The impact that these issues create is not solely on the quality and integrity of the data but affects multiple levels of costs to a research project. In terms of financial cost, funded projects may waste funding on paying non-genuine participants. Timewise, hours and days are spent putting strategies into place and checking the data, which takes away essential time spent on other aspects of the project. It also has a cost in terms of mental wellbeing. As a researcher, spending time investigating sensitive healthcare topics to find out participants have misrepresented their experiences, invalidating others' experiences, makes you doubtful of the data and is potentially heartbreaking. Having to decide on whether to include or exclude data when lacking strict criteria is also very anxiety-inducing and difficult, especially in light of its potential impact on the study findings [38,42], an agreed decision-making protocol can help facilitate this process. It is also important to note that not only are the impacts experienced by the research team but there is also a significant impact on participants. With increased use of advisory groups and focus groups to identify key research topics, having even just one member in this group pretending to have a condition and listening to other's experiences can be very distressing and navigating the facilitation of these situations requires high-

level expertise. Especially if the extent of the fraud leads to non-publication of results and essentially what could be perceived as a “waste of time”.

Finally, as suggested by recent and insightful publications on the impact of these issues on specific groups, it is tricky but essential to strike a balance between data integrity and participants’ vulnerability [38]. It is essential to maintain trust with participants, especially from potentially vulnerable backgrounds in the case of health research. Therefore, exclusion should be dealt with very carefully as while it is important to ensure non-genuine participants’ results are not included in the analysis to minimise its impact on results, it is also essential to give genuine participants the benefit of the doubt and not exclude genuine mistakes or difficulties in communications.

In conclusion, many important considerations need to be given throughout the project to minimise, as much as possible, the impact of multiple responses, bots and non-genuine participants on our data. This is not easy and while many strategies exist and are useful, their efficacy highly depends on the project’s methodology and population of choice. Careful considerations need to be taken when implementing these strategies, ensuring that they are acceptable and feasible within the remits of each project. This is not a quick process and involves time and resources. However, these are essential in conducting online studies as without these checks, it is very unlikely that the data collected will be reliable and/or representative. Reporting of how these risks have been mitigated should become compulsory in upcoming grants and publications to ensure data integrity and credibility for online research.

Acknowledgements

Blandine French is supported by an ESRC-funded grant. Stefan Rennick-Egglestone was supported by the National Institute for Health and Care Research (NIHR), through the Narrative Experiences Online (NEON) programme (RP-PG-0615-20016) and the NIHR Nottingham Biomedical Research Centre (NIHR203310). KB is supported by a grant from Mental Health Research UK. CB acknowledges funding support from the NIHR MindTech Medtech Cooperative, the NIHR Nottingham Biomedical Research Centre, and the UK Research and Innovation (UKRI) Digital Youth Programme award which was part of the AHRC/ESRC/MRC Adolescence, Mental Health and the Developing Mind programme (MRC project reference: MR/W002450/1 and MR/T046864/1). The views expressed here are those of the authors and not necessarily those of the NIHR or the UK Department of Health and Social Care. Mirabel Pelton is supported by a grant from Autism Centre of Excellence at Cambridge (grant number 124306).

Preprint JMIR Publications

Conflict of interest

Dr French reports personal fees and nonfinancial support from Takeda and Medice. All other authors report no conflict of interest.

References

1. Hlatshwako TG, Shah SJ, Kosana P, Adebayo E, Hendriks J, Larsson EC, Hensel DJ, Erausquin JT, Marks M, Michielsen K, Saltis H, Francis JM, Wouters E, Tucker JD. Online health survey research during COVID-19. Lancet Digit Health Elsevier; 2021 Feb 1;3(2):e76–e77. PMID:33509387

2. De Man J, Campbell L, Tabana H, Wouters E. The pandemic of online research in times of COVID-19. BMJ Open 2021 Feb 23;11(2):e043866. PMID:33622948

3. Jones A, Caes L, Rugg T, Noel M, Bateman S, Jordan A. Challenging issues of integrity and identity of participants in non-synchronous online qualitative methods. Methods Psychol 2021 Dec 1;5:100072. doi: 10.1016/j.metip.2021.100072

4. Kramer J, Rubin A, Coster W, Helmuth E, Hermos J, Rosenbloom D, Moed R, Dooley M, Kao Y-C, Liljenquist K, Brief D, Enggasser J, Keane T, Roy M, Lachowicz M. Strategies to address participant misrepresentation for eligibility in Web-based research. Int J Methods Psychiatr Res 2014;23(1):120–129. doi: 10.1002/mpr.1415

5. Routen A, Bambra C, Willis A, Khunti K. Hard to reach? Language matters when describing populations underserved by health and social care research. Public Health 2022 Apr 1;205:e28–e29. doi: 10.1016/j.puhe.2022.02.006

6. Eysenbach G, Wyatt J. Using the Internet for Surveys and Health Research. J Med Internet Res 2002 Nov 22;4(2):e862. doi: 10.2196/jmir.4.2.e13

7. Lefever S, Dal M, Matthíasdóttir Á. Online data collection in academic research: advantages and limitations. Br J Educ Technol 2007;38(4):574–

582. doi: https://doi.org/10.1111/j.1467-8535.2006.00638.x

8. 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 Feb 4;5(1):e12344. doi: 10.2196/12344

9. Gagné N, Franzen L. How to Run Behavioural Experiments Online: Best Practice Suggestions for Cognitive Psychology and Neuroscience. Ubiquity Press; 2023 Jan 4;3(1):1. doi: 10.5334/spo.34

10. Ethics guidelines for internet-mediated research. BPS. Available from: https://www.bps.org.uk/guideline/ethics-guidelines-internet-mediated-research [accessed Feb 23, 2024]

11. Mortensen K, Hughes TL. Comparing Amazon's Mechanical Turk Platform to Conventional Data Collection Methods in the Health and Medical Research Literature. J Gen Intern Med 2018 Apr 1;33(4):533–538. doi: 10.1007/s11606-017-4246-0

12. Peer E, Rothschild D, Gordon A, Evernden Z, Damer E. Data quality of platforms and panels for online behavioral research. Behav Res Methods 2022;54(4):1643–1662. PMID:34590289

13. Yarrish C, Groshon L, Mitchell J, Appelbaum A, Klock S, Winternitz T, Friedman-Wheeler D. Finding the Signal in the Noise: Minimizing Responses From Bots and Inattentive Humans in Online Research. 2019 Oct 1;42:235.

14. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspect Psychol Sci SAGE Publications Inc; 2011 Jan 1;6(1):3–5. doi: 10.1177/1745691610393980

15. Prolific.ac—A subject pool for online experiments - ScienceDirect. Available from: https://www.sciencedirect.com/science/article/pii/S2214635017300989?via%3Dihub [accessed Jan 15, 2024]

16. Ysidron DW, France CR, Yang Y, Mischkowski D. Research participants recruited using online labor markets may feign medical conditions and overreport symptoms: Caveat emptor. J Psychosom Res 2022 Aug 1;159:110948. doi: 10.1016/j.jpsychores.2022.110948

17. Osborne JW, Blanchard MR. Random Responding from Participants is a Threat to the Validity of Social Science Research Results. Front Psychol 2011;1. Available from: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2010.00220 [accessed Mar 1, 2024]

18. Dennis SA, Goodson BM, Pearson CA. Online Worker Fraud and Evolving Threats to the Integrity of MTurk Data: A Discussion of Virtual Private Servers and the Limitations of IP-Based Screening Procedures. Behav Res Account 2020 Mar 1;32(1):119–134. doi: 10.2308/bria-18-044

19. Lie for a Dime: When Most Prescreening Responses Are Honest but Most Study Participants Are Impostors - Jesse J. Chandler, Gabriele Paolacci, 2017. Available from: https://journals.sagepub.com/doi/full/10.1177/1948550617698203 [accessed Jan 15, 2024]

20. Hydock C. Assessing and overcoming participant dishonesty in online data collection. Behav Res Methods 2018 Aug 1;50(4):1563–1567. doi: 10.3758/s13428-017-0984-5

21. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, Thompson SG, Wallace P. Methodological Challenges in Online Trials. J Med Internet Res 2009 Apr 3;11(1):e9. doi: 10.2196/jmir.1052

22. Faking participant identity: Vested interests and purposeful interference - Patricia Fronek, Lynne Briggs, 2018. Available from: https://journals.sagepub.com/doi/full/10.1177/1747016117740177 [accessed Jan 15, 2024]

23. Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying Multiple

Submissions in Internet Research: Preserving Data Integrity. AIDS Behav 2008 Nov 1;12(6):964–973. doi: 10.1007/s10461-007-9352-2

24. Imposter Participants: Overcoming Methodological Challenges Related to Balancing Participant Privacy with Data Quality When Using Online Recruitment and Data Collection - ProQuest. Available from: https://www.proquest.com/docview/2737499952?pq-origsite=gscholar&fromopenview=true&sourcetype=Scholarly%20Journals [accessed Jan 15, 2024]

25. Ridge D, Bullock L, Causer H, Fisher T, Hider S, Kingstone T, Gray L, Riley R, Smyth N, Silverwood V, Spiers J, Southam J. 'Imposter participants' in online qualitative research, a new and increasing threat to data integrity? Health Expect Int J Public Particip Health Care Health Policy 2023 Feb 16;26(3):941–944. PMID:36797816

26. Letter to the Editor: A possible threat to data integrity for online qualitative autism research - Elizabeth Pellicano, Dawn Adams, Laura Crane, Calliope Hollingue, Connie Allen, Katherine Almendinger, Monique Botha, Tori Haar, Steven K Kapp, Elizabeth Wheeley, 2023. Available from: https://journals.sagepub.com/doi/full/10.1177/13623613231174543 [accessed Jan 15, 2024]

27. Salinas MR. Are Your Participants Real? Dealing with Fraud in Recruiting Older Adults Online. West J Nurs Res SAGE Publications Inc; 2023 Jan 1;45(1):93–99. doi: 10.1177/01939459221098468

28. Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsky JA, Hong F, Gundersen DA, Berry DL. 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 Oct 7;22(10):e23021. doi: 10.2196/23021

29. Chandler J, Sisso I, Shapiro D. Participant Carelessness and Fraud: Consequences for Clinical Research and Potential Solutions. J Abnorm

Psychol 2020 Jan 1;129:49–55. doi: 10.1037/abn0000479

30. Strickland JC, Stoops WW. Utilizing content-knowledge questionnaires to assess study eligibility and detect deceptive responding. Am J Drug Alcohol Abuse 2020 Mar 3;46(2):149–157. doi: 10.1080/00952990.2019.1689990

31. Rennick-Egglestone S, Elliott R, Smuk M, Robinson C, Bailey S, Smith R, Keppens J, Hussain H, Pollock K, Cuijpers P, Llewellyn-Beardsley J, Ng F, Yeo C, Roe J, Hui A, van der Krieke L, Walcott R, Slade M. Impact of receiving recorded mental health recovery narratives on quality of life in people experiencing psychosis, people experiencing other mental health problems and for informal carers: Narrative Experiences Online (NEON) study protocol for three randomised controlled trials. Trials 2020 Jul 20;21(1):661. doi: 10.1186/s13063-020-04428-6

32. Robinson C, Newby C, Rennick-Egglestone S, Llewellyn-Beardsley J, Ng F, Elliott RA, Slade M. Statistical analysis plans for two randomised controlled trials of the Narrative Experiences Online (NEON) Intervention: impact of receiving recorded mental health recovery narratives on quality of life in people experiencing psychosis (NEON) and people experiencing non-psychosis mental health problems (NEON-O). Trials 2023 May 20;24(1):343. doi: 10.1186/s13063-023-07246-8

33. Slade M, Rennick-Egglestone S, Elliott RA, Newby C, Robinson C, Gavan SP, Paterson L, Ali Y, Yeo C, Glover T, Pollock K, Callard F, Priebe S, Thornicroft G, Repper J, Keppens J, Smuk M, Franklin D, Walcott R, Harrison J, Smith R, Robotham D, Bradstreet S, Gillard S, Cuijpers P, Farkas M, Zeev DB, Davidson L, Kotera Y, Roe J, Ng F, Llewellyn-Beardsley J, Group the N study. Effectiveness and cost-effectiveness of online recorded recovery narratives in improving quality of life for people with non-psychotic mental health problems: a pragmatic randomized controlled trial. World Psychiatry 2024;23(1):101–112. doi: 10.1002/wps.21176

34. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for

adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ British Medical Journal Publishing Group; 2012 Apr 19;344:e2598. PMID:22517917

35. SPARX-UK: pilot and feasibility trial. Health Res Auth. Available from: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/sparx-uk-pilot-and-feasibility-trial/ [accessed Feb 23, 2024]

36. Lawlor J, Thomas C, Guhin AT, Kenyon K, Lerner MD, Drahota A. Suspicious and fraudulent online survey participation: Introducing the REAL framework. Methodol Innov SAGE Publications Ltd; 2021 Sep 1;14(3):20597991211050467. doi: 10.1177/20597991211050467

37. Teitcher JEF, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, Preventing, and Responding to “Fraudsters” in Internet Research: Ethics and Tradeoffs. J Law Med Ethics 2015;43(1):116–133. doi: 10.1111/jlme.12200

38. EXPAND - pullen-sansfacon-et-al-2024-dealing-with-scam-in-online-qualitative-research-strategies-and-ethical-considerations.pdf. Available from: https://uniofnottm.sharepoint.com/sites/EXPAND/Shared%20Documents/Forms/AllItems.aspx?id=%2Fsites%2FEXPAND%2FShared%20Documents%2FGeneral%2FEXPAND%20%2D%20project%20folder%2F11%2E%20Literature%2Fpullen%2Dsansfacon%2Det%2Dal%2D2024%2Ddealing%2Dwith%2Dscam%2Din%2Donline%2Dqualitative%2Dresearch%2Dstrategies%2Dand%2Dethical%2Dconsiderations%2Epdf&parent=%2Fsites%2FEXPAND%2FShared%20Documents%2FGeneral%2FEXPAND%20%2D%20project%20folder%2F11%2E%20Literature&p=true&ct=1707400852230&or=Outlook%2DBody&cid=88D25DE6%2DE235%2D47FA%2DA890%2D012D16E6FF2A&ga=1 [accessed Feb 8, 2024]

39. Liu E, Sun L, Bellon A, Voelker GM, Savage S, Munyaka INS. Understanding the Viability of Gmail’s Origin Indicator for Identifying the Sender.

40. Wilson PM, Petticrew M, Calnan M, Nazareth I. Effects of a Financial Incentive on Health Researchers' Response to an Online Survey: a Randomized Controlled Trial. J Med Internet Res 2010 May 10;12(2):e1251. doi: 10.2196/jmir.1251

41. Bromberg J, Wood ME, Black RA, Surette DA, Zacharoff KL, Chiauzzi EJ. A Randomized Trial of a Web-Based Intervention to Improve Migraine Self-Management and Coping. Headache J Head Face Pain 2012;52(2):244–261. doi: 10.1111/j.1526-4610.2011.02031.x

42. Data Quality in HIV/AIDS Web-Based Surveys. doi: 10.1177/1525822X12443097

Preprint JMIR Publications