News and Perspectives

Participant Fraud: How Imposters and Bots Undermine Health Databases

Cliff Dominy, JMIR Correspondent

Abstract

Advances in technology are making scientific fraud easier, faster, and harder to detect—from fake papers and fake authors to fake participants. In this *News and Perspectives* article, JMIR Correspondent Cliff Dominy reports on the evolving problem of participant fraud, as well as measures that can be taken to detect and prevent it.

Key Takeaways:

- Increasingly sophisticated AI-driven technologies have contaminated datasets with fraudulent responses, forcing researchers to spend time and money protecting the integrity of their data.
- Addressing participant fraud will require a multipronged approach including both technology and human expert review.

These days, we are often asked to prove our humanity by, say, counting fire hydrants in a graphical screen known as CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart), designed to block malicious bots from infiltrating websites. Despite initial success, the introduction of optical character recognition capabilities has allowed modern bots to evade CAPTCHA detection, necessitating the development of updated approaches such as iReCAPTCHA to outcompete the new threats. For health researchers collecting data via web-based surveys, CAPTCHA failure can be devastating.

The Long COVID Episodic Disability Study

The extent of the AI-infiltration problem was highlighted in a recent paper by Kelly O'Brien, PhD, and colleagues from the Department of Physical Therapy, Temerty Faculty of Medicine at the University of Toronto. The Long COVID and Episodic Disability Study was a web-based survey involving two self-reported health questionnaires administered electronically, one week apart, to assess the symptoms and health challenges of people living with long COVID in Canada, Ireland, the United Kingdom, and United States.

Recruitment was carried out through trusted community organizations, many of which advertised the study, along with its CAD $40 gift card incentive, on their social media channels. These marketing efforts, intended to lower the participation barrier for people with long COVID, ultimately exposed the study to fraud.

Within 24 hours of launch, the survey received 3638 replies; a response that wasn't just good—it was too good. O'Brien was suspicious. "We knew right away that the link we were using had been compromised," she said, continuing, "We really had to shut everything down, and basically we were left with a wealth of data to [manually] sift through to delineate real participant responses from garbage." The result: just 9% of responses came from legitimate participants living with long COVID.

O'Brien's team discovered that the fraudulent responses had left a characteristic trail of clues, including respondents with similar internet addresses or using generic emails that differed by just a single character. Further inspection revealed discrepancies in geolocation, with responses from computers outside the study countries. Another red flag was survey speed—bots are simply faster than humans. The team began manually sifting through the questionnaire responses to separate genuine participant responses from fraudulent responses, which took time and wasted precious research dollars.

The team changed tactics for the relaunch of the survey a few weeks later. This time, fresh weblinks were in place, there was no mention of a gift card incentive, the link was not made openly accessible on social media, and recruitment was contained within networks of community leaders on the team. For successful infiltrators, they added a last trap— honeypot questions, invisible to humans but detectable by bots—to the new survey.

The results were better; 482 (47%) of the 1025 respondents made it past the prescreening procedure, with 377 (69%) legitimate participants completing the survey questionnaire and subsequently being compensated for their participation. In trimming the responses, the team prioritized data validity at the risk of excluding genuine participants—the price of securing trustworthy data that accurately represented the experiences of persons living with long COVID. It was a tough call, says O'Brien, noting "we really erred on the side of being able to ensure that we were retaining responses that

were true... but in so doing, we most likely removed real participant data."

The Problem: Technology

The arrival of the internet opened up new opportunities for these medical fraudsters, with O'Brien recalling, "One of my first experiences conducting a web-based survey was in a study called the HIV Health and Rehabilitation Survey," a 2014 national study in which multiple fraudulent responses were received from a small group of internet (IP) addresses. Today, AI has enabled the mass harvesting of online rewards —a lucrative activity for the ethically challenged. Looking back, she reflects, "I feel like over a decade later, we're still experiencing the same issues, but with much more sophistication."

The unwelcome reality is that the role of health researchers has expanded from improving patient care to playing cybercop. "It's really disrupting and detrimental to science... we often talk about protecting participants, protecting individuals," O'Brien notes, continuing, "I think in this case, we now have to protect the science and the validity of the data itself." Another unrecognized personal cost: fraud fatigue, the emotional consequence of spending time and energy countering malfeasance instead of advancing medical research.

> *I feel like over a decade later, we're still experiencing the same issues, but with much more sophistication.*
>
> Kelly O'Brien, PhD

The Solution: More Technology?

As is often the case with the abuse of technology, better technology is needed to compete in the ongoing arms race against the imposters. Multipronged countermeasures will be needed, involving at least partially automated surveillance to reduce the scale of the problem.

Researchers at Johns Hopkins in Baltimore have provided recommendations for tackling the participation fraud problem, including:

- Website security: outdated CAPTCHA screens need augmentation with newer technology (like iReCAPTCHA) to effectively repel bot attacks.
- Survey questions: honeypot and corroboration questions between different sections of the survey to look for inconsistencies in the answers.
- Web monitoring: ongoing scanning to identify multiple respondents and geolocation discrepancies.

Bots and humans behave differently when interacting with websites, producing measurably different text-entry signatures; behavioral analytics could be informative during the survey phase. Biometric services, for example, can monitor mouse track timing and typing rhythm to help detect and deny bots.

Another clue can be found in the new science of bot fingerprinting. Bot technology creates characteristic operating system signatures not typically used by individuals. For example, bots use scripts and cookies to launch automatic responses on multiple accounts using differing internet addresses. That activity can be tracked and added to the *bot index*, the sum total of all the evidence. It's no longer what you say on the questionnaire but rather how it's said that could reveal your true identity.

Finding Balance in a Brave New World

While many of these approaches can be automated to ease the administrative burden on health researchers, full automation may not be desirable.

Researchers at UC Davis analyzed over 31 fraud detection strategies to determine which combination of strategies optimized for dataset integrity with minimum participant loss. To keep pace with increasingly sophisticated attacks, the group recommended a layered strategy combining several complementary detection technologies in combination with expert human review.

Participant fraud is a significant and growing threat to data integrity in health research. Both automated technology and human oversight will be necessary to safeguard data quality and address that threat.

Keywords: fraud; artificial intelligence; data integrity; research ethics; internet; computer security; bots

Please cite as:
Dominy C
Participant Fraud: How Imposters and Bots Undermine Health Databases
J Med Internet Res 2026;28:e112019
URL: https://www.jmir.org/2026/1/e112019

doi: 10.2196/112019