Identifying and Analyzing Bot-Generated Responses in Healthcare Research

Emily Hamovitch, Kaileah McKellar, Walter P Wodchis

Submitted to: Journal of Medical Internet Research
on: March 07, 2025

Table of Contents

Identifying and Analyzing Bot-Generated Responses in Healthcare Research

Emily Hamovitch[1] MSW, MPH; Kaileah McKellar[1] MPH, PhD; Walter P Wodchis[1] MA, PhD

[1] University of Toronto Toronto CA

Corresponding Author:
Emily Hamovitch MSW, MPH

University of Toronto
155 College Street
Toronto
CA

Abstract

Background: The increasing reliance on online surveys for collecting patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) has led to growing concerns over fraudulent responses generated by bots. These automated responses threaten data integrity by fabricating survey results, distorting statistical analyses, and potentially misguiding policy decisions. Addressing this issue is critical for maintaining the validity of research findings that inform healthcare practice and policy.

Objective: This study aimed to develop a robust set of criteria for identifying bot-generated responses in online healthcare surveys and to examine how these responses impact data quality. We then explored differences in survey results between human and bot respondents in a survey assessing PROMs and PREMs within a geographic region in Ontario, Canada.

Methods: A survey was conducted from July to October 2023 using REDCap, distributed with a generic link via email and later shared on social media. The survey collected data on healthcare use, patient experiences, health outcomes, digital healthcare engagement, and demographics. A three-tier classification system was developed to detect bot responses based on predefined “red flags,” including duplicate open-ended responses, inconsistent demographic reporting, identical timestamps, and location discrepancies. Quantitative analysis included chi-square tests to assess differences between human and bot responses and Spearman’s correlation tests to examine relationships amongst healthcare indicators.

Results: Analysis included 1,154 responses, with 58% (n=668) classified as bot-generated. The most frequent bot-identification criterion was duplicated open-ended responses (n=293). Chi-square tests revealed statistically significant differences (P<.001) between bots and humans across nearly all survey items. Bots demonstrated response patterns concentrated in the middle of Likert scales, whereas humans were more likely to select extreme values. Correlation analyses showed that expected relationships between key health indicators (e.g., depression symptoms) were present in human responses but reversed in bot-generated data, highlighting the potential for compromised validity in unfiltered survey datasets.

Conclusions: The findings underscore the necessity of implementing bot prevention and detection methods in online healthcare surveys to preserve data integrity. Failure to do so risks distorting research conclusions, particularly in health equity studies where demographic misclassification may bias results. The study highlights effective bot detection strategies, including open-text analysis, timestamp evaluation, and geographic validation, and recommends integrating these techniques into survey design. As bots continue to evolve, ongoing advancements in bot prevention and detection will be crucial to maintaining the reliability of digital health research.

(JMIR Preprints 07/03/2025:73622)
DOI: https://doi.org/10.2196/preprints.73622

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
No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Identifying and Analyzing Bot-Generated Responses in Healthcare Research

Abstract

Background: The increasing reliance on online surveys for collecting patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) has led to growing concerns over fraudulent responses generated by bots. These automated responses threaten data integrity by fabricating survey results, distorting statistical analyses, and potentially misguiding policy decisions. Addressing this issue is critical for maintaining the validity of research findings that inform healthcare practice and policy.

Objective: This study aimed to develop a robust set of criteria for identifying bot-generated responses in online healthcare surveys and to examine how these responses impact data quality. We then explored differences in survey results between human and bot respondents in a survey assessing PROMs and PREMs within a geographic region in Ontario, Canada.

Methods: A survey was conducted from July to October 2023 using REDCap, distributed with a generic link via email and later shared on social media. The survey collected data on healthcare use, patient experiences, health outcomes, digital healthcare engagement, and demographics. A three-tier classification system was developed to detect bot responses based on predefined “red flags,” including duplicate open-ended responses, inconsistent demographic reporting, identical timestamps, and location discrepancies. Quantitative analysis included chi-square tests to assess differences between human and bot responses and Spearman’s correlation tests to examine relationships amongst healthcare indicators.

Results: Analysis included 1,154 responses, with 58% (n=668) classified as bot-generated. The most frequent bot-identification criterion was duplicated open-ended responses (n=293). Chi-square tests revealed statistically significant differences (P<.001) between bots and humans across nearly all survey items. Bots demonstrated response patterns concentrated in the middle of Likert scales, whereas humans were more likely to select extreme values. Correlation analyses showed that expected relationships between key health indicators (e.g., depression symptoms) were present in human responses but reversed in bot-generated data, highlighting the potential for compromised validity in unfiltered survey datasets.

Conclusions: The findings underscore the necessity of implementing bot prevention and detection methods in online healthcare surveys to preserve data integrity. Failure to do so risks distorting research conclusions, particularly in health equity studies where demographic misclassification may bias results. The study highlights effective bot detection strategies, including open-text analysis, timestamp evaluation, and geographic validation, and recommends integrating these techniques into survey design. As bots continue to evolve, ongoing advancements in bot prevention and detection will be crucial to maintaining the reliability of digital health research.

Keywords: Bots; Online surveys; Data integrity; Patient-reported outcome measures; Patient-reported experience measures; Healthcare research; Digital health; Survey validity

Introduction

Patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) are invaluable tools for understanding and improving healthcare from the patient's

perspective. PROMs assess health status, functional abilities, symptoms, and quality of life directly from patients, often through self-reported questionnaires, providing insights into the impact of interventions and overall care quality.[1-3] PREMs, on the other hand, capture patients' perceptions of their healthcare experience, focusing on whether key processes occurred and how care was delivered.[2] These measures are essential for implementing person-centered care, as they enable healthcare providers to align their practices with patient needs and experiences.[4] PROMs and PREMs have also been incorporated into care routines to support quality improvement initiatives.[4] By prioritizing patient-reported data, healthcare systems can address inequities, promote patient-centred and value-based care, and ensure that care delivery aligns with the priorities and needs of diverse patient populations.[4]

A common approach to collecting PROMs and PREMs is through online surveys. Online surveys have become an increasingly prominent tool for research, particularly in social and behavioral health fields, due to their efficiency, cost-effectiveness, and ability to overcome challenges faced by traditional data collection methods such as face-to-face, mail, or telephone surveys.[56] These challenges include declining response rates, rising costs, and the need for social distancing during events like the COVID-19 pandemic.[5,7,8,9,10] Internet-based platforms such as Qualtrics, SurveyMonkey, and REDCap have enabled researchers to collect data remotely, making it possible to reach large general populations.[5,7] These tools reduce the resources required for data collection, minimize participant burden, and offer greater anonymity, encouraging honest disclosure of personal or sensitive information.[7,10] Additionally, the limited availability of a sampling frame with accurate contact information has necessitated the use of distribution methods involving generic survey links and promotion through online platforms and social media. Unfortunately, recruiting participants through social media creates a susceptibility to data quality issues.[11]

While early guidance on online surveys for health research warned of the risk of hoax responses,[12] the increasing use of online surveys has been accompanied by a growing concern over fraudulent responses, particularly from bots - applications that automatically complete surveys – producing insincere data and inaccurate results.[13] These bots are capable of mimicking human behavior by selecting multiple-choice answers and generating plausible open-ended responses, making them difficult to detect. Bots can submit hundreds of fabricated responses within hours, potentially overwhelming researchers with invalid data and causing significant financial loss if participant incentives are misallocated.[8] Their presence threatens data quality, which can distort statistical analyses and compromise research findings.[10] Researchers, evaluators, and program analysts also face challenges in identifying fraudulent responses due to the increasing sophistication of bots and the anonymity of online participants.[14] Despite the implementation of data protection mechanisms to prevent bot responses, such as reCAPTCHA, bot programmers easily bypass these safeguards using virtual private networks or servers.[7] As bots become more prevalent and capable, addressing this issue is critical to ensuring data integrity and the validity of research findings that inform policy and practice.

This study goes beyond merely identifying bot-generated responses in online surveys; it systematically examines how bots influence response patterns in a healthcare experience survey. While existing research has focused on detecting fraudulent responses, there is limited understanding of whether and how bot-generated data differ from human responses in patient-reported experience and outcome measures. By analyzing discrepancies in survey responses between bots and real participants, this study provides critical insights into the extent to which bots may distort findings in healthcare research. This knowledge is essential for improving data integrity, refining fraud detection methods, and ensuring that PROMs and PREMs continue to serve their intended purpose—capturing authentic patient experiences to inform healthcare policy and practice.

Methods

Survey

This study was undertaken within a broader evaluation of Ontario Health Teams (OHT) focusing on patient experiences and outcomes. A patient survey was implemented to support planning and evaluation of integrated care for attributed populations in Ontario, Canada. The survey was intended for patients from a large OHT, gathering data on Patient Reported Experience Measures (PREMs) and Patient Reported Outcome Measures (PROMs). Questions included Likert scale items on healthcare use, health outcomes, accessing care, having someone to count on, being heard, managing health, safety, care transitions, digital healthcare use, demographics; as well as an open-text section to provide additional feedback. Six timestamps were programmed throughout the survey. Data was collected between July 2023 to October 2023 using REDCap. The survey was distributed by a generic link through email lists and social media. In order to increase response rates, partway through the distribution period, a Facebook advertisement was employed. Respondents were offered an opportunity to enter a draw for a $25 gift card.

Bot identification

Upon reviewing the qualitative responses and timestamp data, the research team identified clear evidence of bot-generated responses in the survey. Many responses were duplicative, contained overly formal or generic language, or were nonsensical. The response patterns, combined with irregular timestamp data, confirmed the presence of bots. As a result, the team developed specific criteria to systematically identify and remove bot-generated responses from the dataset.

Table 1 illustrates the final criteria for identifying bots. Criteria to identify bots were developed using the following themes: generic and duplicative open-ended responses; concerns with location as determined by postal code (including postal codes that were invalid or nonsensical); timing irregularities; identical timestamps to other participants; illogical demographic reporting; and discrepancies in reporting health care use.

To identify potential bots in the survey data, we developed a three-tier classification system based on specific "red flags" (concerning or questionable features or responses in the data). Tier 1 criteria identified a respondent as a bot if any single criterion occurred on its own (e.g., duplicate open-ended responses, invalid postal codes, or identical timestamps). Tier 2 classified a respondent as a bot if two or more criteria from its list were observed. Finally, tier 3 flagged respondents as bots if any tier 3 criterion was combined with at least two additional criteria from either tier 2 or tier 3, resulting in a total of three or more issues. For example, one tier 2 criteria included having a valid postal code outside of the region where participants received healthcare services. This criterion was not sufficient on its own to categorize a respondent as a bot, as some participants may legitimately live elsewhere and receive care in the area where the participants were sampled. However, when combined with other criteria, it strengthened the likelihood of being a bot-generated response. Hence, Table 1 presents the frequency with which each criterion was met among respondents ultimately classified as bots or humans. While some human respondents exhibited isolated red flags, they did not meet the necessary combination of criteria required for bot classification.

One criterion for identifying bots was the presence of open-text responses that were deemed to be either A) generic or informal, or B) had a similar sentence structure to other responses. To evaluate this more subjective criteria, two reviewers independently coded open-ended responses based on these criteria. Responses were categorized as bots if reviewers agreed they met either criterion. When reviewers disagreed on whether open-ended responses should be classified as generic or informal, a third member of the research team provided the final determination.

Table 1. Frequency of each bot identification criteria

Criteria	Assessed Category		Percentage of respondents identified as bots who met criteria
	Bot (N)	Human (N)	
Tier 1			
Had an identical open-ended question response identical to at least one other respondent (excluding "No)	293	0	44%
Open text is generic/informal or similar sentencing structure	49	0	7%
Duplicate invalid postal codes that were completed on the same day	189	0	28%
Duplicate out of province postal codes (excluding Quebec) that were completed on the same day	152	0	23%
Tier 2			
Identical open-ended question marked as "no" that also fell within the same day and similar time frame	80	0	12%
2 or more survey respondents entered their surveys within 30 seconds of one another (but not identical), and this is the case in 3 or more timestamps	340	2	51%
Had an identical timestamp (including hours, minutes, and seconds) with at least one other respondent on at least one timestamp			
• First timestamp	114	0	17%
• Second timestamp	93	0	14%
• Third timestamp	75	1	11%
• Fourth timestamp	80	0	12%
• Fifth timestamp	69	2	10%
• Sixth timestamp	75	0	11%
Survey was completed between 1:00-5:00am (as identified on ANY of the timestamps)	154	7	23%
Time spent on survey is 3 minutes or less (among particpants who completed the survey)	42	11	6%
Invalid postal code (gibberish, numeric)	189	0	28%
Valid postal code but outside of Ontario	153	0	23%
Valid Ontario postal code but outside of the region of the primary care centre in which participants were recruited	165	5	25%
Discrepancy in reporting ED visits	347	55	52%

Criteria	Assessed Category		
	Bot (N)	Human (N)	Percentage of respondents identified as bots who met criteria
Between the ages of <18- 44 years old but reported living in a retirement home or long-term care facility.	123	1	18%
Tier 3			
Bot takeover day of October 10 or after	618	94	93%
Postal code from neighbouring province	55	3	8%

Bot response analysis

To examine differences in survey responses between humans and bots, bar charts were created to visually compare response patterns. Chi-square tests were conducted to identify significant differences in the proportions of responses between the two groups.

Additionally, Spearman's correlation tests were conducted to evaluate significant relationships between two health indicators, analyzed separately for humans and bots to identify patterns that align with or deviate from expected human behavior. Two sets of correlations were examined: (1) the correlation between "little interest or pleasure in doing things" and "feeling down, depressed, or hopeless," and (2) the correlation between self-rated health and the frequency of interaction with the health system.

Results

Frequency of bot identification

After removing survey responses where consent wasn't given, or no substantive questions were answered (n=231), 1154 survey respondents remained. Among them, 58% (n=668) were categorized as bots, and 42% (n=486) were categorized as humans. Table 1 illustrates the number of respondents that met eligibility for each identification criteria. There were 529 survey respondents that met criteria for Tier 1, 570 respondents met criteria for tier 2 (136 of whom did not meet criteria for tier 1), and 551 respondents met criteria for tier 3 (3 of whom did not meet criteria for tier 1 or tier 2).

In tier 1, the most commonly reported "red flag" was having an identical open-ended response to another survey respondent. In tier 2, the most common 'red flag' was a discrepancy in reporting an emergency department visit, where respondents indicated they had received care from the emergency department in one question but later reported that they had not been to the emergency department in the past 12 months. There was also a high frequency of respondents completing their surveys within similar timeframes (completing three or more sections within 30 seconds of another respondent).

Differences in responses between bots and humans

Overall, humans had a higher percentage of missingness on all questions (an average of 5% for bots vs. 13% for humans).

Chi-square tests revealed significant differences ($P < .001$). in the proportions of responses

between humans and bots for 56 out of 60 survey questions, with only four showing no significant difference.

In terms of demographics (Figure 1), bots had higher proportions of reporting younger age groups (under 18-44), and more frequently reported to be men (whereas humans more frequently reported to be women). While the majority of humans and bots both reported their sexual orientation to be heterosexual, bots more frequently identified as bisexual; homosexual; queer; or two-spirit, compared to humans. Humans and bots both most frequently identified as White, however a higher percentage of bots reported to identify as Black; Asian; Latin; Middle Eastern; Caribbean; or Indigenous.

Visualizations, using bar charts, exposed interesting patterns in the data. We analyzed and report on these patterns according to PROM Likert scale questions, PREM Likert scale questions, ordinal items and nominal unordered items. Overall, we observed that bots were more likely to select mid-scale responses, while humans tended to be more likely than bots to choose options at the extremes of the scale. Regarding PROMs, humans were more likely than bots to report the most favourable outcomes. For example, when responding to concerns about mobility, humans were more likely to report that they had “no problems walking about” whereas bots were more likely to select that they had “slight” or “moderate” problems walking about (Figure 2). For mental health outcomes, humans were more likely than bots to report both the most and least favourable outcomes. For example, when asked about the frequency of feeling down, depressed, or hopeless, humans more frequently responded that they felt this way “not at all” or “nearly every day” whereas bots were more likely to report outcomes in the middle, such as “several days” or “more than half the days” (Figure 2).

In terms of PREMs, when asked about accessing care, human responses were more likely to be evenly distributed across the scale and were more likely than bots to indicate the most and least favourable responses (higher frequencies than bots when indicating it was “very easy” or “very difficult” to access community supports; and describing the length of time it took to get care as “about right” or “much too long).” A higher proportions of bots responded being “somewhat easy” or “somewhat difficult” to get community supports, and the length of time to get care as “somewhat too long.” (Figure 3).

The survey also assessed health service use, including emergency department visits, hospitalizations, and specialist care in the past year. Bots more frequently reported using the emergency department and being hospitalized compared to humans**.** When asked yes/no questions about these services—such as whether an ED visit could have been managed by a regular provider, whether their provider was informed after hospital discharge, or whether a specialist had relevant medical information—bots were more likely to respond "yes" than humans. Consistent with previous patterns, when answering ordinal scale questions, bots tended to select middle-range responses**.** For example, when asked about confidence in managing their health post-ED visit, bots were more likely to report being "somewhat confident" rather than selecting extreme responses.

The survey included one nominal question, which asked about the main reason participants had not looked at their medical records online (for those who indicated this to be the case). There did not appear to be any trend or pattern for the distribution of responses. Interestingly, a substantially higher percentage of bots (compared to humans) indicated that a reason for not having looked at medical records online was not having reliable access to internet (Figure 4).

Correlations

Spearman’s correlations indicated significant correlations between the indicators of having “little interest or pleasure in doing things” and “feeling down, depressed or hopeless,” both of which are indicators of depression. For humans, the Spearman’s correlation was 0.48 ($P < .001$), indicating a moderate positive correlation between the two variables, whereas for bots, the spearman’s

correlation of -0.33 ($P < .001$) indicated a moderate negative correlation between the two variables.

When analyzing the correlation between self-rated health ("how do you rate your own health?") and frequency of interacting with the health system ("how often do you interact with the health system?"), a Spearman's correlation of -0.07 for humans indicated a weak negative correlation, however this was not significant (P=.15). For bots, the Spearman's correlation was 0.06, indicating a weak positive correlation, however this was also not significant (P=.15).

Discussion

This results of this study demonstrate the imperative need for researchers, evaluators and program analysts in the health care field to recognize and address the potential threats posed by bots in online survey research. Failure to detect and remove bot-generated responses can lead to compromised data quality and incorrect inferences. Implementing robust data screening strategies is essential to ensure that only valid responses are retained. These responses respect the integrity of genuine participants whose experiences deserve to be accurately captured.

Implications of including bot data in healthcare surveys

This study uniquely identified variation in response patterns between bots and humans. Across multiple survey domains- health outcomes, healthcare access and experiences - humans were more likely than bots to select the most favourable outcomes (or in some cases, the least favourable response options). In contrast, bots exhibited a tendency to select responses clustered around the middle of the scale. As a result, the presence of bot-generated data could dilute the strength of human responses, potentially skewing average scores toward the midpoint and diminishing the ability to detect meaningful trends. This issue extends beyond scaled responses; even in nominal questions, such as the reason for not accessing medical records online, bots disproportionately selected specific response options, such as lacking reliable internet access. This discrepancy suggests that bot-generated data may introduce artificial patterns in categorical variables, leading to inaccurate estimates of barriers to healthcare access and potentially misinforming policy decisions. This distortion may undermine construct validity, as the data fail to capture reflect real-world experiences.

Beyond response patterns, the inclusion of bot-generated data introduces a significant risk of bias, particularly in research focusing on health equity. The demographics of bot responses differed notably from those of human participants. This discrepancy poses a major threat to the validity of studies that examine healthcare utilization and outcomes among specific demographic subgroups. In equity-focused research, where accurate representation of marginalized populations is crucial, failing to filter out bots could lead to misleading conclusions and inappropriate policy recommendations.

The impact of bot interference is further underscored when examining the relationship between key indicators of depression. Among human respondents, the correlation between "having little interest or pleasure in doing things" and "feeling down, depressed, or hopeless" aligned with established patterns in the literature, reflecting a logical association between these two related constructs. However, among bot-generated responses, this relationship was unexpectedly reversed, indicating a negative correlation that does not align with theoretical expectations. This inconsistency suggests that bot responses do not accurately reflect underlying psychological constructs, reinforcing the need to identify and eliminate bots from survey research.

Bot Detection Strategies

Several methods to identify bots that were used in this study align with those employed in similar research. These include: analyzing open-ended responses for non-sensical, duplicative or inconsistent data; [5,8,1015] tracking survey completion times; [5,7,8,13,16,1715] flagging surveys that were

completed at unusual times of the day;[8] and identifying suspicious patterns related to zip codes and geographic location.[17] We applied all applicable bot detection strategies for which we had data available; however we did not have access to IP addresses, names, or email addresses, which limited our ability to apply certain validation techniques used in prior research. There are additional bot detection techniques which were not used in this study but have been used in prior research. These include: checking participants responses to questions that are asked in two ways[8,14,16] (such as "how old are you?" and "what year were you born?"); using IP addresses to check for responses that were completed in the same household or very close to one-another in relation to the time the survey was submitted; [5,14] and examining email addresses for suspicious wording or similar content.[8,17] Another strategy involves external validation of participant information, such as verifying names, addresses, phone numbers, and birth dates to ensure responses are unique and legitimate.[18] Investigators can also cross-reference participants' details using publicly available databases, social media, or professional networking sites to further confirm authenticity, though ethical considerations must be weighed when implementing such approaches.[18] While these strategies may help reduce fraudulent responses, they can also create additional burdens for legitimate participants, potentially discouraging participation or introducing barriers to informed consent.

Among the bot detection strategies used in this study, some were particularly effective in identifying bot-generated responses. Notably, open-ended responses provided valuable insight, as bot-generated answers often contained repetitive, illogical, or off-topic content. Unusual quantitative responses alone can be difficult to interpret, as they might seem plausible despite being artificially generated. However, open-ended responses offer crucial context that helps verify the authenticity of data. Unlike structured response options, which bots can randomly or systematically select, qualitative responses require coherence and reasoning, making them a powerful tool for distinguishing between genuine participants and automated responses. Geographic inconsistencies—such as invalid or duplicate postal codes, or locations outside the intended survey region—were also strong indicators of bot activity. Additionally, anomalies in time-related metrics, including extremely short survey completion times, completion at unusual hours, or multiple surveys submitted at identical timestamps, further helped differentiate bots from human respondents. Based on these findings, we recommend collecting postal or ZIP codes and incorporating multiple timestamps within surveys to enhance bot detection. While no single strategy is universally effective across all surveys, we recommend employing a combination of detection methods tailored to the specific survey and target population.

While several strategies are outlined to detect bots, it is also important to prevent bots from entering surveys in the first place. Tools such CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) have been developed to offer an additional level of security against bot infiltration.[19] However, research has shown that these tools can be bypassed, as bots are increasingly programmed to recognize and circumvent such challenges.[7,1720]Similarly, honeypot questions (hidden prompts designed to engage only automated respondents) have also proven ineffective, as more sophisticated bots can now avoid these traps.[817] Indeed, researchers who have used this method have reported that bots did not provide responses to these questions.[8] Given the growing complexity of automated survey fraud, it is essential for technology companies to continue developing advanced bot detection solutions. Survey platforms such as REDCap and Qualtrics should integrate these technologies to automate bot prevention, reducing the burden on researchers and ensuring more consistent data quality measures.

It is recommended that bot detection strategies be implemented at multiple stages in the research process. This includes prior to data collection (e.g. using targeted recruitment with specific distribution lists), during data collection (e.g. monitoring data daily to identify potential fraud), and following data collection (e.g. implementing a systemic check of potential fraud).[15] Finally, increasing awareness among researchers about the impact of bots in survey research is crucial. We recommend that training in bot prevention and detection be incorporated into research methods and

survey design coursework to equip researchers with the necessary skills to safeguard data integrity. There will continue to be a value and need to reach out with surveys to broad groups of public and enabling survey research in this context will continue to be important.

Limitations

While this study offers valuable insights, several limitations must be acknowledged. Since bot detection methods were applied to a single case study, their effectiveness has yet to be evaluated across different datasets, which may limit generalizability. Additionally, the data were collected within one region of Ontario, and the characteristics of the human responses—and their differences from bot responses—may not be representative of broader populations. Despite employing various bot detection strategies, there is also potential for undetected sophisticated bots, which could still impact the dataset's validity. Furthermore, the study was conducted at a specific point in time, and as automated response technologies evolve, the effectiveness of these detection methods may diminish. Lastly, the study relied on assumed response patterns to identify bots, but without absolute verification—such as IP tracking or manual validation—there remains some uncertainty in classification.

Conclusions

Ensuring data integrity should be a fundamental component of study design from the outset. By incorporating robust data validation protocols early in the research process, researchers can minimize bias and prevent the inclusion of inauthentic data, strengthening the reliability of findings. As internet-based research (and survey research in particular) continues to expand, there is a growing opportunity for researchers across disciplines to contribute to the development of comprehensive bot detection methodologies. Moving forward, greater efforts should be directed toward developing universal technologies and initiatives for bot detection. Standardized approaches that can be seamlessly integrated across different research platforms will help safeguard the validity of online data collection and ensure that digital survey research remains a trusted tool for generating meaningful insights.

Acknowledgments

Emily Hamovitch and Kaileah McKellar developed coding criteria to identify bots, coded the data accordingly, and take responsibility for the concept and design of the study (with oversight from Walter Wodchis). Emily Hamovitch completed the analysis and drafted the manuscript. Kaileah McKellar and Walter Wodchis provided critical revision of the manuscript.
This research was supported by a grant from the Ontario Ministry of Health to the Health System Performance Network (#694). The funders had no role in data analysis, decision to publish, or preparation of the report. We would like to acknowledge the South Georgian Bay (SGB) Ontario Health Team and their Patient/Client, Family and Caregiver Advisory Council for deciding to undertake the HSPN Patient Survey amongst the entire population of the SGB Family Health Team (FHT) and the SGB FHT for distributing the survey to all patients/clients.

Conflicts of Interest

None declared.

Abbreviations

PROMs: Patient-reported outcome measures

PREMs: Patient-reported experience measures
OHT: Ontario health team
REDCap: Research electronic data capture
CAPTCHA: Completely automated public turing test to tell computers and humans apart

References

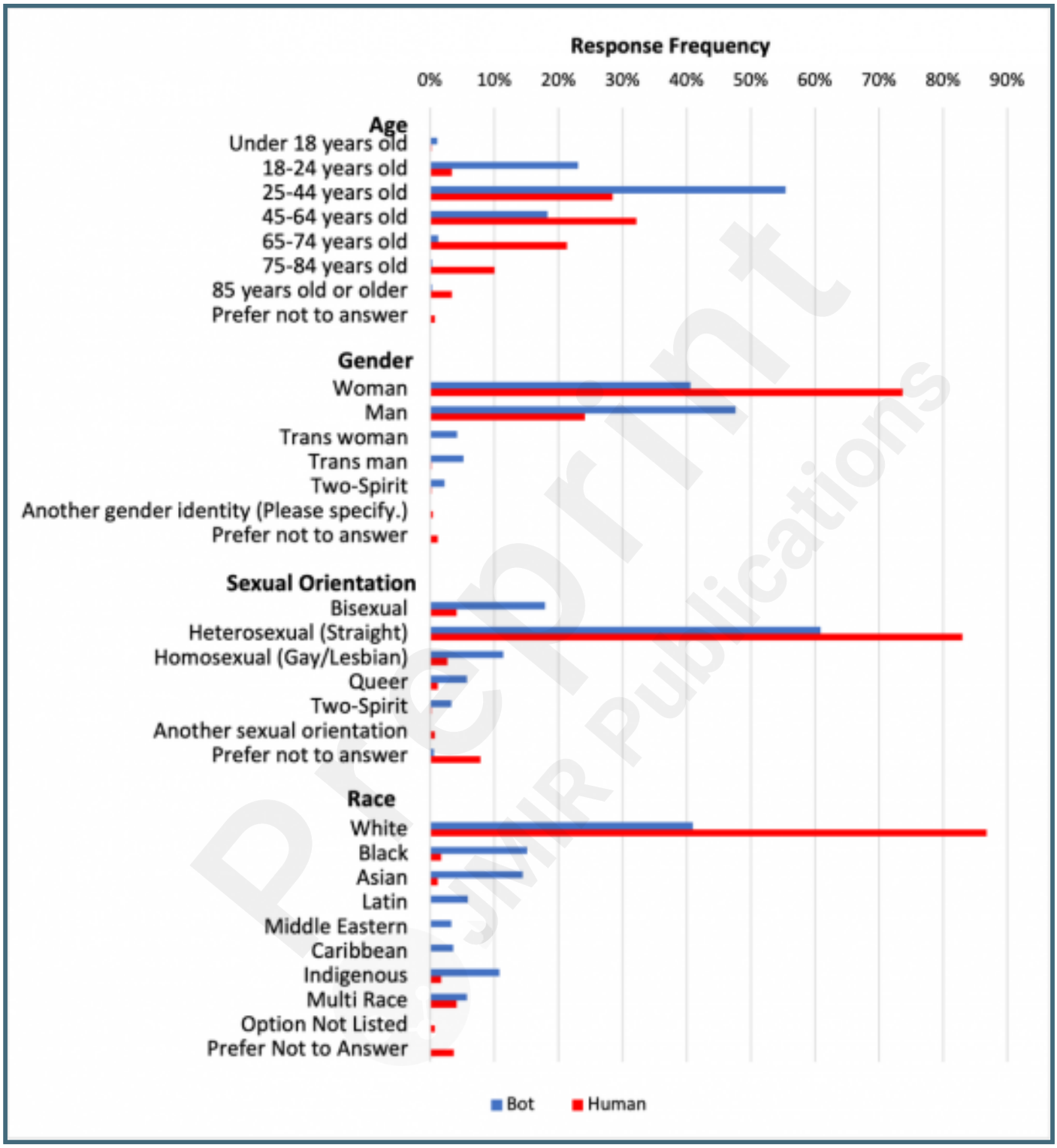
1. Valderas JM, Gibbons C, Porter I, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. *Cochrane DB Sys Rev*. 2021;10: CD011589. doi: 10.1002/14651858.CD011589.pub2.
2. Donald EE, Whitlock K, Dansereau T, Sands DJ, Small D, Stajduhar KI. A codevelopment process to advance methods for the use of patient-reported outcome measures and patient-reported experience measures with people who are homeless and experience chronic illness. *Health Expect*. 2022; 25(5):2264–2274. doi: 10.1111/hex.13489.
3. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Educ*. 2017;17(4):137–144. doi: 10.1093/bjaed/mkw060.
4. Casaca P, Schäfer W, Nunes AB, Sousa P. Using patient-reported outcome measures and patient-reported experience measures to elevate the quality of healthcare. *Int J Qual Health C.* 2023;35(4). doi: 10.1093/intqhc/mzad098.
5. Thompson AD, Utz RL. Online surveys: Lessons learned in detecting and protecting against insincerity and bots. *Quality & quantity*. 2024. doi: 10.1007/s11135-024-01973-z.
6. Singh S, Sagar R. A critical look at online survey or questionnaire-based research studies during COVID-19. *Asian journal of psychiatry*. 2021;65:102850. doi: 10.1016/j.ajp.2021.102850.
7. Griffin M, Martino RJ, LoSchiavo C, et al. Ensuring survey research data integrity in the era of internet bots. *Qual Quant*. 2022;56(4):2841–2852. doi: 10.1007/s11135-021-01252-1.
8. Storozuk A, Ashley M, Delage V, Maloney EA. Got bots? Practical recommendations to protect online survey data from bot attacks. *Quant Meth Psychol.* 2020;16(5):472–481. doi: 10.20982/tqmp.16.5.p472.
9. Dewitt J, Capistrant B, Kohli N, et al. Addressing participant validity in a small internet health survey (the restore study): Protocol and recommendations for survey response validation. *JMIR Res Protoc*. 2018;20(4):e96. doi: 10.2196/resprot.7655.
10. Guy AA, Murphy MJ, Zelaya DG, Kahler CW, Sun S. Data integrity in an online world: Demonstration of multimodal bot screening tools and considerations for preserving data integrity in two online social and behavioral research studies with marginalized populations. *Psychol Methods*. 2024. doi: 10.1037/met0000696.
11. Pozzar R, Hammer MJ, Underhill-Blazey M, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: Cross-sectional questionnaire. *Journal of medical internet research*. 2020;22(10):e23021. doi: 10.2196/23021.
12. Eysenbach G, Wyatt J. Using the internet for surveys and health research. *Journal of medical Internet research*. 2002;4(2):E13. doi: 10.2196/jmir.4.2.e13.
13. Chen AT, Komi M, Bessler S, Mikles SP, Zhang Y. Integrating statistical and visual analytic methods for bot identification of health-related survey data. *J Biomed Inform*. 2023;144:104439. doi: 10.1016/j.jbi.2023.104439.
14. Xu Y, Pace S, Kim J, et al. Threats to online surveys: Recognizing, detecting, and preventing survey bots. *Soc Work Res*. 2022;46(4):343–350. doi: 10.1093/swr/svac023.
15. Johnson MS, Adams VM, Byrne J. Addressing fraudulent responses in online surveys: Insights from a web-based participatory mapping study. *People and nature (Hoboken, N.J.)*. 2024;6(1):147–164. doi: 10.1002/pan3.10557.
16. King-Nyberg B, Thomson E, Morris-Reade J, Borgen R, Taylor C. The bot toolbox: An

accidental case study on how to eliminate bots from your online survey. *Journal for Social Thought*. 2023;7.
17. Goodrich B, Fenton M, Penn J, Bovay J, Mountain T. Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys. *Applied economic perspectives and policy*. 2023;45(2):762–784. doi: 10.1002/aepp.13353.
18. 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.
19. Loebenberg G, Oldham M, Brown J, et al. Bot or not? detecting and managing participant deception when conducting digital research remotely: Case study of a randomized controlled trial. *Journal of medical internet research*. 2023;25(5):e46523. doi: 10.2196/46523.
20. Lebrun B, Temtsin S, Vonasch A, Bartneck C. Detecting the corruption of online questionnaires by artificial intelligence. *Front Robot AI*. 2023;10:1277635. doi: 10.3389/frobt.2023.1277635.

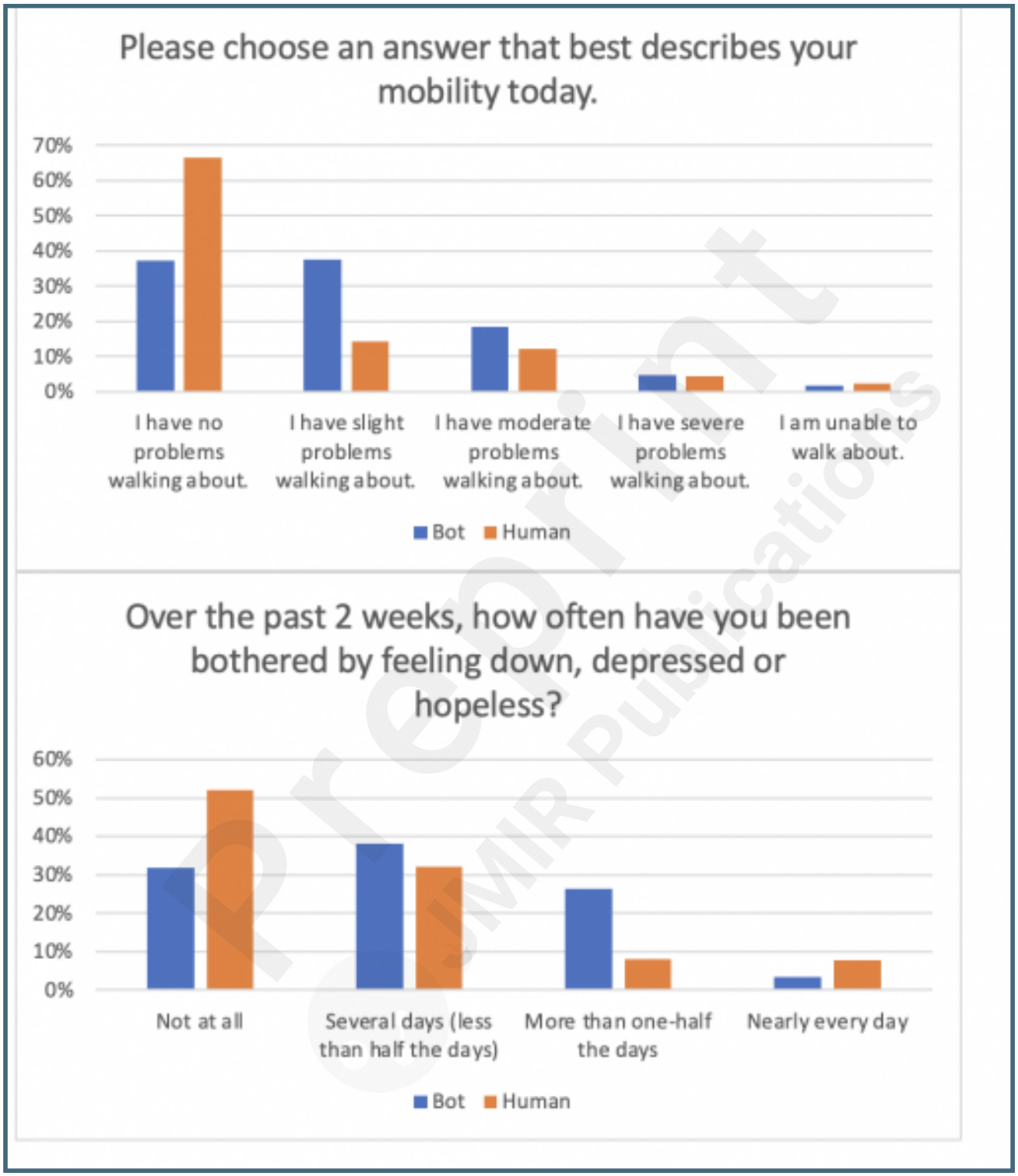
Supplementary Files

Figures

Distribution of demographics for bots vs. Humans.



Distribution of human and bot responses for PROMsUntitled.



Distribution of human and bot responses for PREMs.