Fraud Detection in Incentive-Based Online Surveys Through Follow-up Demographic Verification

Andy E. Lewis[*a], Kevin B. Gittner[ab], & Lauren M. Matheny[b]*

[a]*Department of Health Promotion and Physical Education, Kennesaw State University, Kennesaw, GA*
[b]*School of Data Science and Analytics, Kennesaw State University, Kennesaw GA*

https://doi.org/10.33697/ajur.2025.156

*Student: alewi240@students.kennesaw.edu**
Mentors: kgittner@kennesaw.edu, lmathen1@kennesaw.edu

ABSTRACT
Online survey platforms like Qualtrics, SurveyMonkey, and Amazon Mechanical Turk have become essential tools for fast, cost-effective data collection, but growing concerns over data quality and the risk of fraud have accompanied their rapid growth. The purpose of this study was to determine whether follow-up demographic verification surveys can identify fraudulent participation and improve data quality in survey research conducted through online survey platforms.
An orthopedic activity level survey with 35 embedded data quality checks was distributed using Amazon Mechanical Turk. A follow-up demographic verification survey was sent to 28 participants who contacted the survey administrator after completing the original survey. Five datasets were cleaned and merged using R, allowing for the identification of potential patterns of fraud through the direct comparison of demographic information between the original and follow-up surveys.
Participants who did not complete the follow-up survey exhibited signs of fraud and poor data quality, including one email being linked to multiple participant accounts and higher counts of failed data quality checks. Although the small sample size limited the ability to detect statistical significance, descriptive patterns indicate practically meaningful differences. In contrast, demographic discrepancies were minimal among those who completed the follow-up. Open-text box similarity detection was the most effective individual data quality check.
Integrating follow-up verification surveys into study designs provides a practical and scalable approach to detect fraud and maintain data integrity before compensation is distributed, offering researchers a cost-effective method to protect incentive budgets and data quality within online survey research.

KEYWORDS
Online Survey Platforms; Data Quality; Online Research; Fraud Detection; MTurk; Participant Verification; Demographic Consistency; Incentive Structures; Follow-Up Surveys; Crowdsourced Data

INTRODUCTION

Online surveys play a pivotal role in shaping decisions, and the industry's current value and projected growth demonstrate their widespread adoption and expanding role. Already valued at an estimated three billion dollars in 2024, the online survey industry is expected to continue to grow, projected to reach $36 billion by 2030.[1, 2] Online survey platforms such as Momentive (herein referred to as SurveyMonkey), Amazon Mechanical Turk (MTurk), and Qualtrics allow researchers and other professionals across a broad spectrum of fields to collect and harness data in a quicker, cost-effective, and accessible manner than traditional methods.[3–5]

These platforms have extended their reach into public, private, and academic sectors. Regardless of the field or sector, obtaining good-quality data is integral to making valid conclusions that inform decision-making. Despite these advantages, online survey platforms introduce risks of fraud and careless responses, raising concerns among researchers about threats to data quality and research outcomes.[3, 6, 7] Financial incentives can further motivate bad actors to prioritize quantity over quality, leading to fraudulent activities in survey responses that waste valuable resources and distort study results.[8–10]

Survey incentives have been linked to an increase in multiple submissions,[11, 12] and while strategies have been developed to mitigate fraudulent participation, many remain ineffective against those who continue to adapt their techniques to exploit incentive structures. To combat survey fraud, researchers have explored altering incentives, verifying demographic details to identify repeated attempts, and delaying or withholding compensation until fraud detection checks are complete.[9, 13–15] However, the effectiveness of these strategies remains uncertain and requires further evaluation. Without effective deterrents, researchers risk allocating limited resources to unreliable data. While artificial intelligence (AI) and deep learning offer tools for detecting fraudulent survey responses, both automated and human-generated, their practical use remains limited and not yet widely validated.[16, 17] At the same time, AI lowers the technical barriers for fraudulent actors to exploit surveys. These tools enable automated submissions and allow individuals lacking proficiency in the survey language to use large language models for translation, compromising instruments validated for linguistically proficient respondents and complicating detection of invalid responses.[17, 18]

Given these ongoing challenges, the purpose of this study was to assess and identify fraudulent responses using a follow-up demographic verification survey prior to incentive disbursement. In addition to the embedded data quality checks in the original survey, the follow-up survey functioned as a secondary fraud detection step for a subset of participants who contacted the survey administrator. Instead of relying solely on post-hoc fraud detection methods that can only be used after survey completion and participant compensation, the goal of the follow-up survey was to confirm participant identities through time-delayed verification of reported static demographics and identify discrepancies with their initially reported demographic information. By verifying demographic consistency, this study employs a scalable solution for detecting inconsistencies before incentives are distributed, thereby reducing fraudulent claims, conserving research resources, and improving data quality.

METHODS AND PROCEDURES

Survey design and data collection

An orthopedic activity level assessment survey was distributed in the Summer of 2024 through MTurk. This study was approved by the Kennesaw State University Institutional Review Board (IRB-FY23-373). To examine data quality, 35 quality checks were embedded into the original survey. These checks included logic, demographic consistency, and open-text similarity checks, among others. Participants entered the study with a unique MTurk participant ID. The original survey was divided into two parts, with part one including a set of screening questions and part two containing the primary content for those who qualified. Upon completing the original survey, participants received a survey completion code. This code was required to receive payment through MTurk and was used to verify participation.

Following the completion of the original survey, the data cleaning process was initiated to identify potential fraud. During this time, some respondents contacted the survey administrator about incentive payments, at which point they were sent a follow-up demographic verification survey. These individuals were the only participants to receive the follow-up survey, as the purpose was to confirm their identity before addressing their concerns or distributing compensation. This approach provided a necessary verification step, but it limited the follow-up survey to a self-selected subset, potentially introducing bias. To assess data quality differences, the average number of failed data quality checks in the original survey was compared between those who completed the follow-up survey and those who did not complete it after reaching out to the survey administrator.

Data cleaning and merging process

All data cleaning, handling, and analysis were conducted in R version 4.3.1 using RStudio.[19] Data cleaning and initial quality assessments focused on identifying discrepancies, including missing participant IDs, duplicate entries, and incorrect survey completion codes. These inconsistencies, combined with extensive fraud, made a straightforward merge between datasets impossible. Survey outputs were examined to align participant IDs, survey completion codes, and email confirmation codes used to link the original survey data with the follow-up survey response data. A frequency table was generated to detect duplicate IDs, while other inconsistencies were resolved through manual validation and recoding only when necessary. Email communications were referenced to match records in cases where participant IDs and survey completion codes were missing or incorrect.

Five datasets were imported into RStudio and merged for analysis:

A. Part one of the original survey output (inclusion criteria screening survey)
B. Part two of the original survey output (primary survey)
C. An MTurk file linking each participant to a submission in the original survey, including each Amazon Mechanical Turk participant's unique identifier and their corresponding survey completion code
D. A file containing messages from participants who contacted the survey administrator
E. The follow-up demographic verification survey output

Preparation for analysis involved a series of merges, using both inner and outer joins, to integrate all relevant data into a single dataset (**Figure 1**). First, part one (A) and part two (B) of the original survey output were merged, then linked to MTurk records (C). Duplicated responses were removed, producing a cleaned dataset of unique responses from the original survey. Next, messages from participants who contacted the survey administrator (D) were merged with the follow-up survey responses (E), matching individuals who had reached out to the survey administrator with their follow-up data. Finally, the follow-up survey dataset (D & E) was merged with the cleaned original survey and MTurk records (A, B, & C), creating a single dataset (n = 28) with unique identifiers. The final dataset was used to validate participant demographics between the original and follow-up survey responses, confirming whether respondents reported the same residential ZIP code, birth year, and email address.

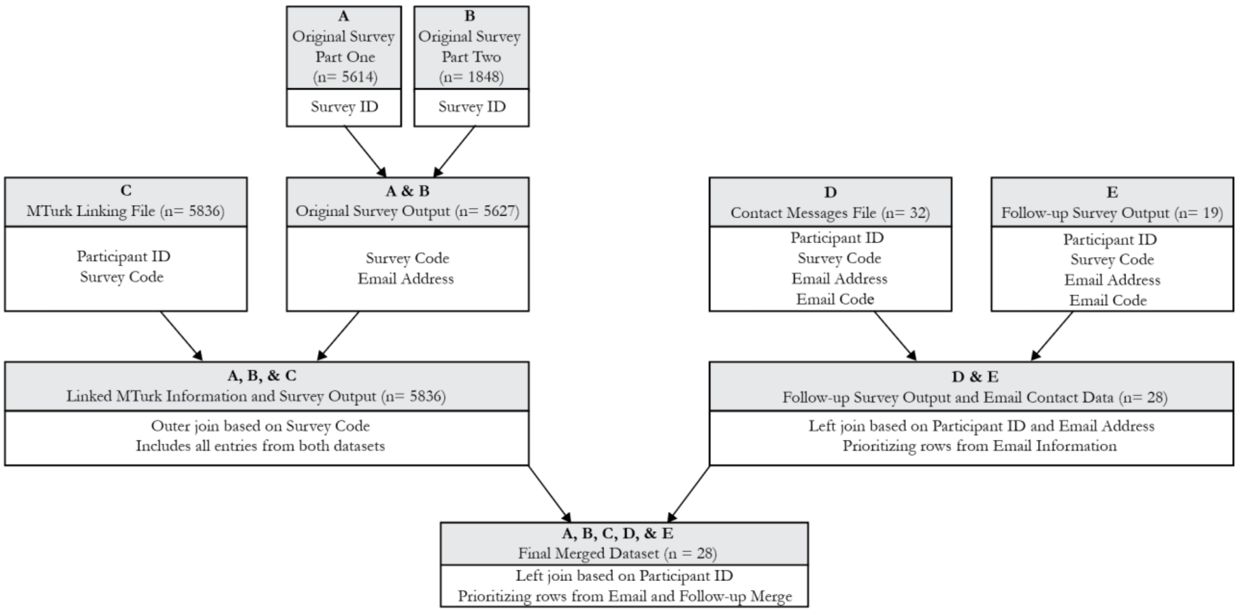

Figure 1. Merging steps. Five datasets combined into a single file for analysis.

Statistical analysis
Descriptive statistics were calculated to summarize demographic consistency checks and original survey quality check failures. Deidentified patterns were visualized to summarize trends observed after the merging process. The mean number of failed embedded data quality checks in the original survey was used as an indicator of data quality and was compared between participants who completed the follow-up survey and those who did not. Histograms of residuals, QQ-plots, and a Shapiro-Wilk test were used to assess the normality of the distribution of counts of failed data quality checks from the original survey for participants who completed the follow-up survey and those who did not. Mean and median counts of failed data quality checks were also compared to assess potential differences between the groups. A Mann-Whitney U test was conducted to determine whether there was a statistically significant difference in the distribution of the number of failed data quality checks between participants who completed the follow-up and those who did not. Finally, summary statistics were used to evaluate the effectiveness of the original survey's 35 individual data quality checks, which were designed a priori to identify the most effective checks.

RESULTS

Demographic comparison
After removing duplicates, a total of 28 emails were sent to the survey administrator. The demographic verification survey was sent to all 28 participants, of whom 53.6% completed the follow-up survey. Among those who completed the follow-up, only minimal demographic discrepancies were observed. Specifically, there was one ZIP code, two birth year, and two email inconsistencies between the original survey and the follow-up survey. Although these inconsistencies raise slight concerns about data reliability from this group, the issues were minor compared to the discrepancies observed among participants who did not complete the follow-up survey. Demographic consistency validation revealed that participants who completed the follow-up survey (n = 15) showed consistency across ZIP codes, birth years, and email addresses, whereas participants who did not complete the follow-up survey (n = 13) demonstrated substantial email inconsistencies (**Table 1**).

Consistency Check	Completed follow-up (n = 15)	Did not Complete follow-up (n = 13)
ZIP Code	14/15 (93.3%)	—
Birth Year	13/15 (86.7%)	—
Email	13/15 (86.7%)	3/13 (23.1%)

Table 1. Demographic consistency across participants who completed and did not complete the follow-up survey. For participants who completed the follow-up survey, demographic consistency was assessed by comparing the ZIP Code, birth year, and email reported in the follow-up survey with those in the original survey. For participants who did not complete the follow-up, email consistency was assessed by comparing the email used to contact administrator with email used in the original survey.

Among participants who did not complete the follow-up survey (n = 13), 76.9% demonstrated email inconsistencies (**Table 1**), including six respondents who were all linked to a single email address across six separate MTurk worker accounts (**Table 2**). This pattern suggests that one individual may have taken the survey multiple times under different identities, with the time-delayed verification introducing recall challenges that reduced the consistency of fraudulent responses.

Participant (n = 13)	Contact Email	Original Survey Email	Email Consistency?
Participant ID 1	Contact-A@example.com	**Survey-A@example.com**	✗
Participant ID 2	Contact-B@example.com	**Survey-A@example.com**	✗
Participant ID 3	Contact-C@example.com	**Survey-A@example.com**	✗
Participant ID 4	Contact-D@example.com	**Survey-A@example.com**	✗
Participant ID 5	Contact-E@example.com	**Survey-A@example.com**	✗
Participant ID 6	Contact-F@example.com	**Survey-A@example.com**	✗
Participant ID 7	Contact-G@example.com	Survey-B@example.com	✗
Participant ID 8	Contact-H@example.com	Survey-C@example.com	✗
Participant ID 9	Contact-I@example.com	Survey-D@example.com	✗
Participant ID 10	Contact-J@example.com	Survey-E@example.com	✗
Participant ID 11	Contact-K@example.com	Contact-K@example.com	✓
Participant ID 12	Contact-L@example.com	Contact-L@example.com	✓
Participant ID 13	Contact-M@example.com	Contact-M@example.com	✓

Table 2. Email inconsistencies among those who did not complete the follow-up survey (n = 13). ✓ indicates that the contact email used to reach the survey administrator matched the email provided in the original survey. ✗ indicates a mismatch between the contact email and the original survey email. The bolded example email (e.g., Survey-A@example.com) was used across multiple unique participant accounts.

Descriptive analysis

To assess whether there were observable differences in data quality, the counts of failed data quality checks from the original survey were compared between participants who completed the follow-up survey and those who did not. Participants who did not complete the follow-up survey had higher mean and median counts of failed data quality checks in the original survey (mean = 5.9, SE = .84, median = 6), compared to those who did complete the follow-up survey (mean = 4.0, SE = 1.03, median = 2.5). Normality assessments indicated a violation of normality in the group that completed the follow-up survey (W = 0.797, p = 0.005). Given the violation of normality and the small sample size, non-parametric measures were used to compare the counts of failed data quality checks in the original survey between participants who completed the follow-up survey and those who did not. The Mann-Whitney U test found no significant difference in failed quality checks between groups (U = 125, p = 0.101). Given the small sample size of this study, the analysis was underpowered, limiting the ability to detect group differences. However, higher counts of failed data quality checks among those who did not complete the follow-up suggest a trend of poorer data quality, while clustering of high-quality check failures among some participants that completed the follow-up survey demonstrates that layered detection approaches remain necessary (**Figure 2**).

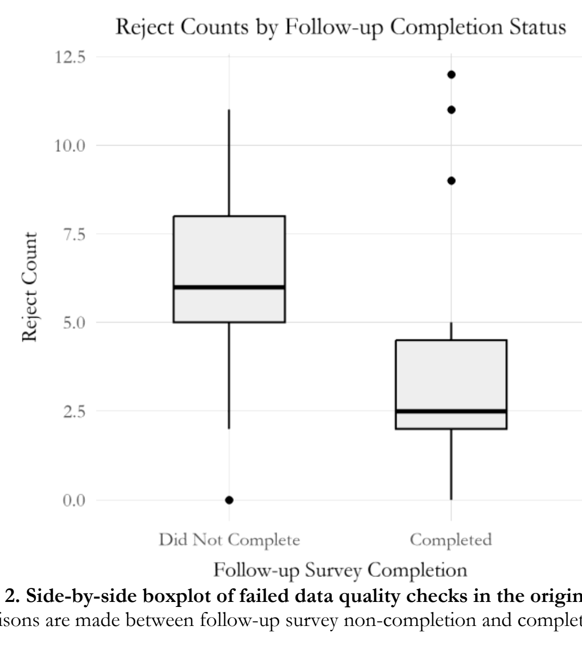

Figure 2. Side-by-side boxplot of failed data quality checks in the original survey.
Comparisons are made between follow-up survey non-completion and completion groups.

Effectiveness of specific data quality checks

A deeper analysis of the 28 follow-up participants revealed that certain data quality checks were more effective at identifying poor data quality in the original survey. Among these, half of the responses failed data quality checks due to email-related issues. Specifically, seven participants had different email confirmation codes reported than the one sent to them, and seven had duplicated email addresses used across multiple submissions. An additional six cases involved the participants' provided age not matching the provided year born with a two-year margin of error. Another six responses failed checks because the first seven words of one open-ended question regarding ankle function were identical to those in another response. Additionally, 14 different cases failed the same seven word similarity check for a separate open-ended question about honesty. Both open-ended checks failed participants whose answers matched text from other responses in the original survey. These counts of failed data quality checks for similarity detection suggest that some participants may have copied text from other responses or reused their answers across multiple entries.

Eight respondents provided a residential ZIP code that was inconsistent with their reported state of residence. In logical comparison data quality control checks, five failed checks related to task difficulty (e.g., reporting that getting out of a tub was more difficult than running a marathon). Twelve respondents failed checks for incorrectly identifying how they received the original survey. Although all participants were recruited through MTurk, they selected options such as Facebook, Instagram, or word of mouth instead of the correct answer, "Other." Additionally, four respondents failed a response strategy check for selecting "careless" or "inattentive" when asked to indicate their strategy for responding to the survey when they should have selected "Other." **Table 3** summarizes the number of respondents that failed each data quality check out of the total sample of 28.

Data Quality Check	Number of Failed Data Quality Checks	Percentage of Participants with Failed Checks (n = 28)
Open-ended honesty similarity checks	14	50%
Survey recruitment method check	12	42.8%
State vs ZIP Code check	8	28.6%
Email code check	7	25%
Duplicate email check	7	25%
Age check	6	21.4%
Open-ended ankle similarity check	6	21.4%
Tub and marathon check	5	17.9%
Response strategy check	4	14.3%

Table 3. Counts and percentages of respondents failing data quality checks (n = 28).

DISCUSSION

Among those who contacted the survey administrator, similarity detection in open-ended responses proved to be the most effective in identifying poor data quality. Participants who did not complete the follow-up survey showed higher reject counts compared to those who did. While this finding was not statistically significant, results indicate a practically meaningful trend that may extend to other online survey platforms, suggesting that participants who avoid verification steps may disproportionately contribute to fraudulent or low-quality data. Replicating findings in a larger sample size would provide the statistical power needed to detect group differences and better evaluate whether the descriptive trend reflects both a practical and statistically significant effect.

Although these findings were observed in MTurk data, concerns about inattentive or fraudulent participation are not exclusive to this platform, and there is no clear consensus on which platform produces the best data. While some studies suggest that Prolific produces higher-quality data than MTurk or Qualtrics,[6, 8] no platform is immune to fraud or poor-quality responses, making verification methods, such as follow-up demographic checks, a solution that can be applied regardless of the platform. One way to scale this approach is to make completion of a delayed time follow-up demographic verification survey a condition for receiving incentives in any study conducted through online survey platforms.

Age and ZIP code consistency could not be fully validated for participants who did not complete the follow-up survey. Analysis was restricted to comparing contact emails with emails from the original survey, limiting the ability to assess fraudulent behavior entirely. This limitation also serves as a key finding, as the high rate of non-response may itself indicate intentional avoidance of verification when dishonest participants are presented with verification before receiving incentives. The follow-up survey was sent after the original survey in response to participant contact, allowing those who did not complete it to opt out of the verification process. This self-selected follow-up group should be interpreted with caution, given the potential for selection bias. Finally, because this study was conducted exclusively on MTurk with a small, nonrandom sample, findings may not be generalizable to other online survey platforms or broader populations.

The follow-up survey enabled the identification of demographic inconsistencies, allowing for the detection of potential fraudulent and dishonest participation. Notably, many of those who did not complete the follow-up survey exhibited suspicious behavior, with nearly half having multiple MTurk participant accounts linked to a single email within the original survey. This supports and expands on other research that examined the use of follow-up surveys as a secondary point of contact, such as demographic verification, as a method to validate participant identity, detect fraud, and reduce fraudulent participation.[9, 14, 15] The results of this study demonstrate how such methods can identify similar deliberate attempts at providing dishonest information, as reported demographic information such as ZIP code, year of birth, or email address should not change in delayed short-term follow-up requests.

Researchers can implement similar follow-up verification procedures within their study planning to serve as an additional fraud detection mechanism after the completion of primary data collection, allowing for a delayed evaluation of respondent legitimacy. This approach is a valuable protective measure for researchers in managing incentive distribution efficiently, as verification prior to payout can reduce unnecessary costs associated with fraudulent claims. This study demonstrates how even non-expert researchers can adopt manual follow-up procedures to detect fraud.

Automation could streamline verification by flagging inconsistent demographic responses, eliminating the need for manual data cleaning. Future research could explore the use of machine learning techniques to automate the detection of repeated responses in open-ended questions. This would improve scalability and reduce manual labor in larger datasets. By incorporating follow-up verification surveys into the research design, investigators can improve data quality while reducing fraudulent claims that exploit survey incentives. As fraud detection methods continue to evolve, integrating multi-layered verification strategies will be essential for maintaining data integrity in online research.

CONCLUSIONS

High-quality data collection through online survey platforms continues to be a challenge, as incentives can increase vulnerability to fraudulent exploitation. When researchers lack a defined sampling frame and are unable to verify participant identity at baseline, a common limitation in online data collection, linking incentives to a follow-up verification step provides a practical and scalable strategy to reduce fraud and preserve data quality. As fraud tactics continue to evolve alongside advancements in artificial intelligence, verification measures should adapt in parallel. This study confirmed that a single individual attempted to manipulate the incentive system by contacting the research team under multiple participant IDs tied to the same survey email address in the original survey. This finding highlights the very real risk of deliberate fraud in online data collection platforms. Future large-scale studies are needed to assess this behavior further and refine follow-up verification as a countermeasure.

ACKNOWLEDGEMENTS

The student author thanks the members of the Kennesaw State University's Data Quality and Survey Methods Lab for their assistance throughout the project. Additional thanks go to Kennesaw State University for providing the resources and environment that made this work possible.

REFERENCES

1. The Business Research Company, Online Survey Software Global Market Report 2025, *https://www.thebusinessresearchcompany.com/report/online-survey-software-global-market-report* (accessed Jan 2025)

2. Knowledge Sourcing Intelligence LLP, Online Survey Software Market Size: Industry Report, 2023–2028, *https://www.knowledge-sourcing.com/report/global-online-survey-software-market* (accessed Jan 2025)

3. Newman, A., Bavik, Y. L., Mount, M., and Shao, B. (2021) Data collection via online platforms: challenges and recommendations for future research, *Appl Psychol Health Well Being* 70(3), 1381–1387. *https://doi.org/10.1111/apps.12302*

4. Callegaro, M., Manfreda, K., and Vehovar, V. (2015) Survey Research and Web surveys, in *Web Survey Methodology* (Metzler, K., Ed.) 1st ed., 1–34, SAGE Publications Ltd., London. *https://doi.org/10.4135/9781529799651*

5. Agans, J. P., Schade, S. A., Hanna, S. R., Chiang, S. C., Shirzad, K., and Bai, S. (2024) The inaccuracy of data from online surveys: a cautionary analysis. *Qual Quant* 58(3), 2065–2068. *https://doi.org/10.1007/s11135-023-01733-5*

6. Douglas, B. D., Ewell, P. J., and Brauer, M. (2023) Data quality in online human-subjects research: comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA, *PLOS ONE* 18(3) e0279720, 1–5. *https://doi.org/10.1371/journal.pone.0279720*

7. Dupuis, M., Renaud, K., and Searle, R. (2022) Crowdsourcing quality concerns: an examination of Amazon's Mechanical Turk, *Proceedings of the 23rd Annual Conference on Information Technology Education*, 127–129. *https://doi.org/10.1145/3537674.3555783*

8. Peer, E., Rothschild, D., Gordon, A., Evernden, Z., and Damer, E. (2022) Data quality of platforms and panels for online behavioral research, *Behav Res* 54, 1643–1653. *https://doi.org/10.3758/s13428-021-01694-3*

9. Johnson, M. S., Adams, V. M., and Byrne, J. (2023) Addressing fraudulent responses in online surveys: insights from a web-based participatory mapping study, *People Nat (Hoboken)* 6(1), 148–161. *https://doi.org/10.1002/pan3.10557*
10. Kumarasamy, V., Goodfellow, N., Ferron, E. M., and Wright, A. L. (2024) Evaluating the problem of fraudulent participants in health care research: multimethod pilot study, *JMIR Form Res* 8 e51530, 2–10. *https://doi.org/10.2196/51530*
11. Chesney, T., and Penny, K. (2013) The impact of repeated lying on survey results, *SAGE Open* 3(1), 1–2. *https://doi.org/10.1177/2158244012472345*
12. Langer, A., Leopold, L., Meuleman, B., Kuppens, L., and Menon, A. (2025) Mitigating fraud in incentivized online surveys: lessons from Facebook recruitment in Nigeria, *Surv Method: Insights Field* 3(2). 4–7. *https://doi.org/10.13094/SMIF-2025-00006*
13. Johns Hopkins University, Guidance on Fraud Prevention Regarding Use of Survey Instruments, Version 1, *https://publichealth.jhu.edu/sites/default/files/2025-01/Fraud-Prevention-Regarding-Use-of-Survey-Instruments.pdf* (accessed Jan 2025)
14. Teitcher, J. E. F., Bockting, W. O., Bauermeister, J. A., Hoefer, C. J., Miner, M. H., & Klitzman, R. L. (2015) Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs, *J Law Med Ethics* 43(1), 119–128. *https://doi.org/10.1111/jlme.12200*
15. Guy, A. A., Murphy, M. J., Zelaya, D. G., Kahler, C. W., and Sun, S. (2024) 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,* 14–16. *https://doi.org/10.1037/met0000696*
16. NORC. Promise and Pitfalls of AI-Augmented Survey Research, *https://www.norc.org/research/library/promise-pitfalls-ai-augmented-survey-research.html* (accessed Sep 2025)
17. Dynata. The Double-Edged Sword of Generative AI: Mitigating the Risks of Respondent Fraud, *https://www.dynata.com/why-dynata/resources/blog/the-double-edged-sword-of-generative-ai-mitigating-the-risks-of-respondent-fraud/* (accessed Sep 2025)
18. Lebrun, B., Temtsin, S., Vonasch, A., and Bartneck, C. (2024) Detecting the corruption of online questionnaires by artificial intelligence, *Front Robot AI* 10, 1–10. *https://doi.org/10.3389/frobt.2023.1277635*
19. R Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, *https://www.R-project.org/* (accessed Feb 2025)

ABOUT THE STUDENT AUTHOR

Andy Lewis graduated from Kennesaw State University in Spring 2025 with a Bachelor of Science in Integrated Health Science. He will be continuing his education at Kennesaw State as a graduate student in the Master's in Data Science and Analytics program. This project served as an important stepping stone in that journey, combining his interests in research, public health, and data analysis.

PRESS SUMMARY

Online surveys are a popular tool for researchers but can be vulnerable to fraud. This study investigated whether a short follow-up survey could help confirm participant identity and improve data quality. Researchers initially conducted an orthopedic survey on Amazon Mechanical Turk to evaluate data quality on the platform. Participants who reached out to the survey administrator about compensation were later sent a follow-up survey to verify demographic details. The analysis revealed signs of possible dishonesty, including one email linked to multiple participant accounts and other inconsistencies. The findings suggest that follow-up verification surveys can help protect data quality and research budgets.