Received: 9 October 2022 | Accepted: 13 February 2023

DOI: 10.1002/aepp.13353

FEATURED ARTICLE

AAEA Agricultural & Applied Economics Association WILEY

Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys

Brittney Goodrich[1] | **Marieke Fenton**[1] | **Jerrod Penn**[2] | **John Bovay**[3] | **Travis Mountain**[4]

[1]Department of Agricultural and Resource Economics, University of California, Davis, Davis, California, USA

[2]Department of Agricultural Economics and Agribusiness, Louisiana State University, Baton Rouge, Louisiana, USA

[3]Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, Virginia, USA

[4]Department of Consumer Sciences, University of Alabama, Tuscaloosa, Alabama, USA

Correspondence
Brittney Goodrich, University of California, Davis, Department of Agricultural and Resource Economics, One Shields Ave, Davis, CA 95616, USA.
Email: bkgoodrich@ucdavis.edu

Funding information
Project Apis m. and National Honey Board; University of California Giannini Foundation of Agricultural Economics

Editor in charge: Daniel Petrolia

Abstract

Declining survey response rates have driven many researchers to seek out cost-effective methods of increasing participation, such as conducting surveys online, paying incentives, and using social media to engage hard-to-reach populations. Malicious actors can exploit the monetary incentives and anonymity of online surveys, threatening the integrity of survey data. We share two recent experiences conducting online surveys that were inundated with fraudulent responses. Our objective is to increase awareness of this emerging issue and offer guidance for others to mitigate the effects of fraudulent responders in their own research.

KEYWORDS

automated bots, data integrity, fraudulent responses, survey methods, survey response rates

JEL CLASSIFICATION

C83, Q00

The quality of social science research depends fundamentally on the quality of the data collected, whether those data come from primary or secondary sources. Recruiting participants is increasingly difficult and expensive. Meanwhile, survey response rates are declining, including at the federal level.[1] Increasing non-response creates issues with representativeness and

potential response bias, leading to inaccuracies and suboptimal policy outcomes due to the use of many survey results as a foundation for policy implementation (Johansson et al., 2017; Sartore et al., 2019; Weigel et al., 2021).

Many researchers have investigated ways to counteract declining response rates, including the use of monetary incentives, multiple modes, points of contact, etc. (Avemegah et al., 2021; Dillman et al., 2014; Hardigan et al., 2012; Weigel et al., 2021). Online surveys distributed through email and anonymous links can be convenient and cost-effective tactics (Hardigan et al., 2012; Schonlau & Couper, 2017). Similarly, many researchers have taken advantage of conducting online surveys by purchasing responses through various marketing panels or platforms, for example, Qualtrics, Amazon's Mechanical Turk (MTurk), or Google Surveys. However, online surveys more acutely face issues affecting data quality, such as inattention or carelessness of participants (Cheng et al., 2022; Gao et al., 2016; Malone & Lusk, 2018a), non-representative samples (Penn et al., 2023; Sandstrom et al., 2023; Whitehead et al., 2023) and fraudulent responses (Belliveau & Yakovenko, 2022; Chmielewski & Kucker, 2020; Griffin et al., 2021; Kramer et al., 2014; Rommel et al., 2022; Storozuk et al., 2020; Teitcher et al., 2015).

Fraudulent responses can come in many forms, some creating more data quality issues than others. Lawlor et al. (2021) distinguish between *unique participant fraud*, individuals who access a survey multiple times for malicious or non-malicious reasons, and *alias fraud*, which involves a single individual using sophisticated techniques to conceal their identity and submit multiple responses to take advantage of participant incentives. Thus, fraudulent responses may range from a small number (Bauermeister et al., 2012; Teitcher et al., 2015) to thousands from automated bots or low-wage laborers abroad (Griffin et al., 2021; Moss et al., 2021; Simone, 2019; Storozuk et al., 2020). "Bots" (short for "robots") are a relatively recent and emerging problem with online surveys. Bots involve computer software designed by a human programmer to perform automated tasks, such as finding and completing online surveys that offer participant incentives (Griffin et al., 2021; Storozuk et al., 2020). Such software is easily available and can be manipulated for specific purposes (Buchanan & Scofield, 2018).

In this paper, we share two recent experiences with suspected *alias fraud*, in two separate online surveys conducted with farm and agribusiness operators. The first survey elicited responses to a discrete choice experiment from the U.S. beekeeping industry. To increase response numbers, participants received a $20 Amazon gift card for completing the survey, and the survey was publicized widely through various beekeeping organizations, including on social media. This resulted in over 2500 responses, more than double the estimated target population, with an estimated 96% of responses being fraudulent. The second survey targeted farm operations and agribusinesses in Virginia, and used a lottery to incentivize and increase participation: among the first 500 responses, 100 randomly selected winners would receive a $10 Amazon gift card. This survey was publicized through Virginia Cooperative Extension (VCE) networks, farm and agribusiness associations in the state, and social media. Out of 444 total responses to the Virginia survey, we assessed 72% to be fraudulent. In general, fraudulent responses to quantitative questions were statistically different from valid responses, which would bias results for any economic analysis, such as estimating willingness to pay.

While several disciplines have documented the growing number of incidents with severe *alias fraud* from automated bots, we were unaware of such documentation, leading to our naïveté while conducting our online surveys. The goal of this paper is to document our experiences to increase awareness of the potential for fraudulent responses and provide practical advice for mitigation strategies in future survey research. Both of these experiences involve distributing convenience surveys to target populations via anonymous links as opposed to purchasing panel data. However, purchased panel data can also face fraudulent responses from sources

such as Qualtrics (Belliveau & Yakovenko, 2022; Johnston et al., 2021) and MTurk (Chandler & Paolacci, 2017; Dennis et al., 2020; Kennedy et al., 2020; Moss et al., 2021). Thus, the risk of fraudulent responses to online surveys is an emerging issue for seemingly all online surveys using non-probability sampling. Our experiences and recommendations may especially help researchers working in or with the Cooperative Extension system, given the need to collect reliable data from stakeholders, often with the support of industry organizations. Industry organizations publicizing a survey on their websites and social media may increase participation but threatens data quality, particularly when providing incentives. The results highlighted in this paper may inform any entities considering collecting data in online formats and offering participant incentives.

SURVEY OF U.S. BEEKEEPING OPERATIONS

In February–April 2021, three co-authors conducted an online survey and discrete choice experiment (DCE) of commercial beekeeping operations in the United States (U.S.). The survey evaluated beekeepers' preferences for different almond pollination contract attributes. Commercial beekeeping operations travel each year from across the U.S. to participate in the almond pollination market, which utilizes roughly 88% of the colonies in the U.S. (Goodrich & Durant, 2020). The U.S. beekeeping industry is highly concentrated, with most of the colonies operated by a small number of relatively large operations. The beekeeping industry consisted of 60,650 operations and 2.9 million colonies according to the 2017 USDA Census of Agriculture, beekeepers with more than 300 colonies made up 2% of operations but 83% of colonies. Goodrich et al. (2019) show that the average size of honey bee colony shipments into California for almond pollination was 394 colonies per truckload. Our target population was large commercial beekeeping operations, primarily those operating more than 300 colonies.

Anticipating the struggles with recruiting commercial beekeeping operations, we knew we needed a robust marketing strategy.[2] We offered a $20 Amazon gift card as a participation incentive. We acquired a list of approximately 250 emails from beekeeping directories enabling direct email requests to potential participants. Additionally, several beekeeping organizations announced the survey through their email listservs, websites, social media, etc. We intentionally timed the release of the survey to match almond pollination season, when beekeepers are relatively less busy once colonies are placed in the almond orchards, and marketed the survey heavily from February 15–March 15.

We received an abnormally large influx of responses roughly a week into the survey.[3] On February 22, Project *Apis m.* (PAm), a non-profit organization that funds honey bee health research, promoted our survey through an Instagram post on its public account. Soon after, responses flooded in, with over 1200 responses on February 22 alone, more than our estimated target population of 1168 beekeepers. It became obvious that most of these were fraudulent responses coming from malicious actors trying to gain participation incentives. Not all publicity had this impact. A website post by *Bee Culture Magazine* on February 19 resulted in only a small increase in responses. Both of these posts contained an anonymous link to the survey and a statement about the $20 Amazon gift card.

We tried several strategies to stop or identify the fraudulent responses but most attempts were ineffective. With responses still flooding in on the morning of February 23, we decided to shut down the original anonymous survey link. We sent out unique, password-protected links to beekeeping organizations to resend to their email lists, and asked them not to post on social media. We later received additional influxes of fraudulent responses when social media posts

were made that included the survey password. These later influxes were controlled more easily with the new, unique links identifying each source. At the end of the distribution, we received 2622 responses, of which we later determined 105 (4%) were legitimate.

SURVEY OF VIRGINIA FARMS AND AGRIBUSINESSES

In Fall 2020, two co-authors and several colleagues designed and implemented a survey to measure the impacts of the COVID-19 pandemic on Virginia farms and agribusinesses. We asked a series of questions designed to estimate the economic impacts of the pandemic on our target audience, including qualitative questions about the nature of disruptions, and their financial literacy and well-being. We planned to carry out three rounds of the survey to gauge the ongoing and evolving impacts of the pandemic.

The first round of the survey was available to respondents from September 21 to October 10, 2020. We announced the survey through several listservs, producer organizations, and on social media. We received 146 responses of which 76 were more than 80% complete, and we deemed this initial response rate lower than necessary for statistical analysis. After discussion with producer organizations and VCE administration, we decided to provide incentives to respondents by offering $10 Amazon gift cards to be randomly offered to 100 respondents among the first 500 valid survey responses for the second and third rounds of the survey.

The second round of the survey was available to respondents from March 15 to April 12, 2021. We recruited respondents through similar (and additional) channels. We announced the survey on Twitter and Facebook but did not post survey links to guard against potentially malicious actors attempting to gain access to gift cards.[4] We did not monitor the responses as they came in, but were pleased to see a much larger number of responses than in the first round; we had 444 total responses, of which 357 were more than 80% complete. However, upon further assessment, we determined 72% of these responses to be fraudulent. After some discussion, we decided not to roll out the planned third round of the survey because of the high incidence of fraud.

METHODS FOR FINDING FRAUDULENT RESPONSES

We used three broad categories to inspect our respective survey datasets to identify legitimate responses: respondent statistics, institutional knowledge, and inconsistencies in the data. These categories are described below. Like previous efforts (see e.g., Griffin et al., 2021; Pozzar et al., 2020; Teitcher et al., 2015), these are based on participants' responses to individual questions and/or metadata collected by Qualtrics.

Respondent statistics

Online survey software automatically generates metadata about the respondents including the IP address, geolocation, and time elapsed taking the survey. More detailed statistics, such as the duration of specific questions, can be programmed when creating the survey. These data can help to identify fraudulent responses. For example, a fraudulent respondent may take much less time than expected to complete the survey. However, fraudsters have likely adjusted their methods to avoid detection (Storozuk et al., 2020), including falsifying geolocation and IP

address data (Dennis et al., 2020). Similarly, some authentic responders may legitimately complete the survey quickly, and IP addresses can often be inaccurate.[5] For this reason, most respondent statistics provide suggestive evidence of falsification best used in tandem with other fraud-detection methods.

Institutional knowledge

Institutional knowledge tests utilize the information that is specific to the surveyed population to determine authenticity. For example, Zhang et al. (2022) targeted computer programmers, so intentionally designed survey questions to test the respondents' knowledge of a specific programming language. Neither of our surveys included institutional knowledge questions deliberately designed to catch fraudulent responders, but the beekeeping survey inherently allowed us to check if responses indicated a poor understanding of commercial beekeeping enterprises. The survey of Virginia farms and agribusinesses did not include questions that would reveal whether respondents knew the industry well.

Inconsistencies in data

Inconsistencies in data are commonly used to detect fraudulent responses (Griffin et al., 2021; Teitcher et al., 2015; Zhang et al., 2022). This type of test flags respondents who give inconsistent responses within the survey itself and works by identifying careless survey takers or bots that often use randomization to answer survey questions (Buchanan & Scofield, 2018; Dupuis et al., 2019; Meade & Craig, 2012; Stantcheva, 2022). For example, Griffin et al. find that some respondents reported working from home during the COVID-19 pandemic, yet later described themselves as essential worker unable to work from home. Inconsistencies can occur anywhere where the answer to a question is partially or fully related to an answer given earlier in the survey. Numerical or logical inconsistencies are the most straightforward and can be purposely built into the survey beforehand. Problematically, legitimate participants can also make mistakes due to inattention or misunderstanding (Teitcher et al., 2015).

Beekeeper survey

We pursued two complementary methods to detect fraudulent responses in the commercial beekeeper survey data: manual elimination of suspicious responses using our best judgment, and a methodological coding of tests and tally of suspicious activity for each response. Each of the approaches was cross-checked against one another to increase our confidence in classifying a response as fraudulent. The methodological coding method used many of the same standards for eliminating fraudulent responses as manual elimination but was less prone to human error.

Table 1 describes each test used to detect and remove fraudulent responses in the beekeeper survey data. Because of the volume of fraudulent responses, we used 20 tests divided into high and low-priority levels. High-priority tests are defined as those in which we expected nearly all real respondents to pass. For example, it is highly unlikely that a U.S. commercial beekeeper would take the survey from outside of the U.S. during almond pollination season, so we could confidently eliminate responses with location coordinates from outside of the U.S. Thus, any response that failed even one of these high-priority tests was flagged as fraudulent. Of the 2622

TABLE 1 Lists of tests for identifying fraudulent responses—Beekeeper survey.

Tests by category	Priority level	Description
Respondent statistics		
Outside continental US	High	Since our survey was open during a busy time in the pollination calendar, we expect commercial beekeepers to be near their colonies. This check screens out responses that do not fit this expectation because they come from outside the continental U.S.
IP address used 3 or more times	High	An IP address uniquely identifies a device on the internet. While these can be manipulated (Dennis et al., 2020), we do not expect the same IP address to be used multiple times by valid respondents.
Finish in <5 minutes	High	Pre-testing showed most respondents needed at least 20 min to complete the survey. Completing the survey in 5 min would be highly unusual.
Location used 3+ times	Low	The location is given by a set of latitude/longitude coordinates. We do not expect multiple beekeepers to respond from the same exact location. We allow room for error due to issues of precision and quality in location data (Dennis et al., 2020).
Finish in <10 min	Low	The expected duration of the survey was over half an hour. Completing the survey in less than 10 min is suspicious.
Compromised link	Low	Unique survey links were provided to individuals and organizations. Responses using links known to have been posted publicly online and noticed by bots are flagged.
Institutional knowledge		
Provide <50 colonies in multiple regions	High	There are considerable economies of scale in commercial beekeeping (Goodrich et al., 2019). The average number of colonies per shipment to California between 2008 and 2018 was 394, with little variation around the mean. Placing considerably less than a full shipment and spacing these low numbers across regions, is unexpected.
Provide a forage discount of >$20 per colony	High	This question asked about discounts provided by the beekeeper to the grower during the 2021 pollination season in exchange for planting bee-friendly forage. This type of discount is not common, and to the best of our knowledge, current instances of this do not exceed $5 per colony.
<50 colonies in a region	Low	There are considerable economies of scale in commercial beekeeping. Placing less than a full shipment of colonies (<400) is therefore undesirable from a logistical and financial perspective.
Colonies in multiple regions	Low	There are considerable economies of scale in commercial beekeeping. Spacing colonies across multiple regions is therefore undesirable from a logistical and financial perspective.
Colonies not multiple of 4 or 6	Low	Commercial beekeepers move honey bee hives around using pallets and forklifts. There are two standardized pallet setups, either 4 or 6 hives per pallet.

(Continues)

TABLE 1 (Continued)

Tests by category	Priority level	Description
Forage discount ≥ $5	Low	This question asked about discounts provided by the beekeeper to the grower during the 2021 pollination season in exchange for planting bee-friendly forage. This practice is not yet common, and to the best of our knowledge, current instances of this do not exceed $5 per colony.
Inconsistencies		
No colonies provided in any region	High	Respondents indicate that they supplied colonies for the 2021 almond pollination season. When asked how many colonies were supplied by region, they indicate no colonies were supplied in all regions.
Supply bees for longer than kept bees	High	Respondents indicate that they have supplied honey bee colonies for almond pollination for longer than they have been keeping bees.
Work with the grower for longer than have kept bees	High	Respondents indicate that they have provided pollination services for their current almond grower for longer than they have been keeping bees.
Base pollination fee > highest possible fee	High	Respondents indicate that the lowest possible fee is higher than the highest possible fee for a pollination contract where the fee is determined by performance.
Lowest advance payment > highest	High	Respondents indicate that the lowest percentage of payment received in advance on any contract during the last pollination season is higher than the highest percentage of payment received in advance on any contract.
Base pollination fee = highest possible fee	Low	Respondents indicate that the lowest possible fee is equal to the highest possible fee for a pollination contract where the fee is determined by performance.
Want more cover than the area available	Low	Respondents indicate that they would like a higher percentage of the orchard planted in cover crops than is available. The question states that the area between rows, where forage may be planted, makes up only 50% of the orchard area.
Leave and do not leave CA	Low	Respondents indicate that most of their colonies do not leave California after almond bloom, and in a later question indicate the opposite.

Note: High-priority tests flagged responses likely to be fraudulent. Low-priority tests flagged suspicious behavior that warrants further investigation.

responses received, 2155 were flagged as very likely to be fraudulent after failing at least one high-priority question.[6]

We further subjected the 478 responses that passed the high-priority tests to low-priority tests. Low-priority tests indicated suspicious behavior, but could also have flagged real respondents. For example, if three or more responses came from the exact same location, these would be flagged as failing a low-priority test. While indicative of fraud, we also know that many beekeepers are within California during the almond bloom and could be taking the survey from the same city or hotel. After summing the total number of low-priority flags for each response, we looked at the responses individually. The more low-priority flags, the more likely a response

TABLE 2 Results of high-priority tests for fraudulent responses—Beekeeper survey.

	Flagged responses	Percentage of Total responses flagged ($N = 2155$)	Responses flagged only by this test	Percentage flagged only by this test
Respondent statistics				
Outside continental US	74	3%	13	18%
IP address used three or more times	118	5%	4	3%
Finish in <5 min	51	2%	13	25%
Institutional knowledge				
Provide <50 colonies in multiple regions	1884	87%	883	47%
Provide a forage discount >$20	418	19%	33	8%
Inconsistencies				
Do not provide colonies	50	2%	31	62%
Supply bees for longer than kept bees	349	16%	26	7%
Work with the grower for longer than have kept bees	35	2%	9	26%
Base pollination fee > highest possible fee	410	19%	36	9%
Lowest advance payment > highest advance payment	315	15%	46	15%

Note: High-priority tests flagged responses likely to be fraudulent.

is fraudulent. We decided to keep or eliminate each response based on its number of low-priority flags, as well as subjective and less easily coded information such as short answer responses that show poor understanding of the industry. In a few uncertain cases, we searched online for the email address the respondent provided to see if it was associated with a beekeeping operation. For any emails we could not verify, we reached out using email, asked the respondents three simple questions from the survey (age, number of years beekeeping, etc.), and cross-checked the answers provided via email with their responses to the survey questions. Among email responses received, inconsistencies showed that some of the associated survey responses were fraudulent.

Tables 2, 3 show the results of the high and low-priority tests. Institutional knowledge most successfully detected fraudulent responses. We were able to reject 87% of the total sample by excluding responses who reported supplying unreasonably small numbers of colonies (<50) across multiple regions, an action which would be logistically and financially difficult for a real commercial beekeeper.

High-priority flags for inconsistencies, that is, tests of numerical responses that should follow an expected order, also performed well, flagging up to 15%–19% of respondents each. These responses commonly failed other high-priority tests as well, confirming that other high-priority flags were accurately identifying fraudulent respondents. High-priority tests using respondent statistics (e.g., IP address, time spent taking the survey, location) were relatively ineffective, catching 2%–5% of the total fraudulent responses. This matches other research that malicious actors use sophisticated techniques to conceal their identities and should not be used exclusively (Dennis et al., 2020; Storozuk et al., 2020).

TABLE 3 Results of low-priority tests for fraudulent responses (478 responses remaining after high-priority elimination)—Beekeeper survey.

	Flagged responses	Flagged responses as a percentage of remaining ($N = 478$)	Legitimate responses	Legitimate responses as a percentage of flagged responses
Respondent statistics				
Location used 3+ times	83	17%	32	39%
Finish in <10 min	127	27%	14	11%
Compromised link	144	30%	12	8%
Institutional knowledge				
<50 colonies in a region	166	35%	5	3%
Colonies in multiple regions	192	40%	19	10%
Colonies not multiple of 4 or 6	262	55%	15	6%
Forage discount ≥ $5	80	17%	3	4%
Inconsistencies				
Base pollination fee = highest possible fee	2	0.4%	1	50%
Want more cover than the area available	195	41%	9	5%
Leave and do not leave CA	43	9%	1	2%

Note: Low-priority tests flagged suspicious behavior that warrants further investigation.

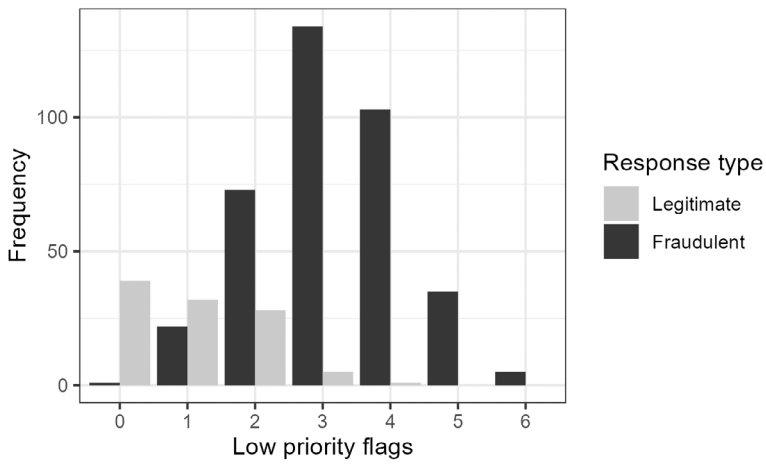

FIGURE 1 Histogram of the number of low-priority flags for each response (478 responses remaining after high-priority elimination)—Beekeeper survey.

Figure 1 shows a histogram of the total number of low-priority flags (maximum of 10) for the 478 responses that passed high-priority tests separated by response type, fraudulent (373), and legitimate (105). Most legitimate responses were only flagged for suspicious behavior two or fewer times.

Similar to the findings of the high-priority tests, low-priority tests associated with institutional knowledge were most helpful in identifying fraudulent responses (Table 3). Of the responses flagged for further investigation in this category, a relatively small number made it into the final dataset. The fraudulent responders gave unconvincing answers with respect to the number of colonies or the size of a discount offered to growers for planting bee-friendly forage. Only some of the low-priority tests for the categories of inconsistencies and respondent statistics were useful. Importantly, tests using respondent statistics flagged a relatively high number of responses as suspicious, but many of these were eventually classified as legitimate, coinciding with our caution about its exclusive use. Fewer responses failed consistency tests, but those that did rarely passed further inspection.

Virginia farm survey

As we discovered unusual patterns in the Virginia farm and agribusiness survey data, we noted the patterns, flagged responses as fraudulent, and eliminated them from the data set we would use for analysis. Figure 2 diagrams the process we went through to filter out fraudulent responses. We primarily filtered responses using respondent statistics rather than institutional knowledge or inconsistencies in data. The first round of the survey had several questions to evaluate inconsistencies in reported business data, but we eliminated such questions in the second round to reduce the expected completion time and respondent fatigue.

We received 444 total responses, with about 25% of responses initiated within 72 h of the survey launch, and 82% within the first 8 days. After this, responses slowed to an average of four per day for the duration of the survey. Our first indication that our survey was targeted by fraudulent responders was upon observing a large number of responses ($N = 187$) had given zip

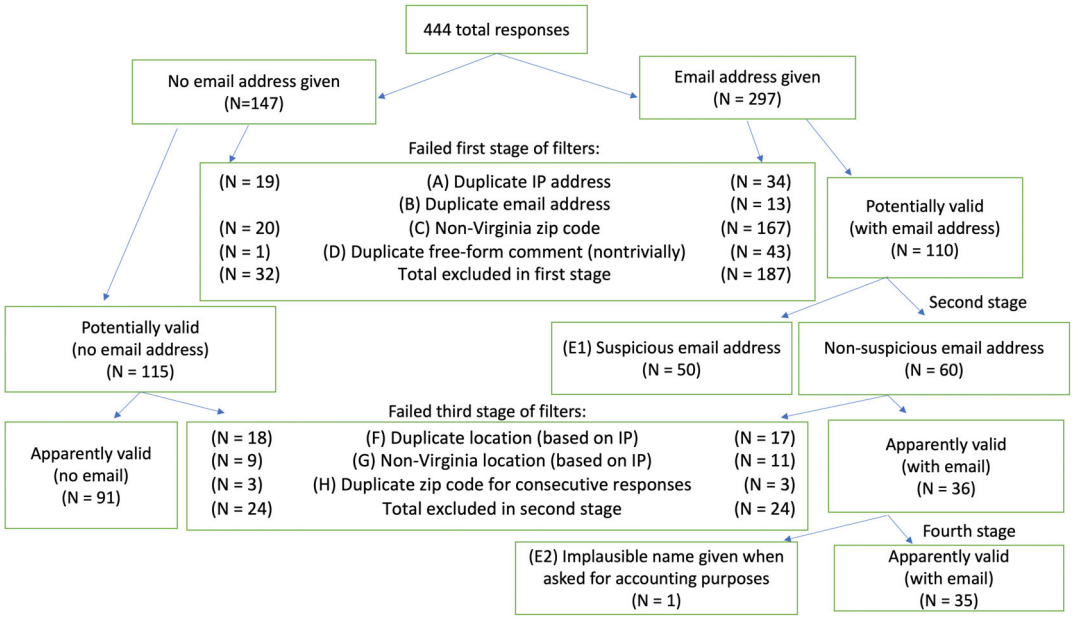

FIGURE 2 Iterative process to eliminate responses from Virginia data set.

codes outside Virginia. As seen in Figure 2, we first disqualified all responses that included any of the following: (A) duplicate IP addresses, (B) duplicate email addresses, (C) zip codes outside Virginia, or (D) duplicate answers to the free-form response question (excluding trivially identical answers such as "No"). These first-stage filters disqualified 219 responses. Respondents' errors might have led to some surveys being misclassified as invalid, and our elimination procedures could be considered overly aggressive. However, we believe these decisions best ensured the highest quality data for our analysis.

Next, we realized that many of the email addresses—particularly for the responses with non-Virginia zip codes—followed unusual patterns, as observed previously (Griffin et al., 2021; Storozuk et al., 2020). The patterns in our survey were email addresses using Gmail, Yahoo, or Outlook with one or two "names" followed by a string of random letters or numbers (or both), or simply a string of random letters and/or numbers.[7] Among the 110 responses with email addresses that survived the first filter, we deemed 50 to have invalid email addresses.[8]

At this point, we had 175 potentially valid responses, with and without email addresses (Figure 2). Of these, 48 had either (F) duplicate latitude and longitude as other responses in the survey, (G) IP address locations outside Virginia, or (H) entered the same zip code as the response immediately before or after it (conditional on using the same survey link). We disqualified all 48 of these responses to ensure that we included only responses that were nearly certain to be from actual Virginia farms and agribusinesses. Over half of the responses (236) came from IP address locations outside Virginia.[9]

At a late stage, we were informed that the Virginia Tech accounting office required a record of each gift card recipient's name. We emailed each of the 36 valid-with-email respondents to retrieve names. We received 24 responses, one of which we deemed unacceptable because—similar to the email addresses we had earlier identified as fraudulent—it included three random consonants at the end of an otherwise normal name. We assume that the remaining 12 respondents simply did not want to share their names but were valid survey responses. We ended up with 126 valid responses, of which 35 had email addresses and 67 were complete.

After completing the iterative process of eliminating responses that failed a single criterion, we evaluated the reliability of the method we employed by tabulating the number of flags for responses that failed a given criterion. Some of the responses that we flagged as fraudulent may have been the result of an honest mistake on the part of respondents (e.g., a typo in the zip code) or technical issues (e.g., IP addresses being identified by the software as being in the wrong state). Table 4 summarizes the number of flags raised for violators of each criterion. The labels on the left side of the table show the criteria, and the remaining columns show the number of flags for each response that violates the criterion names on the left. Based on the patterns of violations and the incidence of multiple violations, we are confident that we have correctly classified nearly all of the responses.

As seen in Table 4, 59 responses (19% of flagged responses) violated a single criterion. Of these, 68% were either duplicate IP addresses (A), gave a non-Virginia zip code (C), or had duplicate geographic coordinates as another response (F). Violations of these three criteria—along with duplicate email addresses (B) and duplicate comments (D)—we considered unambiguous disqualifiers. The remaining 81% of flagged responses had two or more flags. 82% of those with suspicious email addresses or names (E), 78% of those with non-Virginia IP addresses (G), and 78% of consecutive responses that gave the same zip code (H) violated three or more criteria. Some of the 19 responses with a single violation of either criterion E, G, or H may be valid, but these make up only 6% of all flagged responses.

TABLE 4 Number of flagged violations by criterion—Virginia survey.

Criterion	Number of flagged violations						Total violations (sum across columns)
	1	2	3	4	5	6	
(A) Duplicate IP address	9	8	10	2	18	6	53
(B) Duplicate email address	0	0	2	4	4	3	13
(C) Non-Virginia zip code	9	12	96	40	23	7	187
(D) Duplicate free-form comment (non-trivially)	0	3	3	26	7	5	44
(E) Suspicious email address or implausible name	6	37	116	43	25	7	234
(F) Duplicate (IP) location	22	16	17	15	20	6	96
(G) Non-Virginia (IP) location	10	43	110	41	25	7	236
(H) Duplicate zip code for consecutive responses	3	3	12	5	3	1	27
Total number of responses by number of flagged violations	59	61	122	44	25	7	318

Note: Columns show the number of responses by the number of flagged violations that violate the criterion names on the left.

DATA INTEGRITY ISSUES CAUSED BY FRAUDULENT RESPONSES

After identifying fraudulent responses, we can examine their effect on data integrity. Table 5 displays the results of Wilcoxon rank-sum tests for key economic variables in each of our surveys to determine whether the fraudulent and valid responses come from distinct populations. In eight out of ten cases, we are able to reject the null hypothesis that the population distributions for valid and fraudulent responses are identical. In particular, the statistical significance for open-ended questions (number of bee colonies) is stronger than for close-ended questions (pollination fee). While open-ended questions can help identify fraudulent responses, they also contribute to fatigue among legitimate respondents. Table 5 shows that fraudulent responses can substantially change the distributions of survey responses if not detected and removed from the dataset.

Results from the Virginia survey also suggest that fraudulent respondents were more likely to answer multiple-choice questions randomly, consistent with Buchanan and Scofield (2018) and Dupuis et al. (2019). To be specific, we asked five multiple-choice questions to gauge financial literacy. We found that valid responses were more likely to contain four or five correct answers than fraudulent ones (43% vs. 27%), but also that valid surveys were more likely to contain zero correct answers (21% vs. 11%). On the other hand, fraudulent responses were more likely to contain 1–3 correct answers, as would be expected from randomly answering multiple-choice questions. Using Wilcoxon rank-sum tests, these differences were statistically significant at the 5% level. Similarly, the fraudulent responders seem to have chosen their answers about their race and ethnicity randomly. According to the 2017 Census of Agriculture, 96% of Virginia farms have a "White" principal producer, and 1% have a "Hispanic, Latino, or Spanish" principal producer. Of the 60 valid responses that included answers to these questions, 3% indicated

TABLE 5 Results from Wilcoxon rank sum tests of valid versus fraudulent distributions for select economic variables-Beekeeper and Virginia surveys.

Variable	Response type	*N*	Mean	St. dev	Wilcoxon rank sum test *p*-value
Beekeeper survey					
Number of colonies rented out for almond pollination[a]	Valid	102	5729	15,082	0.00
	Fraudulent	2515	493	21,136	
Pollination fee for largest pollination agreement[b]	Valid	94	188	23	0.02
	Fraudulent	2489	194	32	
Virginia farm and agribusiness survey					
Change in revenue, 2019–20 (%)	Valid	66	−11.8	41.4	0.10
	Fraudulent	279	−5.0	28.2	
Change in revenue, 2019–20 ($)	Valid	51	63,521	422,987	0.38
	Fraudulent	205	−31,135	464,001	
Change in revenue, 2020–21 (%)	Valid	66	3.8	19.1	0.00
	Fraudulent	274	13.2	26.0	
Change in revenue, 2020–21 ($)	Valid	56	41,182	149,499	0.34
	Fraudulent	181	32,102	173,219	
Change in expenses, 2019–20 (%)	Valid	60	15.3	27.5	0.00
	Fraudulent	273	282.7	3260.3	
Change in expenses, 2019–20 ($)	Valid	47	43,915	218,750	0.03
	Fraudulent	208	−5992	206,034	
Change in expenses, 2020–21 (%)	Valid	60	8.5	13.9	0.00
	Fraudulent	270	−0.9	23.9	
Change in expenses, 2020–21 ($)	Valid	43	66,538	282,925	0.00
	Fraudulent	163	6695	89,703	

[a]Number of colonies was free entry format. The observation is omitted if the question was left unanswered, but included if 0 was entered.

[b]Pollination fee was entered using a slider, with values every $5 from $100 to 300. The base pollination fee is used when respondents indicate a per-frame bonus contract. Observations that skipped this question were omitted.

that they were a race other than "White," or multiple races, roughly in line with the Census of Agriculture demographics. Of the 292 fraudulent responses that were answered, 74% indicated a race other than "White," or multiple races.

Additionally, we found that DCE results in the beekeeper study would be substantially affected by including fraudulent respondents. Including controls for fraudulent responses leads to significantly improved model fit. Importantly, fraudulent respondents have a negative price coefficient, that is, prefer less money in their pollination contracts all else equal. Thus, fraudulent responders in survey data can lead to inaccuracies in the results of economic analyses.

DISCUSSION AND RECOMMENDATIONS FOR DATA QUALITY CHECKS AFTER SURVEY DISTRIBUTION

By comparing the outcomes of our respective surveys, we largely agree with Zhang et al. (2022) that there is "no perfect strategy for preventing and detecting invalid respondents for online surveys." We recommend using a combination of strategies that will depend on the context of the survey, its targeted population, and the type of incentives used. As outlined above, our two experiences yielded more differences than similarities. The beekeeper survey may have attracted more sophisticated bot programming than the Virginia survey due to a higher guaranteed payout of $20 versus an uncertain $10 payout using a lottery system.

Respondent statistics proved less useful in the beekeeper survey than in the Virginia farm survey. Each of the high-priority respondent statistics flagged only 2%–5% of the beekeeper survey responses, while the three respondent statistics used to identify fraudulent responses in the Virginia survey flagged between 12% and 53% of all responses. One of the key differences was the geographic scope of the two surveys; while the beekeeper survey was intended to collect responses from anywhere within the contiguous U.S., the Virginia survey was restricted to Virginia farms.

Additionally, our two surveys saw differences in the timing and frequency of fraudulent responses. In the Virginia survey, most fraudulent responses occurred at night (between 8 pm and 7 am local time), while most valid responses were initiated during the day. However, this pattern was not present in the beekeeper survey, with fraudulent responses appearing frequently throughout the day until we took action to curtail them. In the Virginia survey, the average number of responses was high at first and then dwindled, even though no actions were taken to combat the fraudulent respondents.[10]

Based on our experience with the beekeeper survey, it is recommended to include institutional knowledge-based questions to ensure that the target audience is being reached. An overwhelming 87% of responses to the beekeeper survey failed a single institutional knowledge question. The Virginia survey, on the other hand, included only one question that could be regarded as institutional knowledge (zip code) but plausible answers to the question could have been determined easily and used by bot programmers. The last point raises questions: How high would payments need to be for fraudulent responders to take time to learn about an industry? And will programmers soon become more sophisticated to deceive researchers, for example, by using machine learning to gain knowledge about industries and institutions?

Our experiences also suggest that bots and their programmers have improved strategies over time (Storozuk et al., 2020). For example, Teitcher et al. (2015) recommended checking the duration of the responses to identify fraudulent responses, though we found only subtle differences in response times of fraudulent and non-fraudulent responses.[11] Relatedly, in both of our cases, we tried to verify the legitimacy of responses by emailing participants using the provided address, and we each received multiple responses that clearly indicated they were fraudulent. This indicates at least some level of human intervention along with programmed bot responses. Thus, as researchers become more skilled at detecting fraudulent responses, it is almost certain that the fraudulent responders will continue to improve methods to evade detection.

RECOMMENDATIONS FOR MITIGATING FRAUDULENT RESPONSES PRIOR TO SURVEY DISTRIBUTION

Fraudulent responses were a problem in each of our cases because incentives were offered for participation and because of our convenience sampling via an anonymous survey link. We

recommend thinking about the potential of fraudulent responses *prior* to distributing a survey that pays participation incentives.

We learned (albeit too late) that the institutional Qualtrics account used to conduct the beekeeping survey did not include enhanced settings for fraud protection. However, Griffin et al. (2021) used enhanced protections in Qualtrics, and more than half of their responses were determined to be fraudulent. Thus, relying on survey software developers is insufficient to guarantee data integrity.

Share on social media (and other websites) with caution

Sharing on social media and posting on other websites (e.g., local government or media websites) seemed a primary cause of fraudulent responses, though fraudsters can enter an anonymous survey in other ways. Yet social media may be the most cost-effective approach to sampling hard-to-reach populations (Gao et al., 2016; Ince et al., 2014; Loxton et al., 2015). For example, in the beekeeper survey, 64% of legitimate responses came from anonymous links, versus 36% from direct emails to beekeepers. In the Virginia survey, although we did not share a link on social media, 61% of valid responses and 91% of invalid responses came from the link that was shared most widely, including on some publicly accessible websites.

Further, not all social media and websites are equally problematic. For example, the beekeeper survey was shared on a private Facebook group page and that link was never inundated by fake responses. Rommel et al. (2022) had a similar experience soliciting a convenience sample of farmers in multiple countries via an anonymous link marketed through farmer networks. In some countries this approach was successful, however, in Scotland, their links were flooded with bot responses. They switched to direct emails to verified farmers in Scotland, but ultimately, were not able to collect enough usable observations for analysis.

Do not automate participant incentive payments

Most universities have many levels of bureaucracy that prevent automated incentive payments. In the small chance that it is possible to automate payments, this should not be done if there is any risk of fraudulent responses. For example, in the beekeeper survey, nearly every fraudulent response contained an email address, as that was required for payment. If the beekeeping survey incentives had been automated, fraudulent respondents would have received over $35,000 during the two worst days of fraudulent responses.

Consider the type of survey incentives

Prior work suggests that offering incentives through a lottery as opposed to guaranteed individual payments might help deter fraudulent responses (Griffin et al., 2021; Kramer et al., 2014; Teitcher et al., 2015). While the lottery format may have deterred some fraudulent responses in the Virginia farm survey, fraudulent responses were still a problem.

Charitable donations could be promising as they can potentially equally attract participation among non-fraudulent responses (Penn & Hu, 2022), but break the incentive for both *unique participant* and *alias fraud* because there is no direct payment to the survey respondent.[12] The

beekeeper survey provides evidence in favor of this. At the conclusion of the main survey, participants were asked if they would answer additional questions for an additional incentive. One of two randomized incentives was offered: a $10 donation to a beekeeping non-profit or a $10 addition to their Amazon gift card. If the respondent rejected the first offer, the alternate option was presented. Of the suspected bots who saw the donation as an incentive first, 40% agreed to answer the optional questions versus 70% of real respondents. Notably, 87% of suspected bots (and 91% of real respondents) agreed to optional questions when offered the gift card incentive first. When shown the alternate option, 80% of suspected bots who had rejected the donation accepted a payment while only 9% of bots who had rejected a payment accepted a donation. Real respondents accepted the second offer around 50% of the time in both cases. The fraudulent responders were clearly driven toward direct payment.

Enable password protection for the survey

Using a password can help limit the number of fraudulent responses. However, asking participants to enter a password also increases respondent burden, potentially increasing non-response rates (Crawford et al., 2001). Also, the password-protected survey only works to deter fraudulent responses as long as the password is not shared publicly.

Use a unique link for each anonymous source

For the beekeeper survey, we used Query Strings in Qualtrics to create a new variable, "Source."[13] This allowed us to assign a unique value to each anonymous source link, for example, the anonymous link sent to Project *Apis m*. ended with ?Source = "PAM" or the link sent to American Honey Producer's Associated ended with ?Source = "AHPA." If those links became flooded with fraudulent responses, this allowed us to shut down individual links or direct those corrupted links through additional authentications.

Use an additional authenticator

Qualtrics has options for different authenticators to be embedded in the survey flow, this includes Single Sign-On (SSO) authenticators through Facebook, Google, and other systems.[14] For the beekeeper survey, we sent links that we suspected were corrupted with fraudulent responses through additional Facebook authentication.[15] This approach was successful in deterring fraudulent responses, though it also potentially discouraged actual respondents.

In addition, Kennedy et al. (2020) provide a tool that can be used to prevent respondents with foreign IP addresses or who are using virtual private servers from completing surveys. This tool can be embedded within Qualtrics.

Use bot-identifying questions

A number of bot-identifying questions can be incorporated into the survey instrument prior to distribution. This facilitates later identification and elimination of fraudulent responses from

the dataset. Some broad categories of frequently used bot-identifying questions are outlined here, but a further dive into the cited literature may reveal additional and more effective approaches. It is also important to note that bot programmers have a large incentive to continually improve upon their bot programming so that they will bypass detection, thus many of these strategies have already become outdated or may become outdated in the relatively near future.

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) questions are security checks with the specific purpose of identifying fraudulent bot responses. Many researchers have recommended the use of CAPTCHA as an effective way to recognize bots (e.g., Charité et al., 2022; DellaVigna & Pope, 2022; Fisman et al., 2020; Stantcheva, 2022; te Velde & Louis, 2022; Teitcher et al., 2015), however, additional recent work has shown that these are no longer effective as bots can be programmed to recognize and bypass these questions (Al-Fannah, 2017; Griffin et al., 2021; Sivakorn et al., 2016; Storozuk et al., 2020). Our experience with the beekeeping survey supports this. Qualtrics' reCAPTCHA question only reduced the rate of fraudulent responses temporarily, resuming previous rates after *only a few minutes*. Zhang et al. (2022) note that the hybridization of automated bot responses and human interventions can allow fraudsters to get around these types of questions.

Attention check questions request that the respondent take a specific action, for example, leave a question unanswered or choose a specific answer option. These have been used to detect inattentive survey-takers (e.g., Blesse & Heinemann, 2020; Cheng et al., 2022; DellaVigna & Pope, 2022; Gao et al., 2016; Lennon et al., 2023; Malone & Lusk, 2018a, 2018b, 2019; Snowberg & Yariv, 2021), but Storozuk et al. (2020) and Zhang et al. (2022) find these types of questions are moderately effective in identifying bots. Zhang et al. (2022) displayed the instructions for attention-check questions in image form to avoid bots that use Natural Language Processing. The authors also note that attentive survey-takers can answer these questions incorrectly, so to combine it with other strategies. In addition, Stantcheva (2022) notes that if inattentive responses are excluded, this could pose threats to external validity (after all, some real respondents are inattentive some of the time).

Open-ended questions can be useful in identifying bots, however as stated previously bots can use Natural Language Processing to fill in open-ended survey questions. Griffin et al. (2021) noted that many fraudulent responses entered exactly identical open-ended responses, so this can be a way to eliminate fraudulent responses. In each of our surveys, we received exact duplicate responses to the open-ended questions. In addition, it may be possible to identify bots by reading individual answers to the open-ended questions and assessing whether the answers make sense, read naturally, and were likely to be written by humans, although this requires researchers to make subjective assessments and is subject to error (Fisman et al., 2020).

"Honey pot questions" are questions that are hidden from human survey takers but visible to bots (Storozuk et al., 2020). Hidden questions can be programmed into surveys using JavaScript or other programming languages, so if a hidden question is answered, it is certainly a bot response.[16] As bots have become more sophisticated the effectiveness of honey pot questions has declined. Only 16% of fraudulent responses to a survey conducted by Pozzar et al. (2020) answered hidden questions, and similarly, Storozuk et al. (2020) find honey pot questions as one of the least effective methods to identify fraudulent responses in their dataset. While honey pot questions may not be very effective, these questions are easy to implement and do not add to the survey taker's burden, so it may be worth including them to assist in fraud identification.

As described previously, checking institutional knowledge and consistency across questions can effectively identify fraudulent responses (Griffin et al., 2021; Teitcher et al., 2015; Zhang

et al., 2022). Purposely building in such questions, such as asking for a respondent's age at two different points of the survey, or asking the same question but phrased differently (see e.g., Zhang et al., 2022), can be an easy way to identify bot responses. Checking for institutional knowledge (a question that only the target population can answer correctly) may be difficult to implement in general-public surveys, requiring more creativity from the researcher. For example, asking for the participant's zip code and then later asking them to name a nearby location (e.g., closest university, adjacent county, etc.) that can be verified. For a consumer study, asking for the number of people in the household and the average weekly grocery bill might illuminate a lack of familiarity with the cost of living in the target country or region. This extra effort in general contexts can help ensure the integrity of their online survey data.

DISCUSSION AND CONCLUSIONS

Evidence of the pervasiveness of fraudulent responses has rapidly expanded, undermining the credibility of online surveys to provide reliable data. These widespread issues may affect a wide array of applied economists who use online survey data. Though still nascent, research showcases the usefulness of fraud detection methods to improve data quality, but with mixed evidence of each technique's efficacy. Like others, we find that multiple methods are critical to detecting fraudulent responses.

Additional research to understand and ultimately reduce the incidence of fraud would also be valuable. In our case studies, the number of fraudulent responses was much higher when we guaranteed payment of $20 to commercial beekeepers compared with the lottery to potentially win $10 for Virginia farmers. We speculate that this is related to the diluted participation incentive of the latter. If incentive structures affect fraudulent attacks, then other incentives may be preferred, or perhaps a researcher may decide the risks associated with incentives are too great and forego incentives entirely. Alternatively, certain marketing techniques may be less prone to fraudulent responses such as posting to private social media groups or not advertising the incentive when posting on social media.

Bot programmers can quickly adapt and make the best practices irrelevant, as is already documented in the literature (Storozuk et al., 2020). Consequently, traditional recruitment methods, such as mailing surveys via address-based sampling, may become increasingly attractive. Mailing can take advantage of online surveys using push-to-web techniques, but still necessitates a credible population frame that may not exist or may be expensive to assemble. Bots also increase the potential attractiveness of probability-based samples, such as IPSOS' Knowledge Panel (previously owned by GfK) or Prolific, but which are much more expensive than non-probability-based samples, for example, MTurk, Qualtrics, Dynata. While such probability-based panels can provide a representative sample, this approach may be infeasible for agricultural producers or other hard-to-reach populations.

Lastly, while preventing and detecting fraudulent responses is central to this paper, we should remain mindful of the actual humans whom we hope to gather information from. Increasing the methods or barriers to detect or reduce fraudulent responses may come at the expense of valid respondents. More open-ended, institutional knowledge or consistency check questions increase respondent fatigue. Further, some strategies, such as asking for a physical mailing address to mail payments, may discourage participation or be blocked by institutional review boards out of concern for privacy or respondent protections (Teitcher et al., 2015). Solutions must be context dependent to find the right balance that provides credible survey data.

ACKNOWLEDGMENTS

The authors are grateful to Madhu Khanna, past president of the Agricultural and Applied Economics Association (AAEA), who supported the Organized Symposium at the 2022 AAEA annual meeting that led to this collaboration. Goodrich, Fenton, and Penn thank the following funding entities that supported the beekeeper survey: Project Apis m., National Honey Board, and the University of California Giannini Foundation of Agricultural Economics. Goodrich, Fenton, and Penn are also thankful to the numerous beekeeping organizations and individuals that helped advertise to increase participation in our survey (and re-advertise once the bots infiltrated) (listed in no particular order): American Honey Producer's Association, California State Beekeeper's Association, Chris Hiatt, Anne Marie Fauvel with Bee Informed Partnership, Jerry Hayes with Bee Culture magazine, Kim Flottum, Jeff Ott with Beekeeping Today podcast, American Bee Journal magazine, Denise Qualls, and Ryan Cosyns. Bovay and Mountain are extremely grateful to their collaborators, Conaway Haskins, Catherine Larochelle, French Price, Nicole Shuman, and Jonathan van Senten, for their work in developing the first and second rounds of the survey of Virginia farms and agribusinesses, and to Dan Goerlich, Rebekah Slabach, and many other Virginia Cooperative Extension colleagues, Tony Banks of Virginia Farm Bureau Federation, Hobey Bauhan of Virginia Poultry Federation, Katie Frazier of Farm Credit of the Virginias, Mary Howell of Virginia Cooperative Council, Kim Hutchinson of Virginia Farmers Market Association, Kyle Shreve and Sarah Jane Thomsen, formerly of Virginia Agribusiness Council, and anyone else authors may have forgotten, for helping us to distribute and increase participation in the surveys. The authors also thank Virginia Tech's Beth Chang, Jennifer Friedel, Crysti Hopkins, Patrick Kayser, Joyce Latimer, Scott Tate, and some of the people mentioned above, for helpful feedback while developing the survey. The authors take responsibility for all errors and omissions.

ENDNOTES

[1] For example, Johansson et al. (2017), Reist et al. (2019), Weber and Clay (2013), and National Research Council (2008) discuss falling response rates to U.S. Department of Agriculture (USDA) surveys, while Czajka and Beyler (2016) and Meyer et al. (2015) document falling response rates to many federal household surveys.

[2] Bee Informed Partnership (BIP) has conducted nationwide surveys of beekeepers for over 15 years and has seen steady declines in response rates from large beekeepers. This survey is highly publicized and results are often cited by the beekeeping industry, yet in 2022 BIP received only 135 responses from beekeepers that operate 50 or more colonies. Similarly, other pollination fee surveys, for example, California State Beekeeper's Pollination Survey and Pacific Northwest Pollination Survey, consistently receive fewer than 50 responses per year.

[3] Table A1 in the Appendix displays a timeline of responses per day and notable distribution events for the first 10 days of the survey.

[4] The Twitter posts directed viewers to contact researchers through Twitter; the Facebook posts directed viewers to one researcher's faculty profile page with contact information. We did not receive any inquiries through Twitter or Email. Also, we asked VCE's communications team to use a specific survey link in responding to any social media inquiries, and no surveys were submitted using that link.

[5] Saxon and Feamster (2022) analyze the accuracy of IP-based geolocation databases and show that fixed-line IP address geolocations are more accurate than mobile IP address geolocations, but that even in New York City, fixed-line geolocations are only accurate within 2.6 km. Also, some respondents might regularly use virtual private networks or virtual private servers to disguise their IP address location. See additional discussion about virtual private servers in Dennis et al. (2020) and Kennedy et al. (2020).

[6] Only 11 of the 2155 rejected responses, or half of 1%, appear to have been rejected in error. This was determined through cross-checking with the manual elimination. Of these 1 responses, nine had completed only

4%–16% of the survey and were rejected due to short completion times. While these were added back to the final tally, these responses are omitted from most analyses due to missing data. The final two appear, to our best judgment, to have been flagged due to honest mistakes leading to inconsistencies on the behalf of legitimate beekeepers. We were able to verify the beekeepers' legitimacy using online searches for their provided emails. Both are reinstated in the final dataset used in analyses. All 11 responses were added back into the dataset to be subjected to the low-priority tests.

[7] Nearly all of the responses we identified as having unusual email addresses also violated other criteria that we deemed necessary for the responses to be considered valid. In addition, there were some striking patterns within the email addresses that cannot be mere coincidence. For example, 69 responses had email addresses of the form firstlastxxx@yahoo.com, where the names are capitalized and xxx is exactly three random lowercase letters (similar to patterns found by Griffin et al. (2021) and Storozuk et al. (2020)). All 50 responses with email addresses consisting of random lowercase letters at outlook.com either had missing or non-Virginia zip codes. The suspicious responses that used Gmail addresses followed a few different patterns.

[8] Later, we assessed the validity of the remaining email addresses of surveys that were eliminated by our first stage filters. Overall, we believe that 231 of the 297 email addresses (including duplicates) were suspicious based on the criteria we describe above. A handful of other email addresses were questionable but did not fit the patterns described above.

[9] Figure A1 in Appendix shows a map of all IP address locations.

[10] On the eighth day of the Virginia farmer survey, we had 129 responses—the most of any day of the survey. Of these, we later identified 125 to be fraudulent. The next 6 days, we received only 14 fraudulent responses. We are unsure why responses slowed; perhaps fraudulent responses were slowed by a Qualtrics filter or mechanism, or strategically by the malicious actor(s) to decrease suspicion.

[11] In the Virginia farm survey, fraudulent respondents took roughly the same amount of time on average to complete surveys (13 m 31 s) as valid respondents (13 m 53 s). In the beekeeper survey, the average response time for fraudulent responses (21 m 28 s) was less than that of valid respondents (35 m 45 s). In both cases, Wilcoxon rank-sum test confirms that the two samples have different distributions. Virginia participants were informed in recruiting materials that the survey "should take no more than 20 min" and beekeeper survey participants were informed the survey would take "approximately 20 min." It is likely fraudulent responders used this information to emulate response times of legitimate responders.

[12] An additional benefit is that using charitable donations avoids processing individual payments per respondent and instead can be carried out with a single payment to the charity.

[13] See "Passing Information via Query Strings" on Qualtrics support website: https://www.qualtrics.com/support/survey-platform/survey-module/survey-flow/standard-elements/passing-information-through-query-strings/.

[14] See "SSO Authenticator" on Qualtrics support website: https://www.qualtrics.com/support/survey-platform/survey-module/survey-flow/advanced-elements/authenticator/sso-authenticator/.

[15] Because many of the links that were corrupted had been shared on Facebook or Instagram, it seemed likely that any legitimate beekeeper trying to respond would have a Facebook account. We also included a message with our contact information in case the participant did not want to log in to Facebook to participate.

[16] See "Hidden question traps for bots" discussion for using JavaScript in Qualtrics: https://community.qualtrics.com/XMcommunity/discussion/6152/hidden-question-traps-for-bots

REFERENCES

Al-Fannah, N. M. 2017. "Making Defeating Captchas harder for Bots." In *2017 Computing Conference* 775–82. London, UK: IEEE. https://doi.org/10.1109/SAI.2017.8252183.

Avemegah, Edem, Wei Gu, Abdelrahim Abulbasher, Kristen Koci, Ayorinde Ogunyiola, Joyce Eduful, Shuang Li, et al. 2021. "An Examination of Best Practices for Survey Research with Agricultural Producers." *Society & Natural Resources* 34(4): 538–49.

Bauermeister, Jose A., Emily Pingel, Marc Zimmerman, Mick Couper, Alex Carballo-Diéguez, and Victor J. Strecher. 2012. "Data Quality in HIV/AIDS Web-Based Surveys: Handling Invalid and Suspicious Data." *Field Methods* 24(3): 272–91.

Belliveau, J., and I. Yakovenko. 2022. "Evaluating and Improving the Quality of Survey Data from Panel and Crowd-Sourced Samples: A Practical Guide for Psychological Research." *Experimental and Clinical Psychopharmacology* 30(4): 400–8. https://doi.org/10.1037/pha0000564.

Blesse, Sebastian, and Friedrich Heinemann. 2020. "Citizens' Trade-Offs in State Merger Decisions: Evidence from a Randomized Survey Experiment." *Journal of Economic Behavior & Organization* 180: 438–71.

Buchanan, E. M., and J. E. Scofield. 2018. "Methods to Detect Low Quality Data and its Implication for Psychological Research." *Behavior Research Methods* 50: 2586–96. https://doi.org/10.3758/s13428-018-1035-6.

Chandler, J. J., and G. Paolacci. 2017. "Lie for a Dime: When Most Prescreening Responses Are Honest but Most Study Participants Are Impostors." *Social Psychological and Personality Science* 8(5): 500–8.

Charité, Jimmy, Raymond Fisman, Ilyana Kuziemko, and Kewei Zhang. 2022. "Reference Points and Redistributive Preferences: Experimental Evidence." *Journal of Public Economics* 216: 104761.

Cheng, Haotian, Dayton M. Lambert, Karen L. DeLong, and Kimberly L. Jensen. 2022. "Inattention, Availability Bias, and Attribute Premium Estimation for a Biobased Product." *Agricultural Economics* 53(2): 274–88.

Chmielewski, Michael, and Sarah C. Kucker. 2020. "An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results." *Social Psychological and Personality Science* 11(4): 464–73.

Crawford, S. D., M. P. Couper, and M. J. Lamias. 2001. "Web Surveys: Perceptions of Burden." *Social Science Computer Review* 19(2): 146–62.

Czajka, John L., and Amy Beyler. 2016. "Declining Response Rates in Federal Surveys: Trends and Implications (Background Paper)." *Mathematica Policy Research* 1: 1–54.

DellaVigna, Stefano, and Devin Pope. 2022. "Stability of Experimental Results: Forecasts and Evidence." *American Economic Journal: Microeconomics* 14(3): 889–925.

Dennis, S. A., B. M. Goodson, and C. Pearson. 2020. "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." *Behavioral Research in Accounting* 32(1): 119–34.

Dillman, D. A., J. D. Smyth, and L. M. Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons.

Dupuis, Marc, Emanuele Meier, and Félix Cuneo. 2019. "Detecting Computer-Generated Random Responding in Questionnaire-Based Data: A Comparison of Seven Indices." *Behavior Research Methods* 51(5): 2228–37.

Fisman, Raymond, Keith Gladstone, Ilyana Kuziemko, and Suresh Naidu. 2020. "Do Americans Want to Tax Wealth? Evidence from Online Surveys." *Journal of Public Economics* 188: 104207.

Gao, Z., L. A. House, and J. Xie. 2016. "Online Survey Data Quality and its Implication for Willingness-to-Pay: A Cross-Country Comparison." *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 64(2): 199–221.

Goodrich, Brittney K., and Jennie L. Durant. "Going Nuts for More Bees: Factors Influencing California Almond Pollination Fees."*ARE Update* 24(1) (2020): 5–8. University of California Giannini Foundation of Agricultural Economics.

Goodrich, Brittney K., Jeffrey C. Williams, and Rachael E. Goodhue. 2019. "The Great Bee Migration: Supply Analysis of Honey Bee Colony Shipments into California for Almond Pollination Services." *American Journal of Agricultural Economics* 101: 1353–72.

Griffin, Marybec, Richard J. Martino, Caleb LoSchiavo, Camilla Comer-Carruthers, Kristen D. Krause, Christopher B. Stults, and Perry N. Halkitis. 2021. "Ensuring Survey Research Data Integrity in the Era of Internet Bots." *Quality and Quantity* 56(4): 2841–52.

Hardigan, P. C., C. T. Succar, and J. M. Fleisher. 2012. "An Analysis of Response Rate and Economic Costs between Mail and Web-Based Surveys among Practicing Dentists: A Randomized Trial." *Journal of Community Health* 37(2): 383–94.

Ince, B. Ü., P. Cuijpers, E. Van't Hof, and H. Riper. 2014. "Reaching and Recruiting Turkish Migrants for a Clinical Trial through Facebook: A Process Evaluation." *Internet Interventions* 1(2): 74–83.

Johansson, R., A. Effland, and K. Coble. 2017. "Falling Response Rates to USDA Crop Surveys: Why it Matters." *Farmdoc Daily* 7(9) Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, January 19, 2017. https://farmdocdaily.illinois.edu/2017/01/falling-response-rates-to-usda-crop-surveys.html

Johnston, Robert, F. Lupi, K. Moeltner, E. Besedin, Z. Yao, T. Ndebele, S. Crema, S. Peery, H. Kim, and J. A. Herriges. 2021. "Do you Know Who's Answering your Survey? Expanding Threats to the Integrity of Online Panel Data in Environmental and Resource Economics." Presented at the AERE Summer Conference.

Kennedy, Ryan, Scott Clifford, Tyler Burleigh, Philip D. Waggoner, Ryan Jewell, and Nicholas J. G. Winter. 2020. "The Shape of and Solutions to the MTurk Quality Crisis." *Political Science Research and Methods* 8(4): 614–29.

Kramer, J., A. Rubin, W. Coster, E. Helmuth, J. Hermos, D. Rosenbloom, Rich Moed, et al. 2014. "Strategies to Address Participant Misrepresentation for Eligibility in Web-Based Research." *International Journal of Methods in Psychiatric Research* 23(1): 120–9. https://doi.org/10.1002/mpr.1415.

Lawlor, Jennifer, Carl Thomas, Andrew T. Guhin, Kendra Kenyon, Matthew D. Lerner, U. C. A. S. Consortium, and Amy Drahota. 2021. "Suspicious and Fraudulent Online Survey Participation: Introducing the REAL Framework." *Methodological Innovations* 14(3): 20597991211050467.

Lennon, Conor, Keith F. Teltser, Jose Fernandez, and Stephan Gohmann. 2023. "How Morality and Efficiency Shape Public Support for Minimum Wages." *Journal of Economic Behavior & Organization* 205: 618–37.

Loxton, D., J. Powers, A. E. Anderson, N. Townsend, M. L. Harris, R. Tuckerman, Stephanie Pease, Gita Mishra, and Julie Byles. 2015. "Online and Offline Recruitment of Young Women for a Longitudinal Health Survey: Findings from the Australian Longitudinal Study on Women's Health 1989–95 Cohort." *Journal of Medical Internet Research* 17(5): e4261.

Meade, A. W., and Craig, S. B. 2012. "Identifying Careless Responses in Survey Data." *Psychological Methods* 17 (3): 437–455. https://doi.org/10.1037/a0028085.

Malone, Trey, and Jayson L. Lusk. 2018a. "Consequences of Participant Inattention with an Application to Carbon Taxes for Meat Products." *Ecological Economics* 145: 218–30.

Malone, Trey, and Jayson L. Lusk. 2018b. "A Simple Diagnostic Measure of Inattention Bias in Discrete Choice Models." *European Review of Agricultural Economics* 45(3): 455–62.

Malone, Trey, and Jayson L. Lusk. 2019. "Releasing the Trap: A Method to Reduce Inattention Bias in Survey Data with Application to U.S. Beer Taxes." *Economic Inquiry* 57(1): 584–99.

Meyer, B. D., W. K. Mok, and J. X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29(4): 199–226.

Moss, Aaron J., Cheskie Rosenzweig, Shalom N. Jaffe, Richa Gautam, Jonathan Robinson, and Leib Litman. 2021. "Bots or Inattentive Humans? Identifying Sources of Low-Quality Data in Online Platforms." Working paper.

National Research Council. 2008. *Understanding American Agriculture: Challenges for the Agricultural Resources Management Survey. Panel to Review USDA's Agricultural Resource Management Survey Committee on National Statistics, Division of Behavioral Social Sciences and Education*. Washington, D.C.: The National Academies Press.

Penn, Jerrod, and Wuyang Hu. 2022. "Payment Versus Charitable Donations to Attract Agricultural and Natural Resource Survey Participation." Working Paper.

Penn, Jerrod M., Daniel R. Petrolia, and J. Matthew Fannin. 2023. "Hypothetical Bias Mitigation in Representative and Convenience Samples." *Applied Economic Perspectives and Policy* 45(2): 721–43.

Pozzar, R., M. Hammer, M. Underhill-Blazey, A. Wright, J. Tulsky, F. Hong, D. Gundersen, and D. Berry. 2020. "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* 22(10): e23021 URL. https://www.jmir.org/2020/10/e23021, https://doi.org/10.2196/23021.

Reist, B. M., J. B. Rodhouse, S. T. Ball, and Linda J. Young. 2019. "Subsampling of Nonrespondents in the 2017 Census of Agriculture." Research and Development Division Washington DC 20250.

Rommel, Jens, Julian Sagebiel, Marieke Cornelia Baaken, Jesús Barreiro-Hurlé, Douadia Bougherara, Luigi Cembalo, Marija Cerjak, et al. 2022. "Farmers' risk preferences in 11 European farming systems: A multi-country replication of Bocquého et al." *Applied Economic Perspectives and Policy* (forthcoming). https://doi.org/10.1002/aepp.13330.

Sandstrom, K., F. Lupi, H. Kim, and J. A. Herriges. 2023. "Comparing Water Quality Valuation across Probability and Non-Probability Samples." *Applied Economic Perspectives and Policy* 45(2): 744–61.

Sartore, L., K. Toppin, L. Young, and C. Spiegelman. 2019. "Developing Integer Calibration Weights for Census of Agriculture." *Journal of Agricultural, Biological, and Environmental Statistics* 24(1): 26–48.

Saxon, James, and Nick Feamster. 2022. "Gps-Based Geolocation of Consumer Ip Addresses." In *International Conference on Passive and Active Network Measurement* 122–51. Cham: Springer.

Schonlau, Matthias, and Mick P. Couper. 2017. "Options for Conducting Web Surveys." *Statistical Science* 32(2): 279–92.

Simone, M. 2019. "Bots started sabotaging my online research. I fought back-STAT." https://www.statnews.com/2019/11/21/bots-started-sabotaging-my-online-research-i-fought-back/

Sivakorn, S., I. Polakis, and A. D. Keromytis. 2016. "I am Robot:(Deep) Learning to Break Semantic Image Captchas." In *2016 IEEE European Symposium on Security and Privacy (EuroS&P)* 388–403. Saarbruecken: IEEE.

Snowberg, Erik, and Leeat Yariv. 2021. "Testing the Waters: Behavior across Participant Pools." *American Economic Review* 111(2): 687–719.

Stantcheva, Stefanie. 2022. *How to Run Surveys: A Guide to Creating your Own Identifying Variation and Revealing the Invisible*. National Bureau of Economic Research Working Paper No. 30527. https://doi.org/10.3386/w30527

Storozuk, Andie, Marilyn Ashley, Véronic Delage, and Erin A. Maloney. 2020. "Got Bots? Practical Recommendations to Protect Online Survey Data from Bot Attacks." *The Quantitative Methods for Psychology* 16(5): 472–81.

te Velde, Vera L., and Winnifred Louis. 2022. "Conformity to Descriptive Norms." *Journal of Economic Behavior & Organization* 200: 204–22.

Teitcher, J. E., W. O. Bockting, J. A. Bauermeister, C. J. Hoefer, M. H. Miner, and R. L. Klitzman. 2015. "Detecting, Preventing, and Responding to "Fraudsters" in Internet Research: Ethics and Tradeoffs." *The Journal of Law, Medicine & Ethics* 43(1): 116–33. https://doi.org/10.1111/jlme.12200.

Weber, J. G., and D. M. Clay. 2013. "Who Does Not Respond to the Agricultural Resource Management Survey and Does it Matter?" *American Journal of Agricultural Economics* 95(3): 755–71.

Weigel, C., L. A. Paul, P. J. Ferraro, and K. D. Messer. 2021. "Challenges in Recruiting US Farmers for Policy-Relevant Economic Field Experiments." *Applied Economic Perspectives and Policy* 43(2): 556–72.

Whitehead, J. C., A. Ropicki, J. Loomis, S. Larkin, T. Haab, and S. Alvarez. 2023. "Estimating the Benefits to Florida Households from Avoiding another Gulf Oil Spill Using the Contingent Valuation Method: Internal Validity Tests with Probability-Based and Opt-in Samples." *Applied Economic Perspectives and Policy* 45(2): 705–20.

Zhang, Z., S. Zhu, J. Mink, A. Xiong, L. Song, and G. Wang. 2022. "Beyond Bot Detection: Combating Fraudulent Online Survey Takers." In *Proceedings of the ACM Web Conference 2022*, 699–709. New York, NY: Association for Computing Machinery.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Goodrich, Brittney, Marieke Fenton, Jerrod Penn, John Bovay, and Travis Mountain. 2023. "Battling Bots: Experiences and Strategies to Mitigate Fraudulent Responses in Online Surveys." *Applied Economic Perspectives and Policy* 45(2): 762–784. https://doi.org/10.1002/aepp.13353