



## Liar! Liar! Identifying eligibility fraud by applicants in digital health research

Jillian V. Glazer, Kirsten MacDonnell, Christina Frederick, Karen Ingersoll, Lee M. Ritterband<sup>\*</sup>

University of Virginia School of Medicine, Center for Behavioral Health & Technology, Department of Psychiatry & Neurobehavioral Sciences, 560 Ray C Hunt Dr, Charlottesville 22903, VA, USA

### ARTICLE INFO

#### Keywords:

Recruitment  
Identity verification  
Fraud  
Digital research  
Internet intervention

### ABSTRACT

Online studies enable researchers to recruit large, diverse samples, but the nature of these studies provides an opportunity for applicants to misrepresent themselves to increase the likelihood of meeting eligibility criteria for a trial, particularly those that provide financial incentives. This study describes rates of fraudulent applications to an online intervention trial of an Internet intervention for insomnia among older adults (ages  $\geq 55$ ). Applicants were recruited using traditional (e.g., flyers, health providers), online (e.g., Craigslist, Internet searches), and social media (e.g., Facebook) recruitment methods. Applicants first submitted an interest form that included identifying information (name, date of birth, address). This data was then queried against a national database (TransUnion's TLOxp) to determine the application's verification status. Applications were determined to be verified (i.e., information from interest form matched TLOxp report), potentially fraudulent (i.e., potential discrepancy in provided information on interest form versus TLOxp report), or fraudulent (i.e., confirmed discrepancy). Of 1766 total interest forms received, 125 (7.08%) were determined to be fraudulent. Enrollment attempts that were fraudulent were detected among 12.22% of applicants who reported learning of the study through online, 7.04% through social media, 4.58% through traditional, and 4.27% through other methods. Researchers conducting online trials should take precautions, as applicants may provide fraudulent information to gain access to their studies. Reviewing all applications and verifying the identities and eligibility of participants is critical to the integrity of online research trials.

### 1. Introduction

Recruiting participants for research studies can be conducted online, and has grown in parallel with significant growth in online health research over the past two decades. Studies that use the Internet and/or mobile phones to provide behavioral and mental health interventions have been particularly prolific (Granja et al., 2018; Marcolino et al., 2018). Effort and expense related to participant recruitment may be reduced by recruiting potential participants online without a need to meet in person (Inan et al., 2020). However, a significant challenge to Internet-based recruitment is the increased risk that online applicants are not who they say they are. This is especially important for Internet intervention research where researchers and participants may never interact in person. Some people use deception to become study participants in an attempt to gain access to a not-yet-available treatment or to

receive payments for completing study milestones. In fact, studies with participation incentives have six times the fraudulent behavior by participants than studies that did not include participant payments (Bowen et al., 2008). This is a critical issue as it directly impacts the integrity of the research. If studies include participants who do not meet eligibility criteria for the trial, the value of the science becomes questionable.

Unfortunately, misrepresentation is not an uncommon occurrence. Wessling et al. (2017) found rates of misrepresentation between 24% and 83% in surveys posted to mTURK on Amazon, with respondents being more likely to answer deceptively on characteristics that may be more difficult to prove, such as product ownership, compared to more identifiable characteristics, like first and last name. Several other studies that monitored misrepresentation and fraud have reported fraudulence rates between 18% and 35% (Ballard et al., 2020; Bauermeister et al., 2012; Bull et al., 2009; Schure et al., 2019; Young et al., 2020).

<sup>\*</sup> Corresponding author at: University of Virginia School of Medicine, Center for Behavioral Health & Technology, Ivy Foundation Translational Research Building, 560 Ray C Hunt Drive, Charlottesville, VA 22903, USA.

E-mail addresses: [jvg3ab@virginia.edu](mailto:jvg3ab@virginia.edu) (J.V. Glazer), [kem6e@virginia.edu](mailto:kem6e@virginia.edu) (K. MacDonnell), [ccf7u@virginia.edu](mailto:ccf7u@virginia.edu) (C. Frederick), [kes7a@virginia.edu](mailto:kes7a@virginia.edu) (K. Ingersoll), [leer@virginia.edu](mailto:leer@virginia.edu) (L.M. Ritterband).

<https://doi.org/10.1016/j.invent.2021.100401>

Received 1 November 2020; Received in revised form 30 April 2021; Accepted 4 May 2021

Available online 9 May 2021

2214-7829/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Some researchers have implemented processes to help address potential fraud (Ballard et al., 2019; Bowen et al., 2008; Teitcher et al., 2015). Teitcher et al., (2015) outlined methods researchers can take to detect and prevent fraudulent online submissions. These include using survey design features to prevent bots (e.g. CAPTCHAs), reviewing questionnaire responses (e.g. looking for those who answer the same for each question), checking respondent computer information to detect duplicate IP addresses, and checking the identity information provided against external sources like social media or WhitePages. Many countries have options for electronic identification, such as through electronic identification cards and BankIDs. Even when processes were put into place to detect fraudulent behaviors, Ballard et al. (2019) found that 28.7% of submitted web-based surveys were “fraudulent” and an additional 10.1% were “potential fraud.” Given fraud detection methods are not infallible, it may be necessary to incorporate both automated detection and human monitoring to reduce inclusion of those who do not meet study criteria.

Despite these risks, online recruitment of study participants is here to stay given the significant advantages to recruiting participants online. Internet-based recruitment decreases barriers to participating in research studies by removing obstacles such as the need to physically go to a clinic or research center. There are also time- and cost-efficiencies when recruiting online in comparison to traditional methods, such as posting paper flyers and print advertisements (Frandsen et al., 2013). Often, greater numbers of people can be recruited, and recruitment periods may be shorter, especially when posting advertisements to classified advertisement websites, such as Craigslist, or social media platforms like Facebook (Adam et al., 2016; Kayrouz et al., 2016; MacDonnell et al., 2019). In addition, online recruitment methods provide a particularly effective way to enroll more geographically-dispersed populations, across age groups, in hard-to-reach populations, and in medically-specific populations (Kayrouz et al., 2016; MacDonnell et al., 2019; Ritterband et al., 2009; Topolovec-Vranic & Natarajan, 2016). Facebook has been a particularly effective source of recruitment for groups that can be hard to reach (Carter-Harris et al., 2016; MacDonnell et al., 2019). Compared to traditional and registry-based recruitment, online recruitment strategies, including Facebook, Instagram, and Craigslist, have been found to yield the greatest number of participants across multiple Internet-based clinical trials, while also being both time- and cost-effective (Lattie et al., 2018).

One group that is growing in both size and research focus is “older adults,” those 55 years of age and older (Fichten et al., 2000). Recruiting older adults online has become a viable option as Internet use within this group climbs (Anderson et al., 2019; Carter-Harris et al., 2016; Cowie & Gurney, 2018). Only 12% of 50–64 year old American adults and 27% of American adults 65 and older remained offline as of 2019, either by choice or lack of access to the needed technology tools (Anderson et al., 2019). Each year, more older adults gain access to computers and the Internet, and the majority of older adult households now have access to both (Anderson et al., 2019; Ryan & Lewis, 2017). However, older adults may be concerned about participating in trials if they are unsure whether an online advertisement for a trial is legitimate, or whether they have the skills to complete the study tasks.

This paper describes and explores the recruitment methods used to target a population of older adults with insomnia, the discovery of fraudulent applicant identities, implementation of identity verification, results of verifications, rate of fraudulent applications by recruitment source, and suggestions for improving the integrity of future research studies.

## 2. Methods

Sleep Healthy Using the Internet for Older Adult Sufferers of Insomnia and Sleeplessness (SHUTi OASIS) is a fully automated, interactive, and tailored intervention based on Cognitive-Behavioral Therapy for insomnia (CBTi). The current randomized controlled study evaluated

the SHUTi OASIS program in an older adult sample who met the DSM-V criteria for insomnia. Individuals who were age 55 years of age and older were eligible, and a particular effort was made to include a sizable percentage of participants over the age of 70. Participants had to live in the United States, be comfortable reading and speaking English, and have regular access to the Internet. Participants meeting criteria for other sleep disorders (e.g. sleep apnea, RLS) without stable treatment, severe cognitive impairment, or medical or psychiatric conditions that put them at undue risk were ineligible. Participants were also excluded for: (1) current psychological treatment for insomnia; (2) recent initiation of psychological/psychiatric treatment for another reason; (3) unstable medication regimens; and (4) shift work interfering with the establishment of regular sleep patterns. The recruitment period spanned 50 weeks, from mid-May 2018 to late-April 2019. The University of Virginia's Institutional Review Board for Social and Behavioral Sciences approved the study (ClinicalTrials.gov ID: NCT03213132).

Recruitment utilized traditional methods, online postings, and social media advertisements, all of which were approved by the UVA IRB. Table 1 lists the categories of recruitment, as well as specific methods, used in this study. One-hundred-twenty-six locations, including senior facilities, community centers, medical offices, and businesses, across the United States agreed to place a flyer for the study at their location. Individuals also learned of this study through popular press articles that mentioned the SHUTi program or the current trial. Health care providers with knowledge of SHUTi also referred patients to the program, while other applicants learned about it through word of mouth (such as the sharing of details by a friend or family member).

Internet-based recruitment efforts centered on advertisements posted to Craigslist and Facebook/Instagram. Given the success of Craigslist in recruiting participants for previous studies (MacDonnell et al., 2019), 38 paid advertisements were posted across 18 states in locations with greater percentages of older and racial-ethnic minority populations. A variety of Facebook interest groups related to the study population, such as AARP and retirement groups, were contacted and asked to share study information. Those interested in the study were directed to the study website where more information was available about the study and those interested could complete an interest form (screener) with personal (name, address, phone number, email, and birth date) and sleep-related information. While all social media methods were posted online, the decision to separate the “online postings” and “social media” was reached to examine possible differences in rates of fraud in domains based on social networking, as opposed to less specific online sites such as through Google searches.

A link to the study site was posted to the research center's Facebook page. It is possible that some interested applicants shared the link on their Facebook feed or directly with others. Paid advertisements on Facebook and Instagram were also posted through the University of Virginia Health System (UVAHS) Marketing and Communications Department's social media accounts. Using Facebook's audience targeting feature, and working within their age categorizations, these posts appeared on the home page of users aged 65 and older (according to user profile information). A total of 4 Facebook ad campaigns were conducted, posting ads in 32 states and the District of Columbia for 14 days each. Each campaign featured between 5 and 14 states/districts and no state receiving multiple paid ads. During the posting period, each ad was visible on both the UVAHS Facebook and Instagram pages. Applicants who reported that they learned of the study via Facebook were asked to provide additional details (if it was through a paid ad, shared by a friend, private message, in a group, etc.). Those indicating they had seen an ad were considered to have heard about the study from one of our paid advertisements. All unspecified Facebook entries are categorized as general “Facebook,” including any applications citing Facebook submitted while a paid ad was active in that region.

A company (BeHealth Solutions, LLC) who had licensed the original version of the SHUTi program (but not tailored for older adults) maintained a database of interested individuals for future studies (LR is a co-

**Table 1**  
Interest forms and participants yielded across recruitment categories, types, and sources.

	Recruitment Type	Location of Ads	Applicants	% of total applicants	Initially Eligible	% of total initially eligible	Enrolled	% of total enrolled that came source	% of applicants from source who enrolled
Traditional	Flyers	Senior Facilities	15	0.85%	13	1.16%	7	2.05%	46.67%
		Businesses	9	0.51%	3	0.27%	1	0.29%	11.11%
		Community Centers	7	0.40%	5	0.45%	1	0.29%	14.29%
		Health Clinics	4	0.23%	3	0.27%	2	0.59%	50.00%
		<i>Flyers Total</i>	35	1.98%	24	2.14%	11	3.23%	31.43%
	Health Providers	Doctor/Health Care Provider	221	12.50%	155	13.80%	53	15.54%	23.98%
		Sleep Specialist	103	5.83%	68	6.06%	17	4.99%	16.50%
		Mental Health Professional	17	0.96%	6	0.53%	2	0.59%	11.76%
		<i>Health Providers Total</i>	341	19.29%	229	20.39%	72	21.11%	21.11%
		Word-of-Mouth	SHUTi User	20	1.13%	11	0.98%	4	1.17%
	Conference Presentation		7	0.40%	4	0.36%	3	0.88%	42.86%
	Family/Friend		130	7.36%	82	7.30%	33	9.68%	25.38%
	<i>Word-of-Mouth Total</i>		157	8.94%	97	8.64%	40	11.73%	25.32%
	Published Materials	Consumer Reports Magazine/	180	10.18%	114	10.15%	19	5.57%	10.56%
		Newspaper Online/	55	3.11%	38	3.38%	13	3.81%	23.64%
		Print							
		Harvard Health Newsletter	11	0.62%	9	0.80%	3	0.88%	27.27%
		Bottom Line	7	0.40%	4	0.36%	2	0.59%	28.57%
		<i>Published Materials Total</i>	253	14.31%	165	14.69%	37	10.85%	14.62%
		<i>Traditional Total</i>		786	44.51%	515	45.86%	160	46.27%
Online	Internet	224	12.68%	151	13.45%	46	13.49%	20.54%	
	Clinicaltrials.gov	25	1.41%	21	1.87%	4	1.17%	16.00%	
	Reddit	2	0.11%	2	0.18%	1	0.29%	50.00%	
	Craigslist (paid)	166	9.39%	120	10.69%	32	9.38%	19.28%	
	Unsolicited Shared Link	82	4.64%	0	0.00%	0	0.00%	0.00%	
<i>Online Total</i>		499	28.28%	294	26.18%	83	24.34%	16.60%	
Social Media	Facebook UVA Ad (Paid)	126	7.13%	93	8.28%	34	9.97%	26.98%	
	Facebook	144	8.14%	84	7.48%	15	4.40%	10.42%	
<i>Social Media Total</i>		270	15.27%	177	15.76%	49	14.37%	18.15%	
Other	Other	24	1.36%	15	1.34%	7	2.05%	29.17%	
	BeHealth Solutions	187	10.58%	122	10.86%	42	12.32%	22.46%	
<i>Other Total</i>		211	11.93%	137	12.20%	49	14.37%	23.22%	
<b>Total</b>			<b>1766</b>	<b>100.0%</b>	<b>1123</b>	<b>100.00%</b>	<b>341</b>	<b>100.00%</b>	

Note: Enrolled numbers reflect all participants who were enrolled, including those later withdrawn.

founder of BeHealth Solutions, see Declaration of Interest). The company sent emails to those believed to be age 55 and older and informed them of the SHUTi OASIS study. Applicants who learned about the study from these email announcements or other methods, such as television programs, a member of the research team, or other sources, were grouped together as ‘Other.’

### 2.1. Recruitment synopsis

A total of 155 advertisements were posted across all advertisement methods. Of these, 55 (35.48%) were online or social media advertisements. All online advertisements included direct links to the SHUTi OASIS website ([www.SHUTiOASIS.org](http://www.SHUTiOASIS.org)), which included details about the study and a study interest form to complete as the first step in determining eligibility for the trial. Additional information about the placements of advertisements, number of applicants received, and number of participants enrolled from each recruitment category can be found in Table 1.

The interest form was a brief Qualtrics survey that captured applicant demographic information and study eligibility criteria. Submitted interest forms were then available to study staff and were manually reviewed to determine initial study eligibility. Eligible applicants were contacted to schedule a phone screening, while ineligible applicants

were informed that they were not eligible through email.

### 2.2. Phone screen/enrollment process

The phone screen was a more extensive process to ensure applicants met study eligibility. During the screening, participants answered questions about their sleep and history of sleep-related issues, and their medical and psychological history. Applicants were enrolled into the study if they met all criteria in the phone screen. Applicants were not compensated for submitting their interest form to the study, completing a phone screening, or completing the baseline assessment after enrollment. Participants were compensated after completing each of the three follow-up assessments for a total of up to \$200 over the course of their participation.

### 2.3. Discovery of fraudulent participants

Concerns were raised when some participants appeared to submit interest forms more than once using the same or different names and provide inconsistent contact information. Applications contained identical phone numbers, addresses, email addresses, and/or names. Some applications appeared to be duplicate submissions, where the applicant may have forgotten that they had previously submitted an application.

Others appeared to be intentionally dishonest, with important eligibility information such as date of birth were changed. During the phone screening process, researchers suspected fraudulent activity due to a number of seemingly suspicious behaviors by applicants, including reporting to be the relative of another participant who sounded identical on the phone. Some voices seemed to sound too young to meet study inclusion criteria. When these participants were questioned, different or inconsistent information was provided, leading the study team to investigate further.

To address these concerns, options were discussed in collaboration with the IRB, and a decision was made to use a tool for identity verification for all applications. The service, TransUnion's TLOxp ([www.tlo.com](http://www.tlo.com)), was selected and pulls together publicly available information such as name, address history, date of birth, Social Security Number, phone and email information, and possible relatives. The expectation was that, through the use of the TLOxp service, fraud would be more readily detected and that verifying identify and contact information would serve as a useful retention strategy for this longitudinal study. The IRB approved use of TLOxp on previously enrolled applicants since the information was in the public record. A statement informing future applicants that this process was in place was included on the interest form.

#### 2.4. Identity verification

Two research team members were responsible for verifying information from interest forms using the TransUnion's TLOxp service. All reports were downloaded and saved to a secured shared drive. Verified information and fraud status were entered into a password protected Excel database.

#### 2.5. Verification process

Fig. 1 shows the process used to determine an applicant's verification status as ultimately either "verified" or "fraudulent." Once an interest form was received, the name, date of birth and current home address of the applicant was entered into the TLO database. "Fraudulent" applications were those that did not pass the verification check by failing to match on one or more of the three identifiers, while "verified" applications matched on all three identifiers. In cases where any identifier(s) did not exactly match (for example, the birth date was one day or one year off), the application would be temporarily assigned the label of "potentially fraudulent." Potentially fraudulent applicants meeting the study's initial eligibility criteria were contacted by research staff and asked to confirm their information. Those who responded with information matching the TLOxp report had their verification status switched

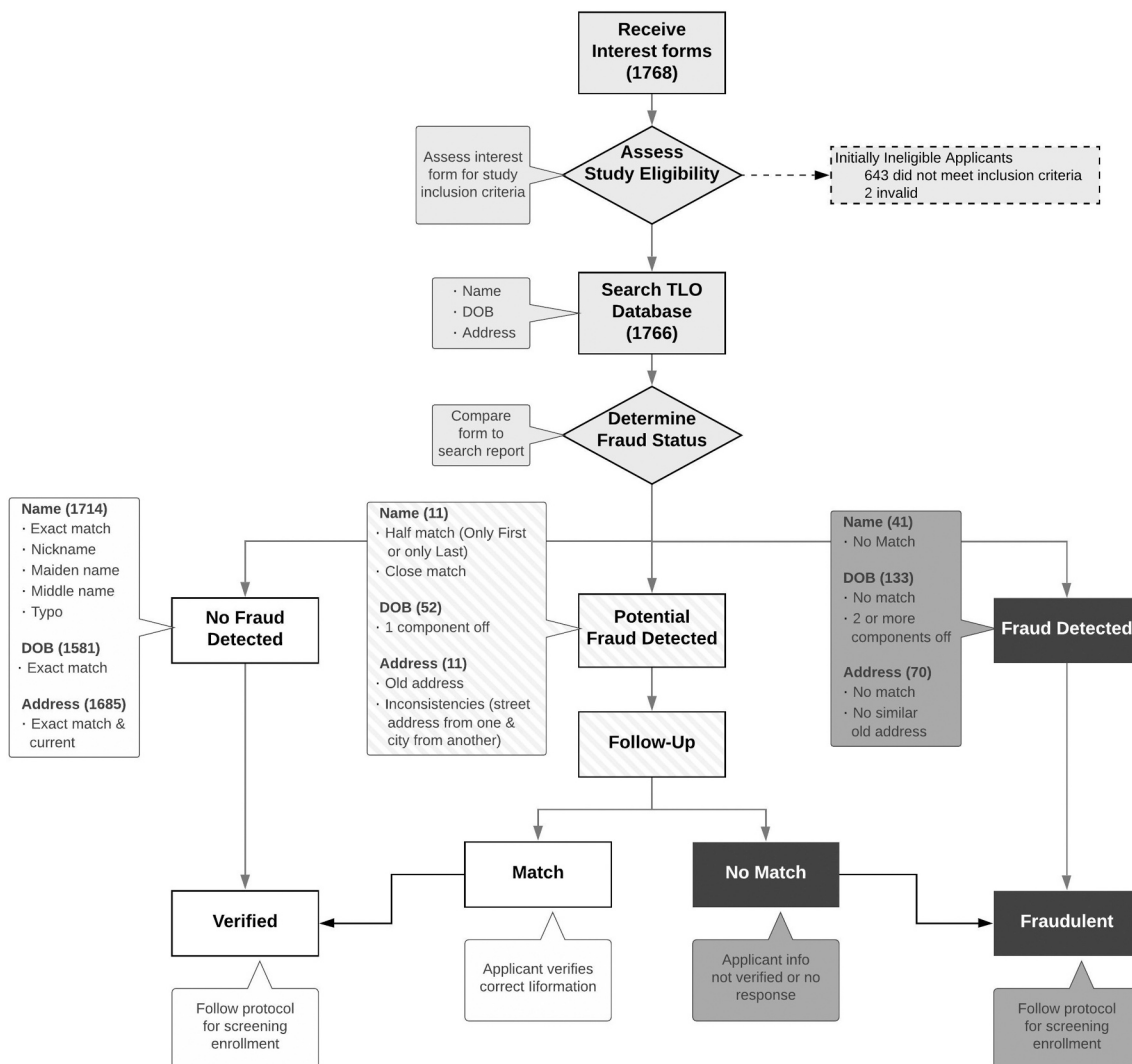


Fig. 1. Flow chart showing identity verification process. Note: All interest forms received were processed for verification except for 2 that did not belong to US residents, listed as invalid above. However, only 1123 applicants were considered to be initially eligible for the study.

to “verified,” and those who did not confirm the matching information or did not respond after two contact attempts were deemed as “fraudulent.” Potentially fraudulent applicants who had not met the study’s initial eligibility criteria were not contacted by research staff to confirm their information and instead kept the label of “potentially fraudulent.”

## 2.6. Analysis

Descriptive statistics were used to calculate means, standard deviations, ranges, and percentages as well as rates of enrollment and fraud across recruitment sources.

## 3. Results

### 3.1. Advertisement yield

Throughout the recruitment period, 1768 unique interest forms were received, averaging 34 interest forms per week. Submissions came from all 50 states in the US and the District of Columbia. The most applications were received from California (198, 11.21%), Virginia (189, 10.70%), and New York (125, 7.08%). Two interest forms were not checked because they were from international applicants and the TLOxp database only includes information for US residents. Results and discussion of analyses include the remaining 1766 interest forms.

As shown in [Table 1](#), among the 1766 interest forms, 786 (44.51%) applicants reported learning about the study through traditional recruitment methods, 499 (28.26%) through online advertisements, 270 (15.29%) through social media, and 211 (11.95%) through other methods.

### 3.2. Study applicant characteristics

The racial distribution was: 1595 (90.32%) White, 77 (4.36%) Black, 36 (2.04%) Multiracial, 36 (2.04%) Asian, 3 (0.17%) American Indian, 3 (0.17%) Native Hawaiian or other Pacific Islander, 2 (0.11%) Alaskan Native, and 14 (0.79%) other. Sixty-five applicants (3.68%) reported Hispanic ethnicity. The average reported age of applicants was 63.16 years ( $SD = 10.74$ , range 18.70–99.55 years of age). However, after reviewing the TLOxp check of each applicant, there were 30 interest forms in which dates of birth were not verifiable. The remaining 1736 applicants had a confirmed average age of 63.26 years ( $SD = 10.86$ , range 21.71–95.45 years of age).

### 3.3. Applicant eligibility

Of the 1766 interest forms, 643 did not meet study inclusion criteria for a variety of reasons, including not meeting the criteria for current insomnia (380, 21.52%), in current counseling for sleep difficulties (76, 4.30%), in new psychological counseling (143, 8.10%), had a current night-shift job (27, 1.53%), and other (210, 11.89%). Eighty-two of the 1766 interest forms were submitted through a duplicated Qualtrics link that occurred after an applicant copied the unique URL associated with their Qualtrics interest form and posted it to Twitter instead of sharing the general study website URL. Others then used this duplicated link to complete their own interest forms, which were not imported by the Qualtrics-SHUTi OASIS API due to the use of the duplicated link. These applications were deemed invalid and not considered for study participation but were assessed for both study eligibility and fraud. Only one of the 82 duplicate link applications was initially eligible. The remaining 1123 submitted interest forms were considered to be initially eligible. Of the 1123 applications appearing to meet inclusion criteria, 183 (10.36%) were lost to contact prior to enrollment and an additional 275 (15.57%) were received after study enrollment ended.

## 3.4. Application verifiability

[Table 2](#) includes the results of the identity verification across recruitment sources and the rates of fraud across major recruitment categories. While the majority of applications were verified, 125 (7.08%) interest forms were deemed “fraudulent.” Applications were identified as fraudulent for one or more reasons, including unverified name (41), unverified date of birth (133), or unverified home address (70). When the information for fifteen applications was entered into the search fields on TLOxp, the service responded with a message saying that there were no reports matching the entered information or similar to it. This was interpreted to mean that the information provided did not match a person who could be verified through the TLOxp service and therefore considered them to be fraudulent. Two of these applications used the personal information of someone who was deceased. The research team also tracked unverified email addresses and phone numbers for applications that already had unverifiable information resulting in the identification of 64 instances of an unverified email address and 47 of an unverified phone number.

Rates of fraud were noticeably different across sources of recruitment (see [Table 2](#)). Online recruitment sources included higher rates of fraudulent applications (12.22%) than traditional (4.58%), social media (7.04%), and other (4.27%), with Craigslist as the recruitment source as having the most number of applicants resulting in failed checks (29 of the 166 total failed checks). Within the social media recruitment methods category, there was considerable disparity in rates of fraud between applications stemming from the two sources (Paid Facebook Ad: 3.97%, General Facebook: 9.72%).

## 4. Discussion

It is critically important that fraud is considered, tracked, and managed when recruiting applicants and conducting research trials online. This study is one of the first to examine rates of fraud by individuals attempting to meet eligibility for a study of an Internet intervention. In the few studies that have examined online research fraud, the focus has been on online surveys ([Ballard et al., 2019](#)). In contrast, this study involved a 62-week clinical trial with multiple assessments, including a phone interview. More than 7% of applicants were identified as fraudulent. While this rate of fraud may seem relatively minor, the consequences of including even a small number of people who are not who they say they are could have considerable impact on the outcomes of and conclusions drawn from research.

Fraudulent behavior is unlikely to be unique to Internet-based studies; however, some people may feel more comfortable providing false or misinformation in these type of trials given the greater sense of perceived anonymity and distance. Data that are more difficult to confirm, such as reports of symptom severity, tend to be more readily misrepresented than those which can be more easily confirmed, such as personal demographics ([Wessling et al., 2017](#)). Additionally, while incentives can be effective in boosting longitudinal participant retention as well as participant responses to online surveys, they are also more likely to increase rates of multiple submissions ([Teitcher et al., 2015](#)). Although the interest form in this study was not incentivized, it was clearly advertised that those enrolled would be eligible for payment and access to the SHUTi intervention, which may have motivated applicants to misrepresent themselves on the interest form in hopes of enrollment. Those who do not perceive punishment may be more inclined to offer fraudulent information in exchange for some desired outcome, whether it be payment or access to a treatment or program ([Bowen et al., 2008](#)). In this study, in which older adults were sought, there was a concern that younger people might inflate their age to gain access. These suspicions were confirmed through verifications using the TLOxp service, where many applicants claimed ages that did not match the corresponding TLOxp reports. Interestingly, while many applicants claimed seemingly inflated ages, there were also several instances of applicants claiming

**Table 2**  
 Fraud across recruitment categories, types, and sources.

	Recruitment Type	Recruitment Source	Applicants	Verified	Potential Fraud	Fraud
Traditional	Flyers	Senior Facilities	15	13 (86.67%)	0 (0.00%)	2 (13.33%)
		Businesses	9	9 (100.00%)	0 (0.00%)	0 (0.00%)
		Community Centers	7	7 (100.00%)	0 (0.00%)	0 (0.00%)
		Health Clinics	4	4 (100.00%)	0 (0.00%)	0 (0.00%)
		<i>Flyers Total</i>	35	33 (94.29%)	0 (0.00%)	2 (5.71%)
	Health Providers	Doctor/Health Care Provider	221	212 (95.93%)	3 (1.36%)	6 (2.71%)
		Sleep Specialist	103	93 (90.29%)	1 (0.97%)	9 (8.74%)
		Mental Health Professional	17	17 (100.00%)	0 (0.00%)	0 (0.00%)
		<i>Health Providers Total</i>	341	322 (94.43%)	4 (1.17%)	15 (4.40%)
	Word-of-Mouth	SHUTi User	20	20 (100.00%)	0 (0.00%)	0 (0.00%)
		Conference Presentation	7	7 (100.00%)	0 (0.00%)	0 (0.00%)
		Family/Friend	130	113 (86.92%)	3 (2.31%)	14 (10.77%)
		<i>Word-of-Mouth Total</i>	157	140 (89.17%)	3 (1.91%)	14 (8.92%)
	Published Materials	Consumer Reports	180	169 (93.89%)	6 (3.33%)	5 (2.78%)
		Magazine/Newspaper Online/Print	55	53 (96.36%)	2 (3.64%)	0 (0.00%)
Harvard Health Newsletter		11	11 (100.00%)	0 (0.00%)	0 (0.00%)	
Bottom Line		7	7 (100.00%)	0 (0.00%)	0 (0.00%)	
	<i>Published Materials Total</i>	253	240 (94.86%)	8 (3.16%)	5 (1.98%)	
	<i>Traditional Total</i>	786	735 (93.51%)	15 (1.91%)	36 (4.58%)	
Online	Internet	Internet	224	193 (86.16%)	8 (3.57%)	23 (10.27%)
		Clinicaltrials.gov	25	20 (80.00%)	0 (0.00%)	5 (20.00%)
		Reddit	2	1 (50.00%)	0 (0.00%)	1 (50.00%)
		Craigslist (paid)	166	133 (80.12%)	4 (2.41%)	29 (17.47%)
		Unsolicited Shared Link	82	72 (87.80%)	7 (8.54%)	3 (3.66%)
	<i>Online Total</i>	499	419 (83.97%)	19 (3.81%)	61 (12.22%)	
Social Media	Facebook UVA Ad (Paid)	Facebook	126	119 (94.44%)	2 (1.59%)	5 (3.97%)
		Facebook	144	129 (89.58%)	1 (0.69%)	14 (9.72%)
	<i>Social Media Total</i>	270	248 (91.85%)	3 (1.11%)	19 (7.04%)	
Other	Other	Other	24	22 (91.67%)	0 (0.00%)	2 (8.33%)
		BeHealth Solutions	187	179 (95.72%)	1 (0.53%)	7 (3.74%)
		<i>Other Total</i>	211	201 (95.26%)	1 (0.47%)	9 (4.27%)
	<b>Total</b>	1766	1603 (90.77%)	38 (2.15%)	125 (7.08%)	

Note: Percentages reflect percent of all applicants from that source. Verified applicants passed the identity verification check, while those deemed fraudulent failed the verification check. Applicants labeled “potential fraud” required clarification on their information to make a decision that was not received, and as such no final decision could be reached.

they were younger than their actual age in fear of being ineligible due to an assumed upper age limit.

Observed levels of fraud were found to differ across various means of recruitment. Traditional methods of recruitment, including the use of physical paper flyers, health provider recommendations, word-of-mouth, and established media outlets, appear to result in fewer cases of misrepresentation by potential participants (<5% in total) in this study. However, recent web-based survey research that used in-person recruitment events and peer referrals found >40% of participants recruited from an in-person event were fraudulent, and 68% of eligible peer referrals were also deemed fraudulent (Young et al., 2020). In this study, online sources and social media channels appear to lead to greater numbers of fraudulent applications than traditional methods, with Craigslist, specifically, having the highest number of fraudulent applications (29, or 17% of applications from this recruitment source). It is possible that applicants perceive these online methods to be more anonymous than traditional ones and may be more inclined to purposefully misrepresent themselves for enrollment. It is also possible that study applicants, especially older adults as those who applied to this study, may perceive online advertisements as less trustworthy, and, therefore, purposefully modify their personal information to protect themselves from potential scams.

Compared to rates of detected fraud in other web-based studies (consistently between 20 and 30%) (Ballard et al., 2019; Schure et al., 2019; Young et al., 2020), the 7.08% found in this study is relatively low. It is possible that this is due in part to the nature of many of these other studies, which were largely online survey research that detected fraud after participants submit their surveys. However, Schure et al. (2019) assessed for fraud among participants enrolled in an Internet-based intervention trial and it was found that 23.5% of the enrolled sample was fraudulent, which is still much higher than the 7.08% fraud

in this study. While that study's design was also interventional and there more similar to the study discussed in this paper, it was of much shorter duration than the SHUTi OASIS trial (8 vs 62 weeks, respectively). It may, therefore, be important to not only consider the design of a study and the number of research requirements to complete it.

Using a verification service (in this case, the Transunion's TLOxp database) to authenticate potential participants and identify fraud does have some shortcomings. Members of certain populations may wish to keep their identity hidden, such as undocumented people who may not want to be identified. Other vulnerable populations, such as those who are transgender or gender nonconforming, may use an identity different than that found on a verification service. Additionally, those living with stigmatized diseases, or those using illicit substances, may fear their identity will be exposed. Further, when informed that the provided data will be verified by TLOxp service, applicants may believe their credit information will be checked and some may be less motivated or more anxious to apply.

A significant limitation of relying on the TLOxp service is that it only accesses public records and is not a background check. Results do not always match, especially phone numbers and email addresses. On some records, multiple entries for each point of information can make the applicant's report difficult to match. While this most frequently happened for names, addresses, and dates of birth, it can also occur with social security numbers. It is possible that some applicants provided fraudulent information that these processes did not detect. While it is not possible to be 100% certain in confirming whether someone is presenting fraudulent information, the processes and protocols put into place in this study provide reassurance that the decisions made to exclude those not meeting study criteria were likely correct.

The TLOxp service used in this research only searched records for those in the US. Many other countries have alternative methods of

electronic identification (eID); including the physical electronic identification cards for both offline and online identity authentication, bank-issued eIDs, and others. In countries where these forms of identification are available, conducting verifications such as the ones described in this research could be substituted with an approved form of electronic identification to verify both participant name and age.

#### 4.1. Recommendations

Based on the results of this study and our experience in conducting this trial, we provide a number of recommendations for researchers preparing to recruit participants for an online study. They are:

1. Establish a method for verifying applicant identity. While it may be impossible to completely eliminate fraud, confirming basic identity information by using identity verification systems, like TLOxp, can help researchers classify clearly ineligible study applicants. Another solution might be to request other documentation of identity, such as drivers' licenses, in lieu of running background checks. This might be a cheaper alternative to a service, but may create more paperwork, result in a time delay for authenticating identities, and potentially establish another technological barrier to enrollment for applicants due to the need for a scanner or camera. Research conducted outside of the US may use the appropriate form of electronic identification, if available, or national identification number. Regardless of what the verification method is, it is important to standardize the process, obtain ethical approvals for it, and train research staff to implement it before study recruitment begins. This recommendation aligns with guidelines set forth by the FDA requiring researchers to verify the identity of each participant in any research conducted remotely ([Food and Drug Administration, 2016](#)) in order to limit identity fraud.
2. Implement a secondary screen. In line with the recommendations of [Teitcher et al. \(2015\)](#) and considering the experiences of this research team, a secondary screen, such as a phone screen or consent review, where researchers engage in synchronous dialogue with the applicant should be considered. A conversation can also build rapport, and it ensures participants are fully informed of the purpose of the study and the reason for verifying identities. This portion of the screening was crucial in the detection of fraudulent applicants for the current trial as researchers noted some applicants in this process who sounded suspiciously similar and too young for the study.
3. Implement contingencies for participant payment. Making payment contingent on both completing study requirements and providing proof of identity may reduce fraudulent behavior. In addition, mailing physical gift cards instead of emailing them digitally might ensure participants are providing a legitimate and verifiable mailing address in order to receive compensation. However, this may have unintended consequences to participant privacy and needs to be considered carefully.
4. Use survey protection features. Researchers should consider enabling any available fraud reduction tools. Qualtrics has a particularly useful tool, "ballot-box stuffing," which helps reduce the likelihood of repeat participants by using HTTP cookies (small pieces of data placed by the website onto the user's computer) that detect browser and device information. Survey tools often offer other helpful features, including ways to prevent indexing, tracking IP address and geo-location, bot-detection, relevantID technology (which detects potential fraud or duplicate participants), and requiring a passcode or unique link in order to participate.
5. Supply ample information. This study explicitly focused on older adults, a population often preyed upon through Internet and phone scams. Many applicants who were contacted to verify their information reported that they were uncomfortable providing personal information on an Internet form without being certain that it was a trusted website. For this reason, they admitted to modifying bits of

information (e.g. used a maiden name, made a slight change in their birth date, etc.) as a way of protecting their identity. This can be problematic for online studies such as this, where verified personal information was crucial to enrolling eligible and unique applicants. As a way to address this, ample information was provided about the research study team and institution. When applicants learn more about the research team and trial, the more likely they may be to provide accurate information because they feel confident the study is not a scam designed to steal their identities.

6. Consider resources needed and plan to fund costs of identity verification. It is important to be sure the study includes the necessary funds and time to implement an appropriate applicant identification process. The TLOxp checks cost this study \$1 each and approximately four minutes of staff time to process each application. This can translate to considerable costs for larger trials. The current trial spent an estimated \$1920 and 7680 min (128 h) conducting checks.
7. Maintain detailed records. It is very important to maintain detailed records of the status of all applicants and the reasons for identity check failures. Researchers should, at a minimum, record the following criteria: date of verification check, criteria being checked (name, date of birth, address, phone, email etc.), outcome of each of those criteria (verified or unverified), overall decision made about each applicant (valid or fraudulent), and a summary/explanation of the decision and reasoning. Using a spreadsheet to document this information will greatly help research teams be consistent in inclusion/exclusion decisions and when documenting recruitment outcomes in papers and reports.
8. Share findings. Researchers must be aware of potential fraud and report their study findings accordingly. This means researchers must also acknowledge the risk of fraudulent participants in their samples and implications about the validity of their findings if identities were not verified. By sharing fraud related data, improvements can be made across the field to limit this potentially harmful issue.

#### 4.2. Conclusion

In a large research trial of an Internet intervention for insomnia for individuals at least 55 years of age, approximately 7% of applicants provided fraudulent information in order to meet eligibility. Participants included men and women from across the US using a variety of recruitment methods. Fraud, which included providing false names, dates of birth, and addresses, was detected using TransUnion's TLOxp service as an identity verification tool. Although fraud has been found to be an issue in online survey research, where there may be fewer research requirements and payment is relatively immediate, this is one of the first studies to examine applicants for potential fraud in a longitudinal trial of an online intervention. Fraudulent applications came from multiple recruitment sources, suggesting that there is no ideal recruitment source; however, a greater percentage of fraudulent applications came from those who identified online advertising methods as the referring source. The rate of potential fraud associated with each recruitment source should be further investigated with other study populations. Verifying identities is critically important to the integrity of online research trials and must be considered and addressed in future trials.

#### Funding

Research reported in this publication was supported by the National Institute On Aging of the National Institutes of Health under Award Number R01AG047885. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

#### Declaration of competing interest

Ms. Glazer, Ms. MacDonnell, Ms. Frederick, and Dr. Ingersoll have no

conflicts of interest to declare. Dr. Ritterband reports having a financial and/or business interest in BeHealth Solutions and Pear Therapeutics, two companies that develop and disseminate digital therapeutics, including by licensing the therapeutic developed, based in part, on early versions of the software utilized in research reported in the enclosed paper. These companies had no role in preparing this manuscript. Dr. Ritterband is also a consultant to Mahana Therapeutics, a separate digital therapeutic company not affiliated with this research. The terms of these arrangements have been reviewed and approved by the University of Virginia in accordance with its policies.

## Acknowledgements

We are grateful to Abigail Williams for her assistance with this research.

## References

- Adam, L.M., Manca, D.P., Bell, R.C., 2016. Can Facebook be used for research? Experiences using Facebook to recruit pregnant women for a randomized controlled trial. *J. Med. Internet Res.* 18 (9), e250 <https://doi.org/10.2196/jmir.6404>.
- Anderson, M., Perrin, A., Jiang, J., Kumar, M., 2019. 13% of Americans don't use the internet. Who are they. *Pew Research Center* 7 (Retrieved October 29, 2019, from). <https://www.pewresearch.org/fact-tank/2019/04/22/some-americans-dont-use-the-internet-who-are-they/>.
- Ballard, A.M., Cardwell, T., Young, A.M., 2019. Fraud detection protocol for web-based research among men who have sex with men: development and descriptive evaluation. *JMIR Public Health Surveill.* 5 (1), e12344 <https://doi.org/10.2196/12344>.
- Ballard, A.M., Haardörfer, R., Prood, N., Mbagwu, C., Cooper, H.L., Young, A.M., 2020. Willingness to participate in at-home HIV testing among young adults who use opioids in rural appalachia. *AIDS Behav.* 1–10. <https://doi.org/10.1007/s10461-020-03034-6>.
- Bauermeister, J.A., Pingel, E., Zimmerman, M., Couper, M., Carballo-Dieguez, A., Strecher, V.J., 2012. Data quality in HIV/AIDS web-based surveys: handling invalid and suspicious data. *Field Methods* 24 (3), 272–291. <https://doi.org/10.1177/1525822x12443097>.
- Bowen, A.M., Daniel, C.M., Williams, M.L., Baird, G.L., 2008. Identifying multiple submissions in internet research: preserving data integrity. *AIDS Behav.* 12 (6), 964–973. <https://doi.org/10.1007/s10461-007-9352-2>.
- Bull, S., Pratte, K., Whitesell, N., Rietmeijer, C., McFarlane, M., 2009. Effects of an internet-based intervention for HIV prevention: the Youthnet trials. *AIDS Behav.* 13 (3), 474–487. <https://doi.org/10.1007/s10461-008-9487-9>.
- Carter-Harris, L., Ellis, R.B., Warrick, A., Rawl, S., 2016. Beyond traditional newspaper advertisement: leveraging Facebook-targeted advertisement to recruit long-term smokers for research. *J. Med. Internet Res.* 18 (6), e117 <https://doi.org/10.2196/jmir.5502>.
- Cowie, J.M., Gurney, M.E., 2018. The use of Facebook advertising to recruit healthy elderly people for a clinical trial: baseline metrics. *JMIR Res. Protocol.* 7 (1), e20 <https://doi.org/10.2196/resprot.7918>.
- Fichten, C., Libman, E., Bailes, S., Alapin, L., 2000. Characteristics of older adults with insomnia. In: Lichstein, K., Morin, C.M. (Eds.), *Treatment of Late-Life Insomnia*, pp. 37–80.
- Food and Drug Administration, 2016. *Use of Electronic Informed Consent in Clinical Investigations Questions and Answers Guidance for Industry, Draft Guidance*. Silver Spring, MD.
- Frandsen, M., Walters, J., Ferguson, S.G., 2013. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. *Nicotine Tob. Res.* 16 (2), 247–251. <https://doi.org/10.1093/ntr/ntt157>.
- Granja, E.G., Janssen, W., Johansen, M.A., 2018. Factors determining the success and failure of eHealth interventions: systematic review of the literature. *J. Med. Internet Res.* 20 (5), e10235 <https://doi.org/10.2196/10235>.
- Inan, O.T., Tenaerts, P., Prindiville, S.A., Reynolds, H.R., Dizon, D.S., Cooper-Arnold, K., Marlin, B.M., 2020. Digitizing clinical trials. *NPJ Digital Medicine* 3 (1), 1–7. <https://doi.org/10.1038/s41746-020-0302-y>.
- Kayrouz, R., Dear, B.F., Karin, E., Titov, N., 2016. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. *Internet Interv.* 4, 1–10. <https://doi.org/10.1016/j.invent.2016.01.001>.
- Lattie, E.G., Kaiser, S.M., Alam, N., Tomasino, K.N., Sargent, E., Rubanovich, C.K., Mohr, D.C., 2018. A practical do-it-yourself recruitment framework for concurrent eHealth clinical trials: identification of efficient and cost-effective methods for decision making (part 2). *J. Med. Internet Res.* 20 (11), e11050 <https://doi.org/10.2196/11050>.
- MacDonnell, K., Cowen, E., Cunningham, D.J., Ritterband, L., Ingersoll, K., 2019. Online recruitment of a non-help-seeking sample for an internet intervention: lessons learned in an alcohol-exposed pregnancy risk reduction study. *Internet Interv.*, 100240 <https://doi.org/10.1016/j.invent.2019.100240>.
- Marcolino, M.S., Oliveira, J.A.Q., D'Agostino, M., Ribeiro, A.L., Alkmim, M.B.M., Novillo-Ortiz, D., 2018. The impact of mHealth interventions: systematic review of systematic reviews. *JMIR mHealth and uHealth* 6 (1), e23. <https://doi.org/10.2196/mhealth.8873>.
- Ritterband, L.M., Thorndike, F.P., Gonder-Frederick, L.A., Magee, J.C., Bailey, E.T., Saylor, D.K., Morin, C.M., 2009. Efficacy of an internet-based behavioral intervention for adults with insomnia. *Arch. Gen. Psychiatry* 66 (7), 692–698. <https://doi.org/10.1001/archgenpsychiatry.2009.66>.
- Ryan, C. & Lewis, J.M. (2017). . United States Census Bureau. Retrieved October 29, 2019 from <https://www.census.gov/content/dam/Census/library/publications/2017/acs/acs-37.pdf>.
- Schure, M.B., Lindow, J.C., Greist, J.H., Nakonezny, P.A., Bailey, S.J., Bryan, W.L., Byerly, M.J., 2019. Use of a fully automated internet-based cognitive behavior therapy intervention in a community population of adults with depression symptoms: randomized controlled trial. *J. Med. Internet Res.* 21 (11), e14754 <https://doi.org/10.2196/14754>.
- Teitcher, J.E., Bockting, W.O., Bauermeister, J.A., Hoefler, 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), 116–133 (doi:10.1111%2Fjlme.12200).
- Topolovec-Vranic, J., Natarajan, K., 2016. The use of social media in recruitment for medical research studies: a scoping review. *J. Med. Internet Res.* 18 (11), e286 <https://doi.org/10.2196/jmir.5698>.
- Wessling, K.S., Huber, J., Netzer, O., 2017. MTurk character misrepresentation: assessment and solutions. *J. Consum. Res.* 44 (1), 211–230. <https://doi.org/10.1093/jcr/ucx053>.
- Young, A.M., Ballard, A.M., Cooper, H.L., 2020. Novel recruitment methods for research among young adults in rural areas who use opioids: cookouts, coupons, and community-based staff. *Public Health Rep.* 135 (6), 746–755. <https://doi.org/10.1177/0033354920954796>.