2024 ■ Vol. 20 ■ no. 2

Dissecting the Onion: Identifying and Remediating Issues Surrounding Data Integrity in Online Survey Research

Xen Modrakovic [a] ✉, Cheriko A. Boone [a], David A. Kalwicz [a], Sharanya Rao [a], Benjamin Parchem [a,b], Natalie M. Wittlin [c,d], Viraj V. Patel [e], Manya Magnus [a], Maria Cecilia Zea [a], Michael Kharfen [f], John F. Dovidio [c] & Sarah K. Calabrese [a]

[a]George Washington University
[b]University of Minnesota Medical School
[c]Yale University
[d]Princeton University
[e]Albert Einstein College of Medicine, Montefiore Health System
[f]New York State Department of Health

Abstract ■ In this non-empirical article, which is intended as a decision-making resource for researchers, we identify issues surrounding data integrity that commonly arise in online survey research and we propose remediation strategies based on challenges we encountered during a particular pilot study as well as our collective experience with conducting online survey research. Using the metaphor of an onion, we peel off the layers of this complex problem, synthesize the various available strategies used across disciplines, and propose some novel ones based on our perspective as psychologists. Corresponding to this multi-layered problem, we propose multi-layered solutions to prevent illegitimate responding—by both humans and non-humans (robots or "bots" for short)—from compromising the quality of data collected via online survey research. The first layer entails strategic item selection and protective programming in survey development. The second layer involves astute advertising and recruitment tactics to minimize illegitimate responses during survey dissemination. The third layer includes algorithms and other mechanisms to identify suspicious responses for possible exclusion during data verification. When we peel off the layers and reach the core problem of illegitimate responses to online surveys—financial incentives—we will propose ways of navigating respondent reimbursement to mitigate their inadvertent harmful impacts on the research process. By proposing these solutions, we aim to protect the integrity of scientific inquiry in psychology, especially given how often this method is used in the discipline.

Keywords ■ Online Surveys, Data Integrity, Survey Development, Respondent Incentives, Bots.

✉ djmodrakovic@gwu.edu

10.20982/tqmp.20.2.p076

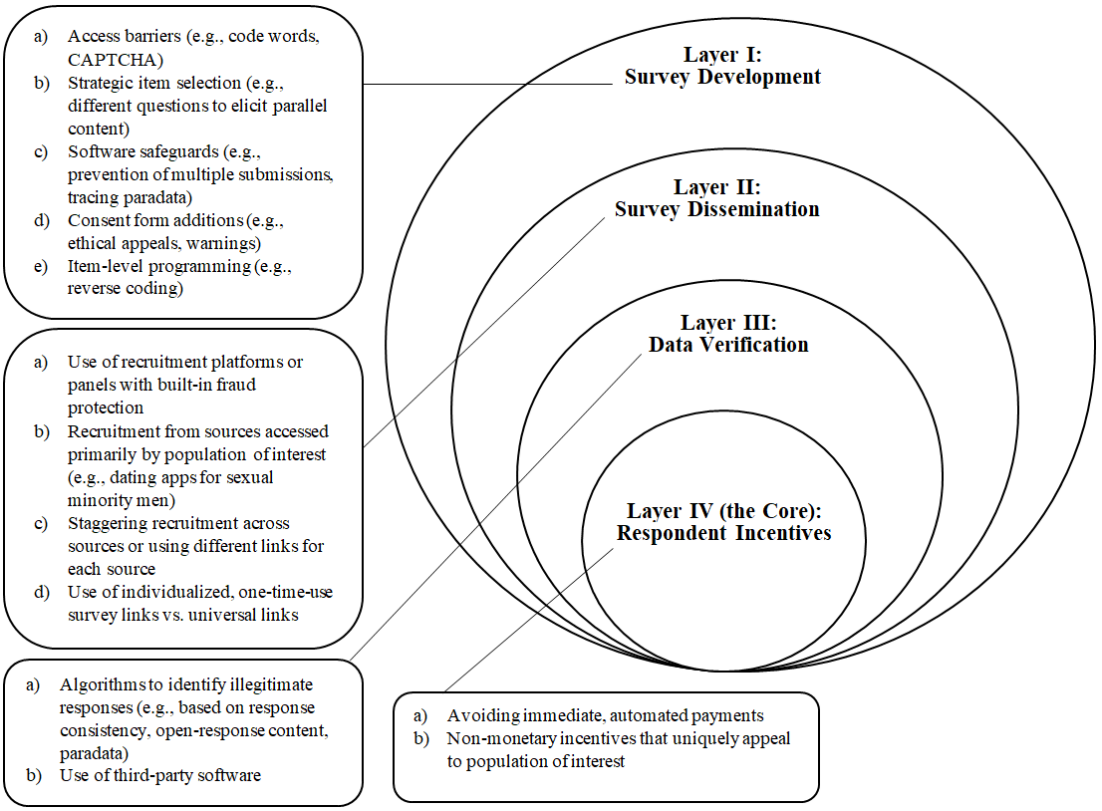
Acting Editor ■ Sébastien Béland (Université de Montréal)

Reviewers ■ Two anonymous reviewers

Introduction

The COVID-19 pandemic demanded unprecedented changes to research protocols and procedures, prompting many researchers to pivot away from in-person and toward online formats of data collection, including the use of online surveys (Saberi, 2020). Online surveys have gained considerable momentum since the early 2000s because of their various advantages, such as affording researchers access to respondents in distant locations; enhancing anonymity, and hence comfort, of populations that are harder to reach due to stigma; decreasing social desirability bias; and enabling convenient, automated data collection procedures that save researchers' resources and reduce human error. The COVID-19 pandemic stimulated a significant interest in the utilization of online data collection, which has continued to remain high because of the advantages it offers. Despite these advantages, online surveys

Figure 1 ■ Dissecting "the Onion" of Online Survey Research. The circles represent layers of the onion, each of which is a stage in online survey research that is vulnerable to threats to data integrity. We recommend multiple solutions to consider at each stage.



pose unique challenges in terms of data integrity (Andrade, 2020; Storozuk et al., 2020; Wright, 2005). Namely, the online format and associated anonymity introduce potential for illegitimate survey responses. By "illegitimate survey responses", we mean: (a) responses provided by people who had previously been eligible for and completed a survey and then opt to retake it (despite being ineligible to take it more than once); (b) responses provided by people who were never eligible, but take a survey one or more times; or (c) responses by non-human respondents or "bots," short for robots, or increasingly sophisticated software applications that perform automated tasks over the Internet at a quicker pace than is humanly possible (Simone, 2019, 2020; Storozuk et al., 2020; Teitcher et al., 2015). Inclusion of illegitimate responses in online survey datasets can yield faulty, unscientific conclusions, which can result in development and implementation of ineffective or otherwise unsound interventions. Given the growing popularity of and reliance upon online survey research methods, accelerated by public health guidelines around social and physical distancing needs during the COVID-19 pandemic and corresponding human subjects protections, it is becoming increasingly necessary to integrate methods for ensuring integrity of data collected online, which is essential for protecting the integrity of scientific inquiry more broadly. We contend that the process of ensuring the integrity of online survey research is analogous to an onion in two ways—it is multilayered, and it has the potential to make a researcher cry. However, just like effective strategies have been developed to prevent tears when handling real onions, there are concrete strategies that can help researchers handle the metaphoric onion as well.

In the present article, which is intended as a decision-making resource for researchers, we identify risks to data integrity and propose mitigating approaches at multiple stages of the online survey research process. Each stage can be understood as a layer of the onion and corresponds with a structural unit of the paper: Layer I or the survey design and programming stage (i.e., Survey Development), Layer II or the respondent recruitment and survey admin-

istration stage (i.e., Survey Dissemination), Layer III or the data cleaning stage (i.e., Data Verification), and Layer IV (the Core) or the respondent reimbursement stage (i.e., Respondent Incentives). We base our recommendations on our experience with a particular pilot project, which we refer to as a case example, as well as our collective experience conducting online survey research. Although in this paper we conceptualize the online survey research process as comprising several distinct stages (corresponding to distinct layers of the onion), we do not view them as mutually exclusive. Rather, we conceptualize them as overlapping with synergistic impacts across stages, operating through an iterative process. For example, upon learning during data cleaning (Layer III: Data Verification) that incoming survey responses were originating in other countries (despite a self-report screening item requiring a US location), we implemented new software-based geographic controls (Layer I: Survey Development) that prevented further fraudulent responding in incoming data (Layer II: Survey Dissemination).

In the evolving world of information technology (IT), online surveys face multilayered threats to data integrity and thus, require a multilayered system of protection, which we characterize using the metaphor of an onion. To protect data integrity in survey research, we will briefly outline the challenges that might arise at each layer of "the onion" and propose ideas to surmount them. When addressing Layer I, we expound upon item selection strategies and IT protections that researchers can introduce during survey development to filter out or prevent illegitimate responses. At Layer II, we discuss how to implement astute advertising and recruitment tactics during survey dissemination to minimize illegitimate response attempts. At Layer III, we propose algorithms and other mechanisms to utilize during data verification to identify suspicious responses for possible exclusion. As we peel off the layers of the onion, we will reach what we propose to be the core of the problem. Namely, we describe how financial incentive structures or respondent compensation might precipitate or encourage illegitimate responses to online surveys, as well as how to reduce their inadvertent harmful impacts on the research process. We visually represent "the onion problem" of online survey research in Figure 1, with each layer entailing unique protections to help filter out illegitimate responses. Building on the solutions to the problem of data integrity proposed in previous literature (Simone, 2019, 2020; Levi et al., 2022; Storozuk et al., 2020; Teitcher et al., 2015), our paper integrates existing solutions and novel ones within a larger conceptual framework, thereby offering a more complete view of the available solutions to date and considerably expanding the resource pool.

Case Example: Challenges Experienced During the PrEP'ing DC Project

Our pilot project titled "PrEP'ing DC Project: Optimizing PrEP Social Marketing to Black MSM in DC to Raise Awareness and Uptake" was designed to investigate attitudes related to HIV pre-exposure prophylaxis (i.e., medication that prevents HIV acquisition) among sexually active, HIV-negative Black men who have sex with men (MSM) living in the Washington, DC metropolitan area.[1] As part of the project, we recruited participants via dating apps to complete a baseline and an 8-week follow up survey, both of which were administered online via Qualtrics software (Qualtrics, Provo, UT). Notably, of the 719 participant responses recorded, each of which should have corresponded to a unique respondent who consented to participate in the baseline survey, we deemed only 13% to be legitimate responses. For the eight-week follow-up survey, there were even fewer legitimate responses (11% of the original total). While implementing the project, we experienced several waves of fraudulent responses that interrupted the data collection process and required notifying the university's institutional review board (IRB) and devising plans and algorithms to address the issue. We identified these waves of fraudulent responses through regular survey monitoring during data collection and immediately paused surveys upon detecting suspicious responses, changing the survey link before resuming data collection. During team meetings, we jointly discussed suspicious responses and troubleshooted our survey design to improve prevention and remediation strategies. As a result of these meetings, we introduced changes to survey programming, survey dissemination, and data verification following survey breaches. As learned, resolving a multilayered problem require swift and sustained intervention with a multilayered system of protection.

Layer I: Survey Development

There is a variety of software and online programs available to program and host online surveys. In addition to Qualtrics (Qualtrics, Provo, UT), the software we used in our case example, and its open-source counterpart, LimeSurvey (LimeSurvey GmbH, n.d.), others include REDCap, SurveyMonkey, Alchemer, TypeForm, SurveyKing, and QuestionPro. While LimeSurvey has a simpler, user-friendly interface, it might lack some of the sophisticated data verification features Qualtrics offers. These features, however, are constantly evolving and may be offered on LimeSurvey and other platforms in the future, so we encourage readers to explore options considering these and other features discussed below.

[1] Our pilot project was funded by District of Columbia Center for AIDS Research (DC CFAR) and conducted at The George Washington University.

Projects involving online survey research may differ in aims, methods, sampling frames and hypotheses, but there are often similarities in terms of survey structure. Many online surveys include an initial welcome page and prompt potential respondents to complete a series of screening items to determine eligibility. Survey software such as Qualtrics offers programming functions that can automatically notify individuals who do not meet eligibility criteria that they are ineligible and prevent their further participation (Qualtrics, Provo, UT). Those respondents who are eligible to participate are typically presented with a general overview of the study, a description of associated risks and benefits, and other information required in order to provide their informed consent for online participation. Once consented, respondents proceed to complete the survey. Display logic; skip logic; branching; and randomization of items, response options, or sections ("blocks") are some of the basic survey programming features used to customize each respondent's online survey experience. After completing the survey, respondents may be given further instructions about obtaining compensation for their participation or being contacted in the future (e.g., if repeating survey across multiple time points). Additionally, they may be offered online resources (e.g., links to local healthcare organizations), provided with the study team's contact information, and debriefed as needed for experimental survey studies.

A few programming strategies can be set up at the outset of the survey to prevent illegitimate responders, help to identify them, and/or mitigate the damage done by them. A standard barrier to protect against fraudulent responders is CAPTCHA (i.e., Completely Automated Public Turing Test to Tell Computers and Humans Apart) in the form of a code respondents need to enter (Levi et al., 2022; Grov et al., 2019; Von Ahn et al., 2003). CAPTCHA commonly requires respondents to type letters or numbers from a distorted image that a bot would not be able to decipher (Levi et al., 2022). However, a potential disadvantage of this strategy is that it could unintentionally exclude people with low computer literacy or people with visual disabilities (Teitcher et al., 2015). Another programming option to mitigate illegitimate responding is integration of a code word at the beginning of a survey. Entry of this code word can subsequently be requested immediately after introducing it to survey respondents or at any later part of the survey to verify that an attentive human is taking it (e.g., "When asked for your favorite color, you must enter the word 'sage' in the text box below… Based on the instructions you read above, what is your favorite color?"). Utilizing this strategy might require introduction of a relatively simple code word that the respondents are likely to remember even if the instructions and follow-up question are spaced out. In our pilot project, we embedded the instructions and follow-up question given above in our eligibility screening items, and participants who did not enter "Sage" or "sage" in the response textbox were automatically screened out. Another strategy to protect survey access that has been recommended by other researchers is the use of SMS verification or email verification by prompting respondents to enter a phone number or email to receive a code (i.e., "one-time password," "one-time PIN", "one-time authorization code"), which they would enter into the survey to proceed (Grov et al., 2019). However, this strategy risks deterring respondents who do not feel comfortable disclosing such contact information from proceeding with the survey.

Besides use of a CAPTCHA and code word, researchers can strategically incorporate survey items and response options that will help to later identify illegitimate respondents during the data verification stage (Levi et al., 2022; Simone, 2019; Storozuk et al., 2020). Inclusion of reverse-coded items within multi-item measures employing Likert scales is a prominent integrity-enhancing tool for survey research (Storozuk et al., 2020; Weijters & Baumgartner, 2012). Reverse coding entails re-coding the responses so that the numerical values assigned to the response options when scoring the Likert scale run in the opposite direction than other scale items. Logical discrepancies in responses to regularly coded items and reverse-coded items (e.g., strong agreement with "I am often anxious" and strong agreement with "I am rarely anxious") can signal fraudulent responding. Besides protecting against fraudulent responders, this strategy can enhance data quality by helping to identify inattentive legitimate responders. However, some evidence suggests reverse coding may impact measure reliability in the sample when mixed stems are used (i.e., positively and negatively worded items) because such items might not always measure the same underlying trait. Thus, we would not recommend using discrepancies on these types of items as sole indicators of illegitimate responding (Weems & Onwuegbuzie, 2001).

Researchers can also embed questions at different points in a survey that are designed to elicit parallel content and then check for inconsistencies during post-hoc analyses (Storozuk et al., 2020). For example, a question about respondents' age and another about their year of birth, or a question about whether respondents have ever had sex and another abou their number of sex partners, can later be assessed for discrepancies. Researchers can also include one or more multiple choice item(s) with response options that are not obviously wrong to the respondent but will invite suspicion of the researchers if selected. For example, in our pilot study, we asked participants how they were recruited into the study and included response options for dating apps that we never used as recruitment sources.

While this strategy may help to screen out humans who are gaming the survey, it would not necessarily be a foolproof way to screen out all bots; hence, as with other strategies, researchers might need to use this strategy in combination with others for maximal effectiveness.

To create even more stringent item-level survey safeguards, researchers might want to consider implementing techniques that have been used to measure validity in psychological assessments. For example, on the Behavior Rating Inventory of Executive Function (BRIEF; Gioia et al., 2005), there are numerical scales that indicate validity based on response infrequency (i.e., expressing agreement with unusual items or items that rarely anyone would agree with, e.g., "I never lie"), negativity (i.e., overdramatizing responses), and inconsistency (i.e., a pattern of selecting a different item answer to semantically equivalent items with slightly altered phrasing). Imbedding such items and using psychometrics to derive validity scales for each respondent could be a more objective method of identifying potential illegitimate responses. However, the same validity flags could also be raised by legitimate respondents who are fatigued, distracted, and/or unmotivated, which researchers would need to consider when deciding whether to use such flags as a basis for exclusion.

The item-level strategies to prevent and identify illegitimate responding that we have described are imperfect. Some can be circumvented by sophisticated bots. Others might raise suspicion about a particular response record without definitively determining its illegitimately. Thus, we recommend using a combination of strategies simultaneously. However, to the extent that each strategy lengthens the survey, inclusion of such strategies must be balanced against the risk of survey fatigue that could result and lead to respondents prematurely discontinuing the survey. Piloting the survey before formally releasing it can help researchers evaluate whether the strategies implemented are overly burdensome to respondents and whether items included to screen for legitimacy have the potential to confuse, alienate, or frustrate legitimate survey takers.

There are specific safeguard options offered by online survey software that researchers can select during survey programming, such as the "Prevent Multiple Submissions" option in Qualtrics (formerly labeled "No Ballot Box Stuffing") or an equivalent option within other survey software. This line of defense protects against respondents who were eligible to take the survey, have already taken it once and are therefore no longer eligible, but nonetheless try to take the survey again (Grov et al., 2019; Nash et al., 2019). The "Prevent Multiple Submissions" option and its equivalents operate by embedding a cookie in a respondent's browser when they complete a survey for the first time, which prevents the survey from freshly loading again on the respondent's device (instead presenting a customizable message or redirecting them to another website) or allows them to repeat the survey but flags such response records as repeats. However, a limitation of this strategy is that people can circumvent this prevention method either by clearing browser cookies from their devices, completing the survey using a different browser, using a different device altogether, or browsing in private mode (Nash et al., 2019; Teitcher et al., 2015). Novel programming strategies available on Qualtrics can detect fraud and bots, and the available protective measures continue to evolve. For example, "Expert Review Fraud Detection" and "Bot Detection" track patterns of data to infer the likelihood that a response was completed by a bot. Furthermore, "Adding Fraudulent Detection Fields to the Survey Flow" even performs automated analyses throughout the survey flow to detect and screen out bots and fraudulent responses. Additional security features are available through Qualtrics (e.g., Security Scan Monitor, RelevantID), and many of these options need to be set up before data collection to function. Some security features are not automatically included in all Qualtrics licenses and may need to be specifically added (Qualtrics, 2018).

Online survey platforms such as Qualtrics and its open source counterpart LimeSurvey also track paradata, or data about the process by which the survey data were collected. These can include timestamps for when a respondent starts and completes a survey, a record of how long it takes a respondent to complete a survey, and information about where a survey is completed (Couper, 1998; Storozuk et al., 2020). Researchers can use paradata to restrict survey access by location (e.g., GeoIP Location in Qualtrics). Survey software typically uses a respondent's Internet Protocol (IP) address, a numerical label associated with each of the devices connected to the same computer network, to determine the location of their device, which researchers can use to exclude respondents outside of a particular geographic area. However, this strategy is vulnerable to the use of a Virtual Private Networks (VPN), which can disguise the original IP address and allow a survey respondent to appear to be located in a place different than where they are actually physically located. When researchers intend to collect data from a specific region, it is important to keep in mind that eligible respondents could be traveling and completing the survey from abroad, which could be the reason why, for example, an international IP address is recorded despite the individual reporting a local address. Software programs have been developed to help researchers identify IP addresses that are likely using a VPN (Waggoner et al., 2019), and researchers can consider adding a stipulation that respondents not use a VPN in the

survey welcome page or screening process. Third-party software may also help screen out illegitimate respondents at the outset of the survey (Winter et al., 2019). However, to ensure protection of respondent confidentiality, we advise researchers to review the policy on data storage when using such third-party software. Strategic use of paradata can also leveraged during the data verification stage, as discussed later. Importantly, because paradata can include information that could be traced to a participant's identity (e.g., IP address, precise geographic location), collection of paradata needs to be disclosed and justified to the institutional review board overseeing the research and should be deleted once data verification (cleaning and checking) has been completed.

There are a number of strategies that researchers can incorporate in the consent process to discourage illegitimate responding. Some survey respondents have altruistic motivations to participate in research (Dubé et al., 2020), and researchers can appeal to these motivations and respondents' sense of morality through explicit communication regarding expectations, which can be a strategy for preventing undesired responding behavior. For example, researchers can integrate an honor code statement during the informed consent process that appeals to respondents' sense of morality about research participation and reminds them that repeat responding or other fraudulent behavior compromises data integrity and resultant scientific conclusions; however, respondents already contemplating fraudulent responding may not be especially responsive to such tactics. Given that a driving motivation for illegitimate responding is often financial compensation, researchers can highlight in the consent form that each participant will be compensated only once. Likewise, communication to prospective respondents can include explicit warnings about consequences of suspected fraud (e.g., compensation withheld) and the types of behaviors that qualify as fraud (Teitcher et al., 2015). Including a statement about monitoring participants' use of VPN could act as a deterrent as well. In combination with other survey programming strategies in Layer I, embedding such disclaimers in consent forms may deter fraudulent responses from human participants or discourage use of bots for the purposes of gaming a research project for financial gain. Development and use of such disclaimers should be based upon consultation with the IRB and community consultation with the population of interest. Researchers should exercise extreme caution about acting on the stated penalties for suspected fraudulent behavior (e.g., compensation withheld) to ensure they are not mistakenly exacted upon legitimate respondents. We will further discuss participant incentives in a later section.

Layer II: Survey Dissemination

Researchers can implement a number of fraud-deterring strategies during the recruitment stage of a project. In this section, we will briefly outline some of these strategies and discuss associated advantages and disadvantages. One of the basic considerations regarding survey dissemination is limiting access to a survey link (Simone, 2019, 2020). If researchers opt to use a single link through which all respondents will universally access a survey, then it should be kept relatively private to prevent ineligible respondents from accessing the survey. However, once the link to a survey has been made public, it is difficult to limit how widely it travels online.

Recruiting through crowdsourcing sites that feature mechanisms to prevent fraudulent responders, such as Amazon Mechanical Turk (MTurk), Qualtrics Panels, and Prolific, is one solution. However, there are tradeoffs associated with these platforms. For example, respondents available through these platforms might be more technologically savvy than others or have other prominent characteristics that render them nonrepresentative of the population of interest. Previous research has suggested that MTurk respondents have higher negative affect and lower social engagement than the general public (McCredie & Morey, 2019). Some platforms may have features that accommodate greater population specificity for recruitment of populations based upon intersecting axes of identity (e.g., Black MSM, as with our project), whereas others might lack this feature. Cost is another consideration related to using these platforms, as different platforms charge researchers at different rates. Given that using online survey platforms featuring data integrity safeguards may limit study feasibility and/or generalizability, the decision to use such a platform warrants careful consideration. We must add that, despite mechanisms for survey protection associated with these platforms, these protections are not 100% effective, as survey fraud still occurs (Ahler et al., 2021; Kennedy et al., 2020).

Researchers might want to consider exclusively or primarily advertising their study through recruitment sources with an audience or patronage that mostly consists of the population who are eligible to participate in a survey. Examples of such sources include closed and/or moderated online groups and forums as well as organization listservs that cater to the researchers' population of interest. For example, given that our focal population was MSM in the PrEP'ing DC Project, we restricted our advertising to apps and social media that were specific to the LGBTQ+ population or that catered predominantly to MSM.

Respondent-driven sampling (RDS) is another useful recruitment strategy that could increase the number of legitimate responders to a survey. RDS is a nonprobability sam-

pling technique whereby respondents recruit other individuals from their networks (as defined by the study) who may be interested in participating and are generally perceived within their social networks to be eligible to do so (Heckathorn, 1997). Researchers typically compensate respondents for their referral of additional participants. Although RDS may reduce illegitimate responding, projects with smaller budgets may be unable to accommodate such expenses. Additionally, this strategy may yield recruitment of a niche, homogenous sample instead of a representative sample of the target population of interest for a particular study, thus reducing external validity.

When running a study with multiple recruitment sources, staggering the timing of recruitment across different sources, coupled with tracking respondents' recruitment source reported in response to a survey item, can be useful for identifying waves of bots and other illegitimate responses. This was a lesson learned early in our pilot study, when we had a sudden influx of suspicious responses following recruitment launched through multiple sources simultaneously, which made it impossible to discern the specific recruitment source(s) that instigated the problem. Subsequently, we initiated recruitment through one source at a time. When an influx of new responses associated with advertising only on Dating App A dwindled, we stopped advertising on Dating App A and switched to only advertising on Dating App B. Consequently, observing suspicious survey responses that newly emerged in our data allowed us to be fairly confident that Dating App B was responsible, especially because Dating App B was commonly reported as the recruitment source within the suspicious response records. Staggered advertising in this way allowed us to make informed decisions about halting recruitment from any specific source. If staggering is not feasible (e.g., due to the study timeline or number of recruitment sources needed), researchers could consider creating multiple copies of the online survey, each with its own link, and then designating a different recruitment link for each recruitment source. Thus, if suspicious response records appear within a given survey, the problematic recruitment source can be readily identified and intervened upon without interfering with ongoing recruitment through other sources. This strategy would require the subsequent merging of survey datasets.

An alternative to a universal survey link disseminated through one or multiple recruitment sources is to generate unique, one-time-use survey links for each respondent, a feature of some online survey platforms that can be advantageous in limiting damage from the continued use of a single compromised survey link. However, dissemination of individual links may increase a researcher's involvement and require additional resources (e.g., coordinated emailing of an individualized link to each respondent). It may also require prospective respondents to expend additional effort and engage in unwanted interaction with the research team, such as emailing to express interest in the study and request a link, and disclosing their email address or other contact information to receive their individualized link. This required effort and engagement could deter participation.

Researchers could also consider video or audio screening to filter out illegitimate responders (Teitcher et al., 2015). An example of this would be recruiting with advertisements that require prospective respondents to contact the research team via email/phone to determine eligibility. Research team members would then conduct screening via phone or videoconference and, if respondents are deemed eligible, send them the unique survey link. The additional step requiring conversation with a human being could protect the survey from bots. A similar alternative could be recruiting with advertisements linking to a welcome page that then requires video- or chat-based verification before participants could advance within the survey. Although these could be effective ways of ensuring a legitimate sample, they also involve disadvantages that are important to consider. One disadvantage is the inconvenience of adding an extra step in the process, which might be cumbersome and could deter legitimate prospective respondents from participating in the study. This is particularly likely to be true if it causes a significant delay between the time a respondent initially expresses interest by responding to the advertisement and the time they are granted access to the survey, which may be inevitable if data are collected on a continuous basis and monitoring the survey and interacting with participants at all times (e.g., overnight) is not feasible for the research team. A second disadvantage is that an additional screening step could unintentionally exclude people with low computer literacy or technology-based limitations (e.g., computer without camera). A third disadvantage is that the screeners could reduce perceived anonymity for potential respondents. In psychology, we often conduct research on sensitive and stigmatized topics (e.g., HIV, sex, trauma) and conduct research with vulnerable populations; compromising anonymity to improve data integrity could deter participation and pose a risk to participants. Additionally, researchers might require additional resources to implement these measures, such as funding to support a research assistant being on call to conduct the video screening. These advanced screening measures are also not foolproof because people who are dishonest on a survey may have no qualms about being dishonest during a video or phone screening session, and some eligibility criteria may depend on self-report and may not be easily verified by the person conducting the screening (e.g., recent sexual activity).

Layer III: Data Verification

The data verification layer includes recommendations for data cleaning and checking patterns of responses to identify and exclude illegitimate responders in the dataset from the final analytic sample. Qualtrics offers a data verification option – "Expert Review Fraud Detection" – that can help weed out illegitimate responses during data collection based on automated analysis of multiple or suspect submissions. Using this option, researchers can discard fraudulent responses, preventing them from being counted against auditable responses or quotas, or redirect these responses for analysis separately. At present, this feature is not yet available on LimeSurvey, so researchers would need to complete a manual screen for suspect responses. When performing data verification independently and manually, we advise researchers to create algorithms or standardized procedures for data cleaning and identifying likely fraudulent responses. In our experience with the pilot project, there were multiple survey records in which we found one or more indications of possible fraudulence described below. However, even then, these survey records were not deemed definitively fraudulent. For these ambiguous survey records, we generally erred on the side of caution in excluding them from the analytic sample (or set criteria for exclusion, such as response records with one "red flag" or two "pink flags"), but nonetheless compensated these potential respondents to avoid undercompensating actual respondents.[2]

Checking for consistency within survey response records is one way of verifying legitimacy. In Layer I, we suggested asking questions designed to elicit parallel content in multiple ways (e.g., year of birth and age). Accordingly, in Layer III, researchers can check for consistency of responses to those questions within each response record (Simone, 2019; Levi et al., 2022). Additionally, unusual patterns within response records (e.g., 1, 2, 3, 4, 1, 2, 3, 4) as well as nonsensical responses to open-ended survey questions (e.g., random letters/digits where words would be expected) are cause for concern. Using data analytic software might facilitate the seemingly cumbersome process of identifying such unusual patterns and nonsensical responses.

Whereas inconsistency in item responses within a survey record can signal possible problems, excessive consistency across multiple response records in a dataset can also be a red flag. Identical responses to a subset of items (e.g., age, race/ethnicity, gender, sexual orientation, income, and geographic location) when more variation would be expected across response records should invite suspicion and further inspection, particularly when timestamps of survey completion are relatively contemporaneous (Simone, 2019; Storozuk et al., 2020). Researchers can sort the dataset and track such identical response patterns across records (Nash et al., 2019; Grov et al., 2019). Similarly, the open-ended questions suggested among our Layer I strategies could help to capture word-for-word repeat responses, which may be automated using bots. For open-ended questions to which one would expect to see diverse responses, nearly identical short responses with only slight variation could also invite suspicion (Simone, 2019, 2020; Storozuk et al., 2020).

When collected in a survey, email addresses warrant close scrutiny, especially if email is the mechanism via which compensation is provided. For example, repeat email addresses can signal fraudulent responding (Simone, 2019; Nash et al., 2019; Grov et al., 2019; Storozuk et al., 2020), as can domain names (the portion of the email address following the "@" symbol) from countries outside the sampling range. Email addresses with matching usernames and domain names (e.g., `username@username.com`), particularly when observed in multiple response records collected in quick succession, may also be a red flag. Another email-based indicator of possible illegitimate responding that we encountered during our pilot study was a series of response records that specified email usernames containing feminine first names inconsistent with the target population gender (i.e., men; e.g., `Jane.Doe@gmail.com`). However, given that names traditionally associated with a given gender may be used by people of other genders, we caution against using a gendered email name as the sole reason to deem a response record illegitimate.

Paradata programmed for collection at the survey development stage can be subsequently examined at the data verification stage. The aforementioned "RelevantID" featured in Qualtrics assesses respondent metadata to assess the likelihood of repeated responses (Qualtrics, Provo, UT). The technology of this feature analyzes the respondent's browser, operating system, and geographic location to generate a fraud score; however, this option does not verify the content of the responses for duplicates. As noted previously, paradata can include timestamps for when the survey was started and completed, and duration of survey completion (Grov et al., 2019; Nash et al., 2019; Simone, 2019; Storozuk et al., 2020); an unusually short duration (e.g., more than two standard deviations below the mean) might suggest a bot or inattentive human respondent (Nash et al., 2019; Storozuk et al., 2020; Schroeders et al., 2022).

[2] In our pilot project, we thought it particularly important to compensate potential respondents even if the legitimacy of associated response records was in question because our study population comprised a stigmatized, underserved community with a long history of mistreatment in scientific research. We did not want to risk cultivating or exacerbating mistrust related to research participation or unfairly denying compensation from legitimate participants. Compensation decisions such as these can be made in consultation with the IRB.

Hence, researchers can sort the dataset by duration of completion and consider excluding response records that were completed unusually fast (Grov et al., 2019; Nash et al., 2019). We recommend research team members taking the survey before Layer II to get a sense of the average response time. Data can also be sorted by IP addresses and corresponding geolocation data (e.g., latitude and longitude where the survey was completed): Duplicates of these paradata across response records may signal fraudulent responding (Nash et al., 2019; Simone, 2019; Storozuk et al., 2020). However, given that respondents using the same device or router can have identical IP addresses, we do not recommend using this strategy independent from other strategies for identifying fraudulent responders in order to avoid false positives. It is important to consider the content of the responses in conjunction with paradata red flags when making determinations regarding data exclusions.

Researchers can consult third party software that unravels true geolocations behind IP addresses and IP address aliases, such as Structon, an approach that uses Web mining coupled with inference to geolocate IP addresses with superior accuracy than existing automated approaches (Guo et al., 2009; Ruiz-Sanchez et al., 2001). Researchers can then manually exclude records that fall out of their desired geographical location ex post facto. We should note that our team does not have direct experience with Structon; instead, we used ActiveCampaign to streamline automatization of communication and stimuli exposure during our longitudinal study in which participants were emailed images weekly over eight weeks. ActiveCampaign includes a feature that documented respondents' true geolocations, which serendipitously revealed that a number of our respondents did not meet location-based eligibility criteria, and such responses were excluded from the sample in the presence of any red flags. However, we recommend asking respondents whether they are travelling or intend to travel during the project to avoid excluding respondents from a certain location who happen to travel versus individuals from outside the sampling range using IP addresses to hide their actual location for whatever reason.

Layer IV (The Core): Respondent Incentives

Incentives are payments or rewards given to research participants to motivate them to take the survey. According to the Office of Human Research Protections, incentive payments extend beyond what participants might be fairly owed for their contribution to studies (i.e., compensation) and instead motivate "speedy and complete study recruitment and retention by making research participation potentially more attractive than alternatives" (Office for Human Research Protections, 2019). The literature on how incentives impact quality of responses is mixed (Göritz, 2010; Heerwegh, 2006; Groves & Peytcheva, 2008). By introducing incentives to respondents in online survey studies, researchers might increase the likelihood of participation, maximize retention, and improve response accuracy (Göritz, 2010).

Despite these benefits, incentives might also elicit adverse outcomes, including inattentive and illegitimate responding. For example, some respondents might skip more items or develop rigid response styles to exert less mental effort and facilitate movement through the survey. It is possible that respondents receiving incentives will answer less conscientiously than groups without incentives because incentives might diminish the intrinsic motivation to perform the task (Heerwegh, 2006). Unmotivated participants who would be inclined to prematurely discontinue a surveyif no incentive were offered may instead push through haphazardly and finish the survey if an incentive is offered. Respondents seeking incentives might fill in meaningless or false data in order to quickly reach the end of a survey or submit the questionnaire multiple times (i.e., engage in illegitimate responding; Göritz, 2010). There is also risk that because of a survey incentive, respondents who do not meet the eligibility criteria may falsely answer eligibility screening items to get access to the study and gain the incentive. Another downside of incentives is that they might attract a particular type of respondent and consequently bias sample composition (e.g., socioeconomically disadvantaged people may be more responsive than advantaged people to monetary incentives; Groves & Peytcheva, 2008).

We recommend that researchers avoid immediate, automatized incentive distribution and instead have a research team member manually dispense each incentive for three reasons. First, it gives the researcher time to first verify the legitimacy of the survey record using strategies outlined in Layer III. Second, more immediate compensation (vs. delayed gratification) may reinforce fraudulent responding more strongly. Third, with automated compensation, a wave of fraudulent responses risks swiftly wiping out the study budget before the fraud is detected and the survey suspended. To circumvent these pitfalls, we suggest notifying respondents at the end of the survey that incentives for participation will not be immediately disbursed, but rather, will be sent in a given timeframe (e.g., 3-5 days). The specific timeframe of incentive disbursement should also be clearly articulated in the informed consent form both for transparency and for the possible deterrence of illegitimate responding.

Although incentivized participation may encourage fraudulent responders to take surveys (Göritz, 2010; Simone, 2020), resolving this issue is not as simple as altogether eliminating incentives from survey research. Af-

ter all, we want to appreciate and thank legitimate respondents for their time and effort. Besides being ethically questionable, eliminating incentives may significantly strain recruitment, and in some cases entirely prevent data collection in the absence of willing respondents. In addressing the core of the problem of fraudulent survey responses, researchers need to recognize the potential tradeoff between minimizing incentivization of fraudulent responses and eliminating appropriate recompense for legitimate responses. Additionally, researchers need to strive to find optimal incentive values with minimal risk of coercion, particularly when surveying socioeconomically disadvantaged communities. For example, the risk for coercion increases when providing incentives excessively higher than average hourly wages because such opportunities are difficult to refuse considering the generally lesser burden to participation in survey studies compared to the burden of many average-wage jobs. We recommend that in the formative stages of the survey study, researchers: (a) Consider the daily realities, challenges, and incentives of their population of interest; (b) Determine incentives relative to effort required of participants and burden to their everyday life; and (c) Devise appropriate incentive disbursement mechanisms for participation before disseminating surveys into communities. Consulting with community members and the IRB during the formative stages of studies might help to establish appealing, but also ethical incentive values for survey respondents.

Another potential solution to the problem of fraudulent survey responses could be non-monetary incentives that would appeal to the surveyed population, but not necessarily the general public. For example, in a different study we conducted, Black MSM were compensated with cash but also entered into a drawing for an all-inclusive weekend getaway to Fire Island, a location recognized at the time to be a top LGBTQ resort destination in New York.[3] The destination in this incentive option might not appeal to people outside of the LGBTQ community, and it would require human interactions to arrange, which would make deception more difficult. Relying less on completely anonymous monetary dispensation and more on such non-monetary rewards may help to remedy the core problem (financial incentives driving illegitimate responding), bearing in mind that such rewards come with novel considerations for protection of respondent confidentiality. We recommend consulting with the IRB regarding ethical administration of these incentives and with the surveyed community regarding cultural relevance and appeal of selected non-monetary incentives.

More broadly, consulting with community members – and, better yet, integrating them within the research team to provide guidance at all stages of the research study– can be invaluable in determining whether online surveys and planned recruitment strategies are the most appropriate method to tackle the research question or approach the population of interest. Community members can provide insight into the practicality of the method and its relevance in understanding the community's lived experience. Depending on the population of interest, online surveys could inadvertently limit access to the study by eligible respondents while simultaneously facilitating access by illegitimate respondents.

Conclusion

Online surveys have been an increasingly popular research method among scientists over the past two decades, and they have become all the more important since the COVID-19 pandemic, when in-person survey administration was especially challenging. Looking ahead, we can anticipate that the need for online surveys in research will continue to grow, in part due to practical reasons. However, the online survey method will continue to face obstacles to data integrity as the technological sophistication of both human respondents and bots continues to evolve and evade safeguards. We recommend that researchers develop a concrete protocol for handling fraudulent responses before they even occur. IRBs are a useful resource in developing such a protocol, and they will help ensure that respondents' rights are protected in addition to protecting data integrity. Box 1 at the end offers sample text that we have begun including in our IRB protocols for online survey studies initiated after the pilot project. Protecting data integrity in survey research represents a multilayered problem, which requires a multilayered system of prevention and remediation strategies analogous to the protective layers of an actual onion.

Authors' note

The study referenced as a case example was a pilot study (PI: S.K. Calabrese) funded by an award from the District of Columbia Center for AIDS Research, an NIH funded program (P30AI117970), which is supported by the following NIH Co-Funding and Participating Institutes and Centers: NIAID, NCI, NICHD, NHLBI, NIDA, NIMH, NIA, NIDDK, NIMHD, NIDCR, NINR, FIC and OAR. *Authors funded by this project: X. Modrakovic, C.A. Boone, D.A. Kalwicz, S. Rao, S.K. Calabrese*.

[3] This study involved qualitative focus groups and did not include an online survey. We refer to it here as an example of using of non-monetary incentives with select appeal, an incentivization strategy that could be applied to survey research.

References

Ahler, D., Roush, C., & Sood, G. (2021). The micro-task market for lemons: Data quality on amazon's mechanical turk. *Political Science Research and Methods*, 1–20. doi: 10.1017/psrm.2021.57.

Andrade, C. (2020). The limitations of online surveys. *Indian journal of psychological medicine*, *42*(6), 575–576. doi: 10.1177/0253717620957496.

Couper, M. P. (1998). Measuring survey quality in a casic environment. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 41–49.

Dubé, K., Perry, K. E., Mathur, K., Lo, M., Javadi, S. S., Patel, H., Concha-Garcia, S., Taylor, J., Kaytes, A., Dee, L., Campbell, D. M., Kanazawa, J., Smith, D. M., Gianella, S., Auerbach, J. D., Saberi, P., & Sauceda, J. A. (2020). Altruism: Scoping review of the literature and future directions for HIV cure-related research. *Journal of Virus Eradication*, *6*, 100008-1–8.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2005). *Behavior rating inventory of executive functions—adult version (brief-a)*. PAR.

Göritz, A. S. (2010). Using lotteries, loyalty points, and other incentives to increase respondent response and completion. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 219–233). American Psychological Association. doi: 10.1037/12076-014.

Grov, C., Westmoreland, D., Rendina, H. J., & Nash, D. (2019). Seeing is believing? Unique capabilities of internet-only studies as a tool for implementation research on HIV prevention for men who have sex with men: A review of studies and methodological considerations. *Journal of acquired immune deficiency syndromes, (1999), 82 Suppl 3(Suppl 3)*, S253–S260. doi: 10.1097/QAI.0000000000002217.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189. doi: 10.1093/poq/nfn011.

Guo, C., Liu, Y., Shen, W., Wang, H. J., Yu, Q., & Zhang, Y. (2009). Mining the web and the internet for accurate ip address geolocations. *IEEE INFOCOM, 2009 (). IEEE*, 2841–2845.

Heerwegh, D. (2006). An investigation of the effect of lotteries on web survey response rates. *Field Methods*, *18*(2), 205–220. doi: 10.1177/1525822X05285781.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P., Jewell, R., & Winter, N. (2020). The shape of and solutions to the mturk quality crisis. *Political Science Research and Methods*, *8*(4), 614–629. doi: 10.1017/psrm.2020.6.

Levi, R., Ridberg, R., Akers, M., & Seligman, H. (2022). Survey fraud and the integrity of web-based survey research. *American Journal of Health Promotion*, *36*(1), 18–20.

McCredie, M. N., & Morey, L. C. (2019). Who are the turkers? A characterization of mturk workers using the personality assessment inventory. *Assessment*, *26*(5), 759–766. doi: 10.1177/1073191118760709.

Nash, D., Stief, M., MacCrate, C., Mirzayi, C., Patel, V. V., Hoover, D., Grov, C., et al. (2019). A web-based study of HIV prevention in the era of pre-exposure prophylaxis among vulnerable HIV-negative gay and bisexual men, transmen, and transwomen who have sex with men: Protocol for an observational cohort study. *JMIR research protocols*, *8*(9), e13715–9.

Office for Human Research Protections. (2019). *Us department of health & human services* [Attachment A - Addressing Ethical Concerns Offers of Payment to Research Participants]. https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-a-september-30-2019/index.html

Qualtrics. (2018). *Qualtrics* [Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA]. Retrieved May 29, 2024, from https://www.qualtrics.com

Ruiz-Sanchez, M. A., Biersack, E. W., & Dabbous, W. (2001). Survey and taxonomy of ip address lookup algorithms. *IEEE network*, *15*(2), 8–23.

Saberi, P. (2020). Research in the time of coronavirus: Continuing ongoing studies in the midst of the COVID-19 pandemic. *AIDS and Behavior*, *24*(8), 2232–2235.

Schroeders, U., Schmidt, C., & Gnambs, T. (2022). Detecting careless responding in survey data using stochastic gradient boosting. *Educational and Psychological Measurement*, *82*(1), 29–56. doi: 10.1177/00131644211004708.

Simone, M. (2019). *How to battle the bots wrecking your online study behavioral scientist*. https://behavioralscientist.org/how-to-battle-the-bots-wrecking-your-online-study

Simone, M. (2020). *Melissa Simone – survey sleuth. Nature* [https://media.nature.com/original/magazine-assets/d41586-020-00768-0/d41586-020-00768-0.pdf].

Storozuk, A., Ashley, M., Delage, V., & Maloney, E. A. (2020). Got bots? Practical recommendations to protect online survey data from bot attacks. *The Quantitative Methods for Psychology*, *16*(5), 472–481. doi: 10.20982/tqmp.16.5.p472.

Teitcher, J. E., Bockting, W. O., Bauermeister, J. A., Hoefer, C. J., Miner, M. H., & Klitzman, R. L. (2015). Detecting, preventing, and responding to "fraudsters" in internet research: Ethics and tradeoffs. *The Journal of Law,*

Medicine & Ethics, *43*(1), 116–133. doi: 10.1111/jlme.12200.

Von Ahn, L., Blum, M., Hopper, N. J., & Langford, J. (2003). Captcha: Using hard ai problems for security. *International conference on the theory and applications of cryptographic techniques*, 294–311.

Waggoner, P. D., Kennedy, R., & Clifford, S. (2019). Detecting fraud in online surveys by tracing, scoring, and visualizing ip addresses. *Journal of Open Source Software*, *4*(37), 1285.

Weems, G. H., & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. *Measurement and Evaluation in Counseling and Development*, *34*(3), 166–176. doi: 10.1080/07481756.2002.12069033.

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, *49*(5), 737–747.

Winter, N., Burleigh, T., Kennedy, R., & Clifford, S. (2019). *A simplified protocol to screen out vps and international respondents using Qualtrics* (tech. rep.).

Wright, K. B. (2005). Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of computer-mediated communication*, *10*(3), 1034–1034.

Box 1. Sample Text to Include in IRB Protocols. Including planned methods of detecting and addressing illegitimate survey responses in IRB protocols enables researchers to respond promptly to suspected fraudulent responding when it occurs, thereby minimizing interruptions to data collection.

To prevent fraudulent responses, we have included several measures, including:

- Using the " Prevent Multiple Submissions" feature in Qualtrics
- Using a CAPTCHA code
- Using a geolocation screening feature
- Using an attention check open-response screening item

Despite the above measures, fraudulent responses are possible and will be identified by features such as: repeat IP addresses (despite "Prevent Multiple Submissions" setting in Qualtrics being active), respondents reporting their recruitment sources to be a source we did not use, numerous suspicious email addresses being reported in a pattern (e.g., email addresses all following the format `FirstnameLastnameXx@...`), and an unusual pattern of open response answers.

Per IRB guidance on managing fraudulent responding, individuals associated with fraudulent responses will not be counted towards our target enrollment. Individuals associated with response records identified as definitively fraudulent will not be paid. When fraudulent responses are identified, the survey will be closed and the survey link will be changed before the survey is reactivated.

Citation

Modrakovic, X., Boone, C. A., Kalwicz, D. A., Rao, S., Parchem, B., Wittlin, N. M., Patel, V. V., Magnus, M., Zea, M. C., Kharfen, M., Dovidio, J. F., & Calabrese, S. K. (2024). Dissecting the onion: Identifying and remediating issues surrounding data integrity in online survey research. *The Quantitative Methods for Psychology*, *20*(2), 76–87. doi: 10.20982/tqmp.20.2.p076.

Received: 13/12/2023 ~ Accepted: 28/05/2024