



# HHS Public Access

Author manuscript

*Psychol Sex.* Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

*Psychol Sex.* 2022 ; 13(4): 901–911. doi:10.1080/19419899.2021.1936617.

## **Bots and notes: Safeguarding online survey research with underrepresented and diverse populations**

**Sara Bybee,**

University of Utah, College of Nursing, 10 S 2000 E, Salt Lake City, UT, 84112

**Kristin Cloyes,**

University of Utah, College of Nursing

**Lee Ellington,**

University of Utah, College of Nursing

**Brian Baucom,**

University of Utah, Department of Psychology

**Katherine Supiano,**

University of Utah, College of Nursing

**Kathi Mooney**

University of Utah, College of Nursing

### **Abstract**

The COVID-19 pandemic has intensified the use of online recruitment and data collection for reaching historically underrepresented minorities (URMs) and other diverse groups. Preventing and detecting responses from automated accounts “bots” and those who misrepresent themselves is one challenge in utilizing online approaches. Through internet-mediated methods, interested LGBTQ+ and non-LGBTQ+ couples facing advanced cancer completed an interest form via REDCap®. Eligible participants received a direct link to electronic consent and surveys in REDCap®. Once responses to the interest form (N = 619) were received, the study PI: 1) assessed participants’ entries and non-response survey data (time of completion, rate of recruitment, etc.), 2) temporarily postponed recruitment, 3) sent eligibility questionnaires, consent documents, and validated surveys to N= 10 couples and scrutinized these data for suspicious patterns or indications of untrustworthy data, 4) responded to potential participants via email, and 5) implemented additional strategies for detecting and preventing untrustworthy survey responses. Investigators must consider multi-step eligibility screening processes to detect and prevent the collection of untrustworthy data. Investigators’ reliance on internet-mediated approaches for conducting research with diverse, hard-to-reach populations increases the importance of addressing threats to data validity. Ultimately, safeguarding internet-mediated research supports research accessibility and inclusion for URMs while also protecting participant data integrity.

---

corresponding author: Sara Bybee sara.bybee@utah.edu.

Declaration of Interest Statement

There are no known conflicts of interest associated with this publication. This is an original manuscript and it not being considered elsewhere for publication.

## Keywords

Internet-mediated research; data collection; surveys and questionnaires; minority groups; LGBTQ persons

---

## Introduction

The global pandemic has illuminated the increasing need for conducting internet-mediated research. The social distancing guidelines resulting from COVID-19 have forced many researchers to conduct their studies virtually, increasing the use of the internet for recruitment and conducting survey-based research. In mid-March, when the coronavirus pandemic forced many Americans to work from home, internet use in America increased by 25% within a few days (Rizzo & Click, 2020).

Health and social sciences researchers are also increasingly relying on internet-mediated research methods due to the ease and efficiency in recruiting large samples (Pew Research Center, 2019). Online recruitment is often used with historically difficult to reach groups such as underrepresented minorities (URMs) (Kaplan et al., 2018). URMs can include African Americans, Native Americans, Mexican Americans, mainland Puerto Ricans, and additional non-racial/ethnic factors such as socioeconomic status, sexual orientation, disability status, rural origin, first-generation college graduate, and religion (Page et al., 2013). For some URMs, such as LGBTQ+ individuals who may mistrust the medical system (Whitehead et al. 2016), participating in online research may feel more comfortable, as it may provide an environment that feels safer. For researchers conducting studies with URMs, internet-mediated research may be critical for obtaining a large enough sample size to conduct a rigorous study.

Typical internet-based outreach and recruitment strategies include emailing lists of eligible participants, website posting, and using social media platforms like Facebook to deliver recruitment information to targeted groups, or Twitter and Instagram for broader outreach. The use of internet-mediated research via surveys that are directly accessible to participants is increasingly commonplace as these methods allow researchers to accrue more diverse and representative samples (McInroy, 2016; McInroy & Beer, 2021; Watson et al., 2018), often outside of local, geographic boundaries (Watson et al., 2018). Internet-mediated research offers convenience for participants (McInroy, 2016), which can reduce barriers to participation and lower implementation costs for researchers (Watson et al., 2018).

Despite the many advantages, conducting internet-mediated research may increase individuals' potential to misunderstand inclusion criteria or purposefully misrepresent their sociodemographic or medical information (Grey et al., 2015). Prominent national participant registries used for health and social science research operate on the "honor system" by allowing individuals to sign up by providing an email address and name, then self-report their demographic, medical, and health behavior information. With study methods such as online surveys directly accessible via public links, participants can remain anonymous and may never communicate with researchers (Dewitt et al., 2018; Teitcher et al., 2015). If researchers do not implement strict screening procedures or restrictive software features for

online surveys, there is a risk of multiple submissions and potentially untrustworthy data. Some researchers have found that multiple submissions are associated with the promise of incentives for participation and recommend limiting such incentives (Quach et al., 2013).

To date, the majority of peer-reviewed articles regarding preventing and detecting untrustworthy survey responses are conducted with younger individuals, including many studies conducted in the area of sexual health (Teitcher et al., 2015). This may be due to the idea that: sexual health is a sensitive topic, groups recruited for studies related to sexuality may be historically more difficult to reach, and using internet-mediated methods increases a sense of privacy and autonomy for participants, thereby facilitating participation. Despite published information regarding threats to data integrity in studies of sexual health (Dewitt et al., 2018; Teitcher et al., 2015), there is scant literature regarding detecting and managing untrustworthy data in internet-mediated dyadic research with adults. Dyadic research may encounter different patterns of responses (since both partners in a couple may need to provide their name and contact information) that can be examined to determine if they represent unique and eligible participants. There is also limited information regarding how to address software robots (commonly referred to as bots) in studies with these populations. Bots are computer-generated algorithms designed to mimic human-like behavior (Ferrara et al., 2016), and they may present an especially pernicious threat to data validity.

For scientists researching URM populations, challenges in recruiting a large enough sample size combined with untrustworthy data can contribute to underrepresentation in research and continued health disparities. For LGBTQ+ populations, it is critical to gather sexual orientation and gender identity (SOGI) data in order to accurately assess health disparities and plan effective interventions for reducing them (Centers for Disease Control and Prevention [CDC], 2020). If data integrity is in question, research questions about URMs such as LGBTQ+ populations will remain unanswered, and the extent of health disparities may remain unknown and unaddressed. Therefore, it is critical to address data integrity concerns in internet-mediated studies of URMs to work towards health equity for these populations.

The purpose of this manuscript is to: 1) illuminate potential concerns regarding data integrity when conducting internet-mediated research, 2) demonstrate how researchers can begin to determine if data is trustworthy, 3) describe strategies for ensuring data integrity through the study design process, and 4) discuss specific considerations when conducting internet-mediated research with LGBTQ+ dyads. This paper uses data collected from an initial interest form ( $N = 619$ ), and surveys completed online ( $N = 20$ ) with LGBTQ+ and non-LGBTQ+ adult couples facing cancer.

## Methods

The data reported in this paper stems from a larger ongoing research project which examines posttraumatic growth (1)—the positive psychological change that may occur through the struggle with a highly stressful or traumatic event (Tedeschi & Calhoun, 1995)—among couples facing cancer. We included LGBTQ+ and non-LGBTQ+ couples (spouse, partner, or significant other) to examine any potential differences in their PTG

experiences and identify any health disparities faced by LGBTQ+ couples. The larger study uses a cross-sectional, convergent mixed-methods design which collects quantitative data through Research Electronic Data Capture (REDCap®) surveys and qualitative data through semi-structured interviews with couples (Harris et al., 2009).

The research team utilized paper and electronic study recruitment flyers developed to assist in recruiting LGBTQ+ and non-LGBTQ+ couples. Flyers contained a brief description of the study, the PI's contact information, the eligibility criteria, and described the \$20 electronic gift card compensation for study participation. The eligibility criteria included: 1) age 40 years or older, 2) have a diagnosis of advanced cancer (stages III or IV, metastatic, unresectable, or have stopped curative treatment), and 3) have a significant other who also agrees to participate.

The research team performed study recruitment by 1) Posting announcements to four special interest group online forums through the Gerontological Society of America, 2) Posting announcements to four cancer-support Facebook groups, 3) Posting public announcements on Facebook, Twitter, and Instagram, and 4) Asking members of the research project's advisory board to share study information with their networks. The board was comprised of community members who had experience with cancer and/or who self-identify as part of the LGBTQ+ community, experts in psycho-oncology, and experts in LGBTQ+ aging. The University of Utah Institutional Review Board approved all procedures.

Individuals interested in participating in the study completed a brief interest form through REDCap®, which asked for their name and email and their partner's name and email. The PI informed interested participants that each partner listed on the interest form would receive a link to the eligibility questionnaire within 24-48 hours. The public link to the interest form was included in all recruitment efforts, whether via email or social media.

Because the survey responses and response patterns alerted the team to possible problems with legitimacy of the data, a stepwise process was developed to establish data trustworthiness. Once a large number of responses to the interest form were received within a short period of time, the study PI: 1) assessed participants' entries on the interest forms and examined other factors from the interest forms (e.g. time of completion, rate of recruitment, etc.), 2) temporarily postponed recruitment, 3) sent eligibility questionnaires, consent documents, and validated surveys to N= 10 couples (the first five who submitted interest forms and five other couples in which both partners report the same last name) and scrutinized these data for suspicious patterns or indications of untrustworthy data (e.g. large sections of unanswered questions), 4) responded to potential participants via email, and 5) implemented additional strategies for detecting and preventing untrustworthy survey responses. These actions were implemented based on peer-reviewed literature regarding strategies for improving the trustworthiness of data collected via the internet (Ballard et al., 2019; Godinho et al., 2020; Teitcher et al., 2015).

## Results

### Data collected

There were  $n = 370$  responses to the interest form within 24 hours of recruitment, representing  $n = 370$  dyads or  $n = 740$  individuals. On day two, there were an additional 164 responses (total  $N = 534$ ). On day three, there were 66 responses (total  $N = 600$ ), on day four, there were four responses (total  $N = 604$ ), and on day five, there were 15 responses (total  $N = 619$ ). After day five, the research team temporarily postponed recruitment due to the high response rate, high number of blank responses, and limited number of couples reporting the same last name. These unusual patterns elicited concerns regarding trustworthiness of the data, including potential participant deception and the use of bots.

The high response rate was incongruent with previous literature documenting the challenges in recruiting cancer patients and caregiver dyads and LGBTQ+ individuals (Heckel et al., 2018; Whitehead et al., 2016). Also, for a dyadic study on coupled significant others, of the  $N = 619$  total responses, only  $n = 7$  (0.16%) had the same last name (and three of these submissions were recorded within two to four minutes of one another).

Many email addresses did not match the participant's name but used a different first and last name in their email address. For example, a participant might have entered their name as John Doe but listed their email as JaneSmith@gmail.com. There were two duplicate responses and numerous batches of interest forms completed in rapid succession of one another. For example, on the first day of recruitment, there were six interest forms submitted simultaneously at 8:21 PM, three forms submitted at 8:22 PM, two forms submitted at 8:23 PM, and nine forms submitted at 8:24 PM. The PI also identified 113 blank responses completed in rapid succession, confirming attempts at multiple entries (Quach et al., 2013). Of particular importance for dyadic research, none of the  $N = 619$  couples reported the same email address. The Pew Research Center (2014) states that 27% of internet users in a marriage or committed relationship have a shared email account with their partner and older adults, and those in relationships for more than ten years are more likely to share an email account.

During this postponed recruitment and assessment of the data's trustworthiness, the PI received six emails—each from the individual identified as the cancer patient within the couple. The PI received these six emails in the exact order in which interest forms were submitted and with the same subject line. For example, the patient listed in the first submitted interest form emailed the PI asking about study participation. Then, the patient from the second submitted interest form emailed the PI, followed by the patient from the third completed interest form, and so on. Were these unique eligible participants, there should be more natural variation in the timing and order in which participants contacted the PI.

To further examine the data's trustworthiness, after receiving the 619 interest forms, the PI continued the enrollment process with couples from the first five interest forms (See Figure 1 for a visual of the steps taken by the research team). As an additional method of testing

couples' veracity, the PI also completed enrollment with five of the couples whose partners reported the same last name on their interest forms. (The research team was suspicious of the trustworthiness of the couples reporting the same last names since all forms were received at approximately the same time.) Through REDCap®, the PI sent a link to the eligibility questionnaire to both partners in the ten couples. All 20 individuals appeared to be eligible based on their responses to the eligibility questionnaire and thus were sent a link to the consent document and surveys in REDCap®. All 20 individuals completed the consent document and all the validated surveys.

The following data reported represent an analysis of the  $n = 20$  individuals (10 couples) who submitted interest forms, eligibility questionnaires, consent documents, and study surveys. All of the first five couples who submitted interest forms left the same section of one validated survey blank, signaling the potential that responses may be untrustworthy (Quach et al., 2013). Furthermore, on a different validated survey, two couples (who reported the same last name as their partner) gave identical responses. The remaining three couples who reported the same last name gave nearly identical answers to one another.

The time in which it took participants to complete surveys was also an indication of untrustworthy data (Teitcher et al., 2015). Many surveys were completed in one minute when the estimated time of completion was 15 minutes. Prior research suggests that compared with individuals who made unique survey submissions, those who made multiple submissions spent less time on each survey (Quach et al., 2013). The timing of survey completion within couples was also a concern for untrustworthy data. For example, many couples had one partner complete all the surveys at 2 pm, and then the next partner would begin their surveys at 2:01 pm. While some couples may share a device and may be near each other while completing the surveys, the frequency with which this occurred suggests that these surveys may be completed by one person posing as two different people. Therefore, these data added to the research team's suspicion of untrustworthy data.

### Examining Data Trustworthiness

The PI wrote emails to all  $n = 20$  participants with the following message: "Good afternoon, I apologize for the delay in responding to your interest in the study. There seems to be an issue with your participant IDs. If you are still interested in participating, could you please call me during business hours so we can sort it out? It should only take a minute. You can reach me at [telephone number]. I apologize for the inconvenience" (Teitcher et al., 2015). None of the 20 participants who received this request called the PI. Based on the  $n = 10$  couples' lack of response, the remaining  $n = 609$  couples who submitted interest forms were emailed the same message and asked to call the PI. Again, none did, and therefore these responses were considered invalid, and participants were not enrolled in the study (Ballard et al., 2019). While some emails from the PI may have not been received by the participants due to email or spam filters, the fact that not a single of the 1,238 partners responded to the PI's email to their self-identified preferred email address was alarming. All data from the remaining couples ( $n = 609$ ) and the  $n = 10$  couples who completed the surveys (total  $N = 619$ ) were considered untrustworthy and not included in the larger study.

Before resuming recruitment, the PI took additional measures to ensure future participants' credibility (See Table 1 for a complete list of recommendations). First, in the public link to the interest form, the PI enabled the Completely Automated Public Turing Test to tell Computers and Humans Apart (CAPTCHA) feature, which some experts say may reduce the possibility of fraudulent participation (Watson et al., 2018; Teitcher et al., 2015). This was executed in REDCap® by navigating to the survey distribution tab and under the public survey link, clicking the box to enable CAPTCHA. The public link to the interest form was excluded in all future recruitment efforts, requiring interested participants to directly contact the PI (Pozzar et al., 2020).

Additional fields were added to the eligibility questionnaire to prevent further untrustworthy responses. The PI added the following statement to the beginning of the eligibility questionnaire: "Please note that participants who are suspected of fraudulent behavior or falsifying information will NOT receive compensation. Multiple entries are not allowed" (Teitcher et al., 2015). While the eligibility questionnaire initially asked interested participants if they were 40 years of age or older, the revised questionnaire asked participants to enter their age in a text box and subsequently provide their date of birth to verify eligibility based on age (Konstan et al., 2005). The PI added a question regarding how the individual heard about the study, providing two incorrect answer choices to screen out untrustworthy responses (Chandler et al., 2020).

The addition of the CAPTCHA occurred on November 9<sup>th</sup>, 2020, and all other changes were completed by November 12<sup>th</sup> after receiving REDCap® approval. With these additional measures in place, the responses to the interest form substantially declined. There was one response per day to the interest form in the following six days. Within 30 days of these changes, an additional 11 responses were received and appeared to be credible based on initial interest form data. Twelve couples were randomly invited to take the new eligibility questionnaire. Of these 12 couples, 11 (91.7%) reported hearing about the study via Facebook, and one couple reported hearing about the study via a Craigslist ad (which was not utilized in this study).

## Discussion

This paper provides additional information regarding potentially untrustworthy responses in internet-mediated research and offers amelioration strategies. The implications of these findings extend to scientists conducting online surveys with diverse populations and to studies utilizing dyadic recruitment. The current study did not initially incorporate strategies for preventing untrustworthy responses into the study design and therefore was limited in its ability to follow many preventative strategies. However, by implementing revisions to the surveys and changes to the recruitment methods, the study team significantly reduced the number of responses to the interest form received per day, thereby increasing our trustworthiness of future data collected. We highly recommend that researchers consider the possibility of untrustworthy responses during the design phase of their studies and implement as many strategies as possible to reduce the likelihood of collecting data from bots or individuals misrepresenting themselves.



There is some discussion in the literature on provision of compensation and the amount of compensation that should be offered to participants (Teitcher et al., 2015; Quach et al., 2013). However, especially among diverse populations, there is increasing consensus that participants should be paid for their time and for providing their personal expertise (Beckdor et al., 2020). With the economic downturn resulting from COVID-19 and the social distancing guidelines emphasizing working remotely, participating in online research for compensation may have increased deceptive research behavior and the use of bots.

One important consideration for data integrity is determining if any specific recruitment forums were more likely to result in untrustworthy data. Complete metadata for all 619 interested couples is not available. However, once we added a question regarding where participants heard about the study, the majority of participants stated that they heard about the study through Facebook. The high response rate from Facebook may indicate a greater vulnerability to bots and individuals who misrepresent themselves, however many researchers find Facebook to be highly effective in recruiting participants (Christensen et al., 2017; Watson et al., 2018). In the future, researchers should carefully consider the benefits and drawbacks of using Facebook in recruitment efforts.

To combat the threat of bots, the computing community created advanced methods of detection that utilize “highly predictive features that capture a variety of suspicious behaviors and separate social bots from humans” (Ferrara et al., 2016, p. 101). Other techniques to identify data quality issues include data analytics, visualization, and large databases to aggregate and normalize data from multiple sources (Knepper et al., 2016). For a more detailed discussion of detecting social bots, readers are encouraged to consult Latah (2020). Some companies also offer verified online participants who have been pre-screened (Prolific, 2020). Despite advances in detection and pre-screened participants, many of these strategies remain unavailable to those without the funding to purchase them. For unfunded studies or studies with constrained budgets (e.g., student projects, pilot studies), the inability to pay for complex techniques to detect bots and individuals misrepresenting their eligibility may further add to the health disparities that URM face.

### Limitations

This paper examines data obtained during a psychosocial study recruiting LGBTQ+ and non-LGBTQ+ couples facing advanced cancer. The findings from this study may therefore not apply to other populations. Given the use of REDCap® for survey distribution, the researchers were unable to collect IP addresses as a form of verification, limiting the ability to identify multiple responses using this method. While the study recruitment method and eligibility screening process were substantially revised in response to concerns regarding untrustworthy data, other study design features (such as the use of IP addresses and data analytics) could not be altered. It is highly encouraged that researchers consider the potential threat of bots and individuals misrepresenting themselves during study design to limit the potential of untrustworthy data.



## Conclusion

As the internet becomes increasingly critical in conducting research with URMs such as LGBTQ+ populations (and during the COVID-19 pandemic), it is even more essential that participant data is trustworthy. Researchers must use prevention and detection strategies for deterring dishonest responses when conducting online surveys. While some may employ experts and advanced algorithms, financial constraints impede their use for many (especially student researchers and early-career investigators). The strategies outlined in this manuscript offer preventive low-cost ways to prevent and detect untrustworthy data collection in online surveys. The authors hope that this manuscript serves to introduce the possibility of untrustworthy data when conducting internet-mediated research and to prepare social scientists for dealing with these threats to their data integrity.

Social scientists should document any untrustworthy data in studies conducted with LGBTQ+ populations to determine if these populations are targeted purposefully, and if so, for what purposes. Just as the computing community continues to develop new algorithms to detect bots, the scientific community needs to report concerns regarding untrustworthy data and develop amelioration strategies to address these threats' constant evolution.

## Acknowledgments

The authors would like to acknowledge the National Center for Advancing Translational Sciences and the National Institute of Nursing Research of the National Institutes of Health for supporting this research.

This work was supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002538; and by the National Institute of Nursing Research of the National Institutes of Health under Award Number T32NR013456. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Author bios

Sara Bybee, LCSW is a Ph.D. Candidate at the University of Utah College of Nursing. She earned a Master's degree in Social Work from Portland State University in 2012 and Bachelor's degrees in Community Health and Spanish from Tufts University in 2008. She has been employed as an oncology social worker at Huntsman Cancer Institute for the past five years. During her doctoral studies, Sara earned Gerontology Interdisciplinary Program and Nursing Education Graduate Certificates. She conducts research with interdisciplinary teams studying family and cancer caregiving, hospice visits, hospice team attitudes towards LGBTQ families, and dyadic interactions among cancer patients and their partner caregivers. She has presented research at: The Society of Behavioral Medicine, Social Work in Hospice and Palliative Care Network, the Gerontological Society of America, and the Western Institute of Nursing conferences. Sara is a Jonas Nurse Scholar alumni and is currently a predoctoral fellow in Interdisciplinary Training in Cancer, Caregiving, and End-of-Life Care through the National Institutes of Health. Sara serves as a co-convenor of the Rainbow Research special interest group of the Gerontological Society of America, and as Utah's Advocacy Leader with the Graduate Nursing Student Academy.

Kristin Cloyes is an associate professor at the University of Utah College of Nursing. She investigates family and informal caregiving in communities with flexible and adaptive

family and kinship structures that have been underserved or marginalized by health care and research. She is especially interested in how networks of family, friends, and close others provide social support and caregiving for members with chronic illness and at end of life. Since 2014, her program of research has focused on LGBTQ+ individuals with chronic and life-limiting illness and their caregivers. She is leading an NIH Administrative Supplement on Dr. Ellington's R01 (NR016249) investigating support and communication needs for LGBTQ hospice family, and is the PI on a multi-site project utilizing mEMA with LGBTQ and AYA advanced cancer patients and their primary caregivers to assess patterns of support interactions within their social networks. Recently, she led a mixed-methods study to investigate how distal sources of minority stress—including interactions with health care providers—affect chronic illness, pain, and mental health outcomes for older adults in these groups.

Lee Ellington is a Professor at the University of Utah College of Nursing, a licensed clinical psychologist, and a Huntsman Cancer Institute Investigator. She holds the Robert S. and Beth M. Carter Endowed Chair. Her interdisciplinary program of research focuses on the impact of interpersonal health communication on adjustment, decision-making, and health. Dr. Ellington has had 20 years of continuous funding from NIH and the American Cancer Society, and for the last decade her team has focused their research on hospice care of families facing life-limiting cancer. She was a Project Leader on a Program Project Grant (5P01CA138317-06; PI Mooney) on hospice family caregivers of cancer patients and nurse communication. She is currently the PI of a multi-site R01 in partnership with the Palliative Care Research Cooperative (NR016249) assessing the responsiveness of hospice care team members to the daily needs of cancer family caregivers. She is also the PI of an Administrative Supplement, led by colleague Dr. Kristin Cloyes, to examine the experiences of sexual and gender minority hospice family caregivers. Additionally, Dr. Ellington co-directs a T32 in Interdisciplinary Training Program in Cancer, Caregiving, and End-of-Life Care (T32 NR013456; mPIs Ellington & Mooney). Dr. Ellington was recently awarded a 5-year K07 Academic Leadership Award—The Family Caregiving Research Collaborative (K07 AG068185).

Brian Baucom is an Associate Professor of Psychology at the University of Utah. His research integrates trait questionnaires, daily diary, psychophysiological, observational, and sensor-based measurement of behavior and emotion using data acquired in the laboratory as well as outside of the laboratory during daily life. He has extensive expertise in collecting, processing, and dyadic observational and self-report data. He is currently the m-PI of a \$2,355,470 project funded by the Department of Defense that develops signal processing technologies and machine learning algorithms for risk for suicide based on observationally coded dyadic behavior and individual performance on cognitive performance tasks. He is also currently a co-I on eight different federally funded and foundation grants five of which involve the study of dyadic processes linked to mental and physical health outcomes including four different grants that examine close relationships. His role in each of these projects includes both substantive and statistical/methodological roles that draw on his substantial quantitative expertise in complex dyadic analysis. In addition to these eight current grants on which he is a co-I, he serves as a consultant on three more federally or internationally funded grants that examine dyadic processes and has also been a co-I

or consultant on eight completed federally or internationally funded grants, six of which focused on dyadic research.

Katherine Supiano is an Associate Professor and the Director of Caring Connections: A Hope and Comfort in Grief Program, the bereavement care program of the University of Utah serving the intermountain west. She earned her PhD from the University of Utah College of Social Work in 2012. She has been a practicing psychotherapist in geriatrics, palliative care and grief therapy for over 35 years. She is a Fellow in the Gerontological Society of America, a Fellow of Thanatology-Association of Death Education and Counseling, and a member of the Social Work Hospice and Palliative Care Network Board of Directors. Her research focuses on developing, implementing and evaluating training of clinicians to address the Grief of Overdose Death and the Grief of Suicide Death (Utah Division of Substance Abuse and Mental Health). She is also working towards evaluating a prevention model for dementia caregivers at risk for complicated grief, funded by the Alzheimer's Association. She has received an Innovations in Care grant from the Rita and Alex Hillman Foundation to conduct a feasibility study of Complicated Grief Group Therapy and has received 1U4U Innovation grant from the University of Utah.

Kathi Mooney is a Distinguished Professor at the University of Utah College of Nursing and holds the Louis S. Peery and Janet B. Peery Presidential Endowed Chair in Nursing. She is an investigator and co-leader of the Cancer Control and Population Sciences Program at the Huntsman Cancer Institute. Her research focuses on cancer patient symptom remote monitoring and management, technology-aided interventions, cancer family caregivers, improving supportive care outcomes for cancer patients in rural/frontier communities and models of cancer care delivery. Dr. Mooney has been continuously funded through National Cancer Institute grants for the past 20 years. She has published numerous book chapters and journal articles and is a frequent speaker on topics related to cancer symptom management, quality cancer care, new delivery models, technology-aided interventions/telehealth, and creativity and innovation in science.

## References

- Ballard AM, Cardwell T, & Young AM (2019). Fraud detection protocol for web-based research among men who have sex with men: Development and descriptive evaluation. *JMIR Public Health Surveillance*, 5(1), e12344–e12344. 10.2196/12344 [PubMed: 30714944]
- Beckford Jarrett S, Flanagan R, Coriolan A, Harris O, Campbell A, & Skyers N (2020). Barriers and facilitators to participation of men who have sex with men and transgender women in HIV research in Jamaica. *Culture, Health & Sexuality*, 22(8), 887–903. 10.1080/13691058.2019.163422
- Centers for Disease Control and Prevention [CDC]. (2020, Feb 13). Collecting sexual orientation and gender identity information: Importance of the collection and use of these data. <https://www.cdc.gov/hiv/clinicians/transforming-health/health-care-providers/collecting-sexual-orientation.html>
- Chandler J, Sisso I, & Shapiro D (2020). Participant carelessness and fraud: Consequences for clinical research and potential solutions. *Journal of Abnormal Psychology*, 129(1), 49–55. 10.1037/abn0000479 [PubMed: 31868387]
- Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, Toft Sørensen H, & Mikkelsen EM (2017). Costs and efficiency of online and offline recruitment methods: A web-based cohort study. *Journal of Medical Internet Research*, 19(3), e58–e58. 10.2196/jmir.6716 [PubMed: 28249833]

- Dewitt J, Capistrant B, Kohli N, Rosser BRS, Mitteldorf D, Merengwa E, & West W (2018). Addressing participant validity in a small internet health survey (The Restore Study): Protocol and recommendations for survey response validation. *JMIR Research Protocols*, 7(4), e96–e96. 10.2196/resprot.7655 [PubMed: 29691203]
- Ferrara E, Varol O, Davis C, Menczer F, & Flammini A (2016). The rise of social bots. *Communications of the Association for Computing Machinery*, 59(7), 96–104. doi: 0.1145/2818717
- Godinho A, Schell C, & Cunningham JA (2020). Out damn bot, out: Recruiting real people into substance use studies on the internet. *Substance Abuse*, 41(1, 3-5). 10.1080/08897077.2019.1691131
- Grey JA, Konstan J, Iantaffi A, Wilkerson JM, Galos D, & Rosser BR (2015). An updated protocol to detect invalid entries in an online survey of men who have sex with men (MSM): How do valid and invalid submissions compare? *AIDS Behavior*, 19(10), 1928–1937. 10.1007/s10461-015-1033-y [PubMed: 25805443]
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, & Conde JG (2009). Research electronic data capture (REDCap®)-A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377–381. 10.1016/j.jbi.2008.08.010 [PubMed: 18929686]
- Heckel L, Gunn KM, & Livingston PM (2018). The challenges of recruiting cancer patient/caregiver dyads: Informing randomized controlled trials. *BMC Medical Research Methodology*, 18(1), 146. 10.1186/s12874-018-0614-7 [PubMed: 30463520]
- Kaplan CP, Siegel A, Leykin Y, Palmer NR, Borno H, Bielenberg J, Livaudais-Toman J, Ryan C, & Small EJ (2018). A bilingual, Internet-based, targeted advertising campaign for prostate cancer clinical trials: Assessing the feasibility, acceptability, and efficacy of a novel recruitment strategy. *Contemporary Clinical Trials Communications*, 12, 60–67. 10.1016/j.conctc.2018.08.005 [PubMed: 30272035]
- Knepper D, Fenske C, Nadolny P, Bedding A, Gribkova E, Polzer J, Neumann J, Wilson B, Benedict J, & Lawton A (2016). Detecting data quality issues in clinical trial: Current practices and recommendations. *Therapeutic Innovation & Regulatory Science*, 50(1), 15–21. doi: 10.1177/2168479015620248 [PubMed: 30236017]
- Konstan JA, Simon Rosser BR, Ross MW, Stanton J, & Edwards WM (2006). The Story of subject naught: A Cautionary but optimistic tale of internet survey research. *Journal of Computer-Mediated Communication*, 10(2), 00–00. 10.1111/j.1083-6101.2005.tb00248.x
- Latah M (2020). Detection of malicious social bots: A survey and a refined taxonomy. *Expert Systems with Applications*, 151. 10.1016/j.eswa.2020.113383d
- McInroy LB (2016). Pitfalls, potentials, and ethics of online survey research: LGBTQ and other marginalized and hard-to-access youths. *Social Work Research*, 40(2), 83–94. 10.1093/swr/svw005 [PubMed: 27257362]
- McInroy LB & Beer OW (2021): Adapting vignettes for internet-based research: Eliciting realistic responses to the digital milieu. *International Journal of Social Research Methodology*. DOI: 10.1080/13645579.2021.1901440
- Page KR, Castillo-Page L, Poll-Hunter N, Garrison G, & Wright SM (2013). Assessing the evolving definition of underrepresented minority and its application in academic medicine. *Academic Medicine*, 88(1). 10.1097/ACM.0b013e318276466c
- Pew Research Center. (2014, February 11). Couples, the internet, and social media. <https://www.pewresearch.org/internet/2014/02/11/couples-the-internet-and-social-media/>
- Pew Research Center. (2019, June 12). Internet/Broadband fact sheet. <https://www.pewresearch.org/internet/fact-sheet/internet-broadband/>
- Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulskey JA, Hong F, Gundersen DA, & Berry DL (2020). Threats of bots and other bad actors to data quality following research participant recruitment through social media: Cross-sectional questionnaire. *Journal of Medical Internet Research*, 22(10), e23021–e23021. 10.2196/23021 [PubMed: 33026360]
- Prolific. (2020, February 13). How it works. <https://www.prolific.co/#how-it-works>
- Quach S, Pereira JA, Russell ML, Wormsbecker AE, Ramsay H, Crowe L, Quan SD, & Kwong J (2013). The good, bad, and ugly of online recruitment of parents for health-related focus groups:

Lessons learned. *Journal of Medical Internet Research*, 15(11), e250. 10.2196/jmir.2829 [PubMed: 24231040]

Rizzo L & Click S (2020, August 15). How Covid-19 changed American's internet habits.

The Wall Street Journal. <https://www.wsj.com/articles/coronavirus-lockdown-tested-internets-backbone-11597503600>

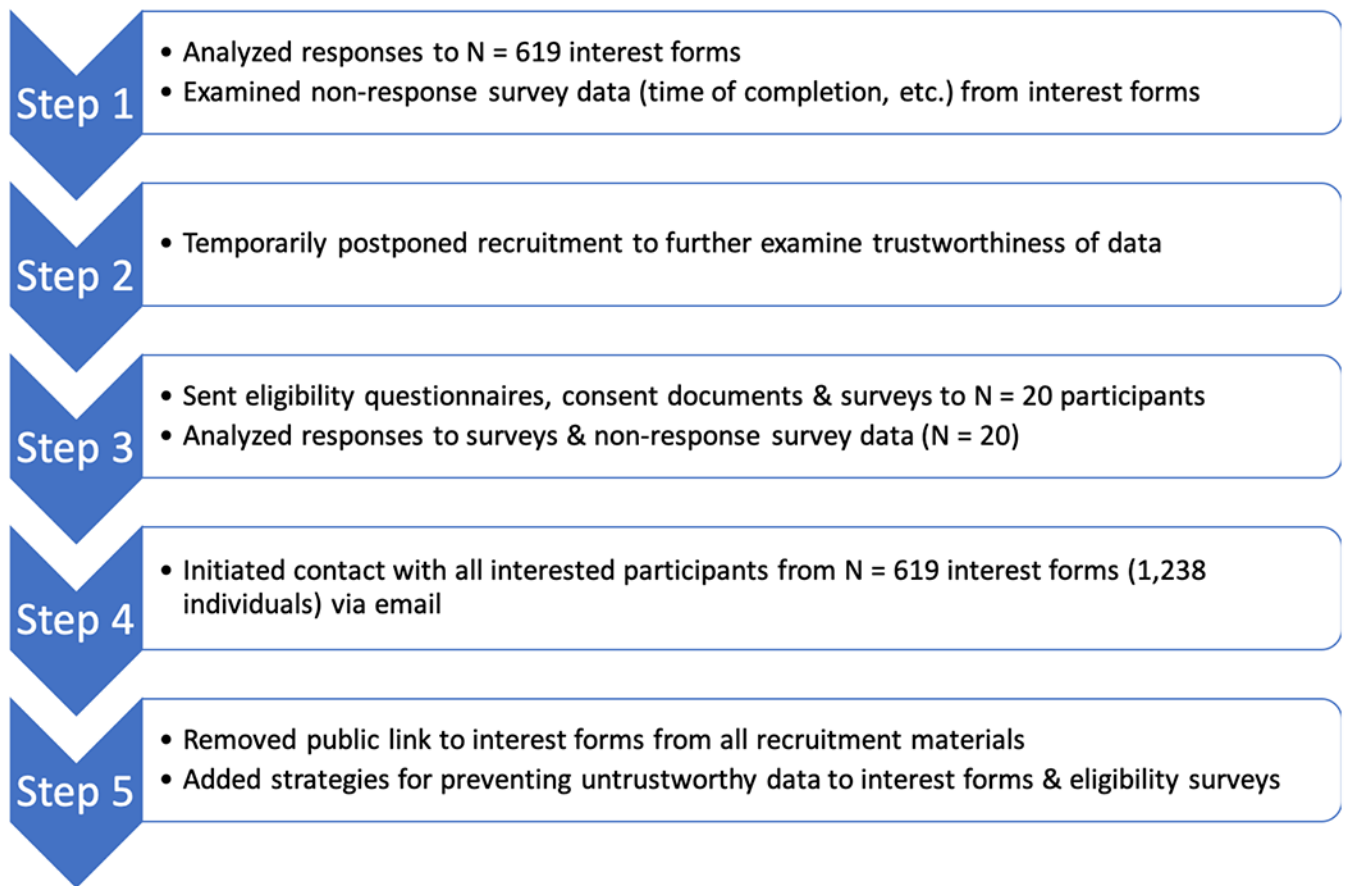
Tedeschi RG, & Calhoun LG (1995). *Trauma and transformation: Growth in the aftermath of suffering*. Sage Publications, Inc.

Teitcher JE, Bockting WO, Bauermeister JA, Hoefler CJ, Miner MH, & Klitzman RL (2015).

Detecting, preventing, and responding to "fraudsters" in internet research: Ethics and tradeoffs. *Journal of Law, Medicine, and Ethics*, 43(1), 116–133. 10.1111/jlme.12200

Watson NL, Mull KE, Heffner JL, McClure JB, & Bricker JB (2018). Participant recruitment and retention in remote eHealth intervention trials: Methods and lessons learned from a large randomized controlled trial of two web-based smoking interventions. *Journal of Medical Internet Research*, 20(8), e10351–e10351. 10.2196/10351 [PubMed: 30143479]

Whitehead J, Shaver J, & Stephenson R (2016). Outness, stigma, and primary health care utilization among rural LGBT populations. *PLoS One*, 11(1), e0146139. 10.1371/journal.pone.0146139 [PubMed: 26731405]



**Figure 1.**  
Steps taken by the research team

**Table 1.**

## Recommendations for improving the trustworthiness of data

<b>Area of Application</b>	<b>Recommendation</b>
<i>Study design phase</i>	
Social media recruitment	Strongly weigh the pros and cons of using social media platforms for recruitment, as these forums may be more susceptible to the collection of untrustworthy data
Link to public survey	Enable a CAPTCHA feature on the public link and/or consider removing a public link on all recruitment materials
Eligibility questionnaire	Consider adding a question that contains untrue options such as asking where the participant heard of the study.
Eligibility questionnaire/ Demographic questionnaire	Include a blank text box for participants to list their age AND ask a blank text box for them to write their date of birth to compare their responses.
Participant contact	Prior to enrolling eligible participants, set up a brief video chat/phone call (with IRB approval) to improve the trustworthiness of their data
Survey introduction	Consider adding a statement at the beginning of the survey stating that multiple entries are not allowed and will not be compensated
Collect IP addresses	Use online survey methods that allow you to collect IP addresses so you can screen for multiple submissions
Collect verifiable information	Collect phone numbers or physical addresses that can be easily verified
Include time stamps	Including a time stamp at the beginning and end of each instrument may help determine if responses are trustworthy based on estimated completion time
<i>Data collection and analysis phase</i>	
Accrual rates	Monitor recruitment/accrual rates of participants to compare rates/sample sizes to sample sizes typical for your specific research domain/population
Compensation	If compensation is provided, consider excluding this information from initial recruitment efforts, or use a lottery system where not every participant receives this incentive
Timing of survey completion	Frequently examine data to determine the average time for survey completion and consider eliminating data 2 standard deviations above or below this threshold
Blank sections of survey	Frequently examine the data for any patterns such as large blocks of blank questions, or multiple surveys with the same exact responses one after another
Responsiveness of participants	Consider contacting participants (with IRB approval) if there are concerns about the trustworthiness of data; ask participants to call the study number during business hours in order to receive compensation; require a home address for delivery of compensation rather than an email address so you can easily compare addresses of participants