Evaluation and Program Planning 80 (2020) 101794

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Matching anonymous participants in longitudinal research on sensitive topics: Challenges and recommendations

Jane E. Palmer[a,*], Samantha C. Winter[b], Sarah McMahon[c]

[a] American University, School of Public Affairs, Department of Justice, Law & Criminology, 4400 Massachusetts Ave NW, Kerwin Hall, Washington, DC, 20016, United States

[b] Columbia University, School of Social Work, 1255 Amsterdam Ave, Rm 801, New York, NY, 10027, United States

[c] Rutgers University, School of Social Work, Center on Violence Against Women & Children, New Brunswick, NJ, 08901, United States

ARTICLE INFO

Keywords:
Self-generated identification codes
Campus sexual violence
Longitudinal research
Anonymity
Sample matching

ABSTRACT

The purpose of this study was to examine the final analytic sample of a longitudinal randomized control trial (RCT) evaluation of a sexual violence prevention program at a university after facing challenges with the implementation of a self-generated identification code. The matched and unmatched samples (e.g., all unique surveys across all time periods) included 10,135 surveys. Eighty-eight percent of these surveys were matched into the final longitudinal dataset. Findings suggest that students with certain characteristics were more likely to be matched over time (i.e., students who participated in student government, Latino/a students, and Asian students). In addition, students who did not comply with RCT protocol were less likely to be matched. Student history of victimization or perpetration of sexual violence was not associated with being matched over time. This study provides recommendations for preventing matching problems in longitudinal studies, a process for rectifying matching issues and a critique of studies that do not address issues of matching-related sample bias in their final analytic sample.

1. Introduction

Randomized controlled trials (RCTs) and other types of longitudinal designs strengthen the validity of evaluation research. However, these designs can present a number of challenges for researchers (Grossman & Mackenzie, 2005), especially when the focal issue is a sensitive topic such as sexual violence (e.g., Ullman, 2011). Major challenges can include participant adherence to protocol, missing data (including attrition) and ensuring that participation remains anonymous or, at the very least, confidential (Ahren & Le Broque, 2005). In particular, there is a dearth of research on longitudinal studies' matching success rates when using anonymous strategies for identification (DiIorio, Soet, Mater, Woodring, & Dudley, 2000). Therefore, the focus of the current study is to describe best practices for matching techniques in longitudinal studies, describe the process used to rectify matching issues in a longitudinal RCT university-based sexual violence prevention program evaluation, determine the extent and nature of sample bias due to matching challenges, and to describe procedures for examining sample bias in longitudinal studies.

2. Literature review

There are few longitudinal studies examining campus sexual violence prevention or incidence (Banyard, 2014; McMahon, Wood, Cusano, & Macri, 2019; Voth Schrag, 2017). A recent systematic review of campus sexual assault studies from 2000 to 2015 found that only 32 % of studies were longitudinal, ranging from 8 weeks to 4 years (Fedina, Holmes, & Backes, 2018). Of these longitudinal studies, most used confidential, not anonymous, methods of tracking participants over time (e.g., randomly assigning students a code that was associated with a key). Only one used an anonymous self-generated identification code (Messman-Moore, Coates, Gaffey, & Johnson, 2008) and the research team faced minimal challenges with matching over time (Messman-Moore, 2019).

It is clear from experimental research that anonymity, relative to confidentiality, increases the validity of self-report in survey research on sensitive topics (Beatty, Chase, & Ondersma, 2014; Durant, Carey, & Schroder, 2002; Ong & Weiss, 2000; van de Looij-Jansen, Goldschmeding, & de Wilde, 2006). In order to increase the number of studies using anonymity instead of confidentiality, to reduce bias, and increase validity, it is important for researchers to address practical

* Corresponding author.
E-mail addresses: jpalmer@american.edu (J.E. Palmer), scw2154@columbia.edu (S.C. Winter), smcmahon@ssw.rutgers.edu (S. McMahon).

https://doi.org/10.1016/j.evalprogplan.2020.101794
Received 22 October 2019; Received in revised form 5 February 2020; Accepted 11 February 2020
Available online 12 February 2020

considerations and challenges posed by using anonymity in longitudinal research (Bouchard, 2016). One of the most promising and common methods used to keep participants anonymous involves using self-generated identification codes.

2.1. Self-generated identification codes

Carifio and Biron (1978) first introduced the idea of self-generated identification codes (SGICs) as a method to allow for anonymity in a longitudinal study. Since then, studies have established strong test-retest reliability of these methods with sensitive survey topics (such as substance abuse or sexual behavior) and diverse populations of middle, high school and postsecondary students (see Schnell, Bachteler, & Reiher, 2010 for a review). When using SGICs, the code should be easy to reproduce with accuracy by longitudinal study participants and the determinants of the code should be static during the study period (Damrosch, 1986; DiIorio et al., 2000; McGloin, Holcomb, & Main, 1996). The elements of SGICs are typically derived from survey questions that ask respondents to indicate items like their mother's name, home street name, birth month, the number of older siblings, or similar questions (Ripper, Ciaravino, Jones, Jaime, & Miller, 2017; Schnell et al., 2010).

Although SGICs help to reduce social desirability bias when the topic of a study is sensitive, SGICs are not a perfect solution (Schnell et al., 2010). Inconsistent answers can lead to an inability to match a participant longitudinally. In the current study, we outline our matching strategies, matching process, and matched analytic sample to help other researchers consider how to maximize matching among their anonymous, longitudinal participants. We include several correlates to examine if there are certain types of students who are more likely to be matched, so that interventions could be targeted toward the populations most "at-risk" of being unmatched.

3. Current study

Several studies have found that there are demographic correlates (such as race, gender, or class) and activities related to delinquent activity that are significantly associated with anonymous participants being matched (or unmatched) across waves of longitudinal data. For example, unmatched participants tend to engage in more delinquent behaviors, be male and have a lower socioeconomic status (DiIorio et al., 2000; Grube, Morgan, & Kearney, 1989; Kristjansson, Sigfusdottir, Sigfusson, & Allegrante, 2013; Pérez, Ariza, Sánchez-Martínez, & Nebot, 2010). DiIorio et al. (2000) found that matched participants were likely to be white, which supported McAlister and Gordon's (1986) finding that unmatched participants tended to not be white. Given that the research is limited in this area, in this exploratory study, we examine whether demographic correlates (ethnicity and gender), engagement in campus life (involvement in Greek life, athletics or student government), delinquent behavior (perpetration behavior) and salience of the topic (victimization experiences) were associated with being matched longitudinally. These correlates are associated with the main dependent variables in the larger study (citation removed for anonymous peer review), so it was essential for the research team to evaluate matching success on each of these variables. The findings of this study could be helpful to researchers who are trying to plan for a strong longitudinal design or who are trying to understand who they lost due to problems with matching.

4. Method

4.1. Data

The data for this study are from an evaluation of a university-based sexual violence prevention program. This study was a longitudinal RCT designed to examine the impact of a sexual violence prevention program on undergraduate students' attitudes and behaviors related to sexual violence and bystander intervention. There were six data collection points over an 18-month period. The first survey (pretest) was administered in person to incoming first year students at the new student orientation. The remaining five surveys were conducted online using a web-based survey program.

The population for this study included all incoming first year students who attended orientation at a large Mid-Atlantic university in summer, 2010. At pretest, participants were asked to create a self-generated identification code (SGIC) that included their birth month, day, and the first three letters of their mother's first name (e.g., 0923ANN). This allowed for anonymous participation in the study while also providing the mechanism to track and match individuals' surveys over time.

A total 4390 incoming first year students completed the pretest and, subsequently attended the prevention program presentation at new student orientation. Of these students, 2,002 completed the first posttest at the start of the fall semester. Next, these participants were randomly assigned to either the experimental group receiving two additional doses of the intervention or the control group receiving no additional doses. The two additional doses of the intervention were administered to the experimental group before the second posttest, which was in December 2010.

4.1.1. Matching strategies

Although SGIC methods from previous studies were followed (e.g., McGloin et al., 1996), like several past studies (Schnell et al., 2010), there were inconsistencies and duplications among the SGICs generated by the study participants. Additionally, similar to Garvey Wilson et al. (2010), there were individuals who failed to read instructions and entered their code to be the same as the example code. Therefore, after the first posttest, the research team received Institutional Review Board approval to ask participants to voluntarily provide their email addresses and high school attended to assist with matching participants across time.

Upon conclusion of the study, a "two-phase matching process" was followed to link participants' SGICs longitudinally (Carifio & Biron, 1978; Damrosch, 1986; Grube et al., 1989). In Phase I, we matched the SGICs that were exact matches. For Phase II, we used the "off-one procedure." This involves retaining cases that can be matched based on all but one element of the code. Decisions about which elements can be altered have varied in previous studies. For example, McGloin et al. (1996) allowed age to vary but not birth month. In the study by Garvey Wilson et al. (2010) "gender and race had to be an exact match, while age, education level and marital status were allowed one move since theoretically these demographics could have slight changes in the few months between surveys" (p.933). After the SGIC matching process, it is recommended that tests of sample bias are conducted (e.g., Grube et al., 1989; Kristjansson et al., 2013; Yurek, Vasey, & Sullivan Havens, 2008) and that information about unmatched sample is included when reporting results (see Kristjansson et al., 2013).

4.1.2. Matching process

The complete matching process for the current study is summarized in Fig. 1. In total 10,998 surveys were collected over the length of the study Step 1. Matching based on SGIC Phase I involved several intermediate steps. First, prior to matching on SGIC, 40 exact duplicate surveys were permanently removed from the study sample because they did not represent unique cases e.g., participants clicked "submit" more than once when sending their survey data to the server Step 2. Next, a total of 836 surveys across all time periods were temporarily removed because they had duplicate SGICs Step 3. SPSS software was then used to match surveys with non-duplicate SGICs from all five posttests to surveys with unique SGICs from the pretest 3,966 cases. In total 8,840 80 % of the total surveys collected could be matched based only on the seven-digit SGIC Step 4. Thus, at the end of Phase I, a total of 2,118

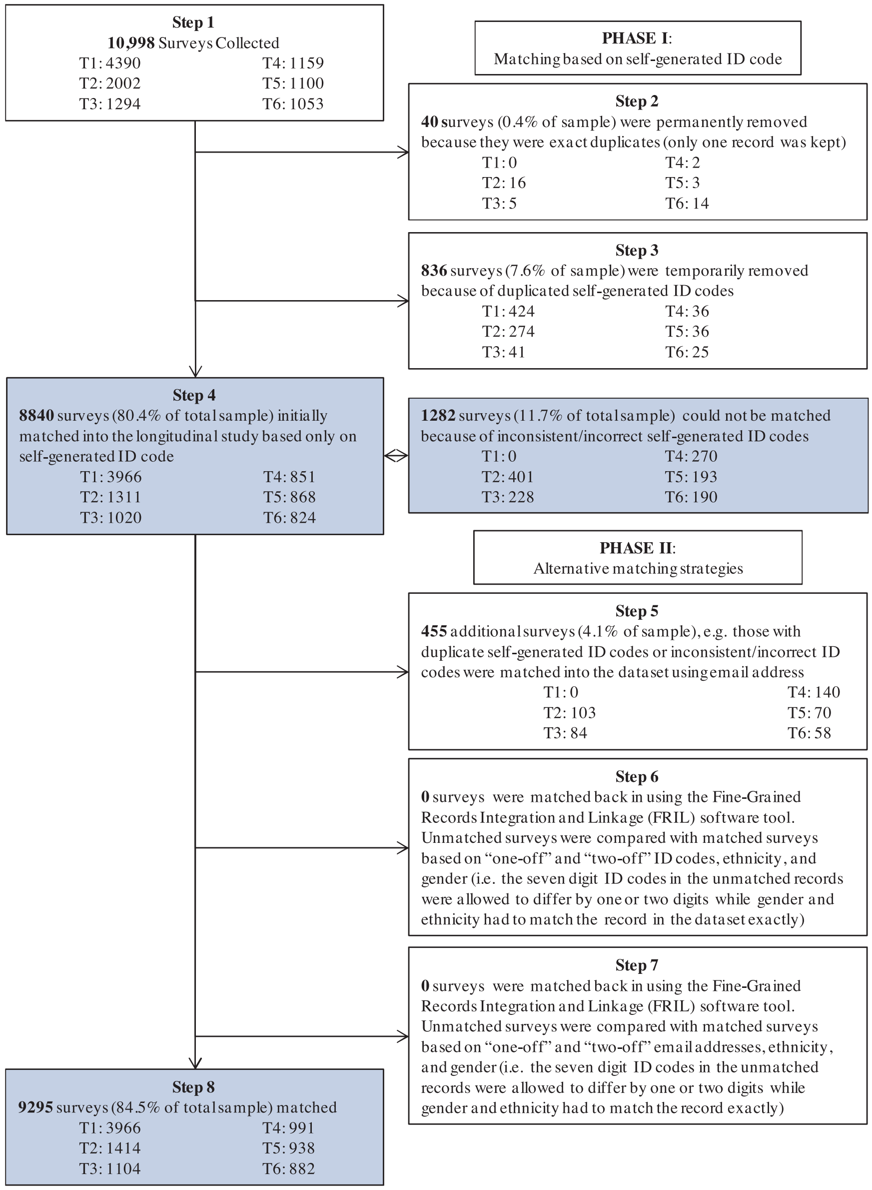

Fig. 1. Survey Matching Process.

surveys 836 surveys with duplicate SGICs from Step 3 and the 1,282 surveys that did not match a pretest SGIC remained unmatched from the longitudinal sample.

In Phase II, the research team used three strategies to match remaining surveys into the longitudinal sample. First, of the remaining 2,118 unmatched surveys, 455 surveys were matched into the longitudinal sample using email addresses collected from the students during the five posttests T2-T5 Step 5. Second, Fine-Grained Records Integration and Linkage FRIL software was used to try to match additional surveys with "one-off" SGICs Jurczyk, Lu, Xiong, Cragan, & Correa, 2008). Gender and ethnicity were assumed to remain constant over the study period; therefore, unmatched surveys with SGICs that differed from matched surveys by one or two digits, but had identical gender and ethnicity values, were identified using FRIL. All "one-off" records were compared with the longitudinal, matched records (Step 6). Surveys with discrepancies in email address or other demographics were not added back into to the dataset despite "one-off" match results. Lastly, records with "one-off" email addresses, but identical gender and ethnicity, were identified using FRIL (Step 7). However, the "one-off" SGIC and email address comparisons in Steps 6 and 7 did not yield any additional surveys that were not already matched back into the dataset using exact-match email address. Therefore, after the matching process was completed, the final longitudinal dataset included 9,295 matched surveys—representing 84.5 percent of the total surveys collected (Step 8). All email addresses were removed from the dataset once the matching by email address process was complete.

4.1.3. Matching analysis sample

Fig. 2 displays the process of obtaining the final analytic sample for the matching analysis for the current study. The matched analysis compares surveys that can clearly be linked to an individual participant with those that cannot, therefore the unit of analysis is completed surveys rather than individual participants. The sample for the matched/unmatched analysis included 10,998 completed surveys (pretest n = 4,390, posttest 1 n = 2,002, posttest 2 n = 1,294, posttest 3 n = 1,159, posttest 4 n = 1,100, posttest 5 n = 1,053 over all six data collection periods. As above, a total of 40 surveys were removed because they were exact duplicates Step 2 in Phase I. Thus, the analytic sample represents the total number of unique surveys N = 10,958 collected in the study see Fig. 2 for number of surveys by time period).

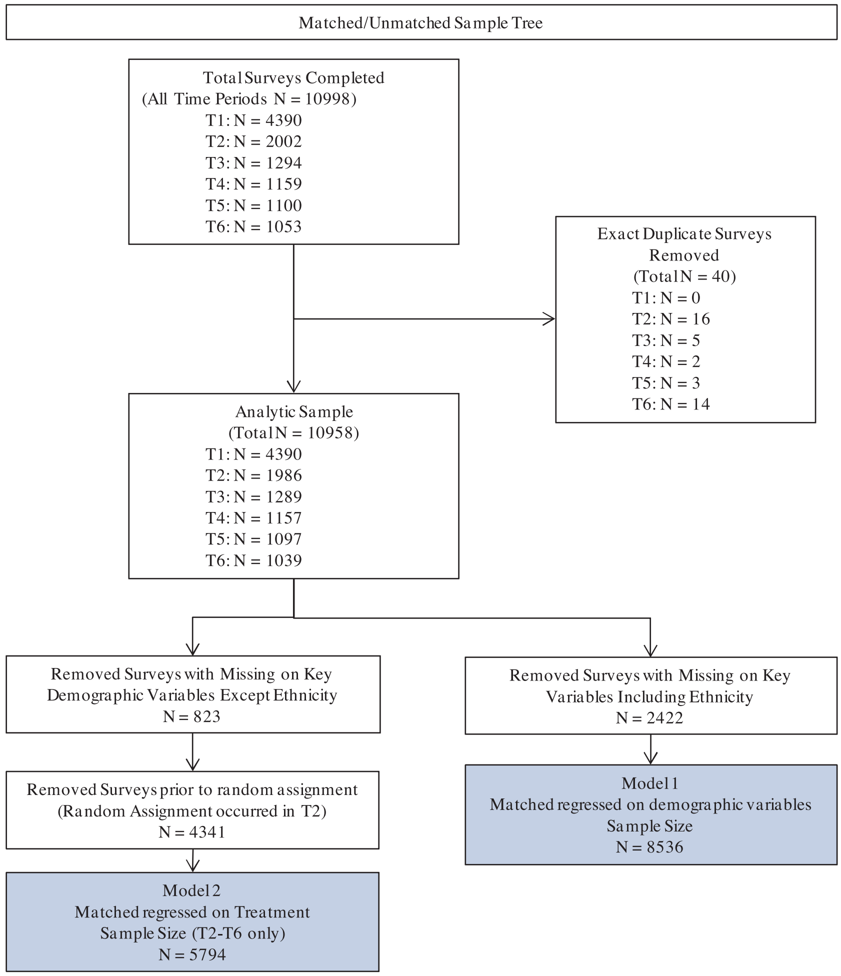

Fig. 2. Matched Analysis Sample Tree.

5. Measures

5.1. Dependent variable

5.1.1. Matched

All non-duplicate, pretest and posttest surveys that were matched into the longitudinal dataset (e.g., matched a pretest identification code) were coded as matched (1). Surveys that remained unmatched (e.g., had an incorrect, inconsistent, or duplicate SGIC or email address) were coded as unmatched (0).

5.2. Independent variables

5.2.1. Survey collection period

Student surveys were collected at six different time periods over the course of the eighteen-month study period. Surveys took place in summer 2010 (pretest), September 2010 (posttest 1), December 2010 (posttest 2), February (posttest 3), September 2011 (posttest 4), and February 2012 (posttest 5). The survey collection period variable was coded as follows: Pretest (T1) = 1, Posttest 1 (T2) = 2, Posttest 2 (T3) = 3, Posttest 3 (T4) = 4, Posttest 4 (T5) = 5, and Posttest 5 (T6) = 6.

5.3. Random assignment and adherence to treatment protocol

5.3.1. Random assignment

Stratified random assignment procedures (based on gender and ethnicity) were used to randomly assign students to the treatment and control groups directly following completion of the first posttest (T2). Students who were assigned to the treatment group were invited to attend two additional sessions (or doses) of the intervention. The control group did not receive any additional doses. Therefore, for the treatment variable, individuals who were in the control group were assigned a value of 0. In order to test for whether receiving additional doses of the treatment affected matching, participants who received at least one additional dose of the intervention were assigned a value of 1. Some students were assigned to receive additional doses of the program, but did not attend the doses; yet, these individuals completed posttest surveys beyond the second posttest. In order to account for whether *assignment* to the treatment group, alone, affected individuals' matching in the study, participants who were assigned to the treatment group, but did not attend any additional doses of the intervention after random assignment were given a value of 2.

5.4. Demographic variables

5.4.1. Gender

In the pretest, survey participants were asked to specify their gender (0 = female, 1 = male, 2 = other). Gender demographics were recoded to be 1 = female, 0 = not female because only a small number (n = 3) of participants identified as "other."

5.4.2. Ethnicity

Participants were provided with a wide range of race/ethnicity identification possibilities including White/European, Asian, Latina/Latino, South Asian, Black/African American, Caribbean, Mid-Eastern, Pacific Islander, African, American Indian/Alaskan, Multi-Ethnic and Other. For the purposes of the current study and due to the nature of the sample, these options were collapsed into five categories (1 = White/European, 2 = Black/African/African American, 3 = Latino/a, 4 = Asian/South Asian/Pacific Islander, and 5 = Other, which included participants that identified as Caribbean, Mid-Eastern, American Indian/Alaska Native, Multi-Ethnic or other).

5.5. Engagement in campus life

5.5.1. Greek life

In T2 - T6 participants were asked: "*Since coming to college, have you pledged a fraternity or sorority?*" where responses were coded as follows: Yes = 1 and No = 0. For longitudinal analyses a cumulative Greek life variable was created, i.e., if a participant responded "yes" to the question in any of the time periods between T2-T6, he or she received a Greek Life value of 1. The Greek life variable was coded as 0 for those participants who answered "no" in all posttests (T2-T6). Since the pretest took place prior to the start of the students' first semester in college, this variable was coded as 0 for all participants in T1.

5.5.2. Student government

In T2 – T6 participants were asked: "*Since coming to college, have you participated in student government?*" where responses were coded as follows: Yes = 1 and No = 0. For longitudinal analyses a cumulative student government variable was created. Participants who responded "yes" to the student government question in at least one posttest survey (T2-T6), received a student government value of 1. The student government variable was coded as 0 for those participants who answered "no" in all posttests (T2-T6). Again, since the pretest took place prior to the start of the students' first semester in college, the student government variable was coded as 0 for all participants in T1.

5.5.3. College sport

In T2 – T6 participants were asked: "*Since coming to college, have you joined an athletic team?*" where responses were coded as follows: Yes = 1 and No = 0. For longitudinal analyses a cumulative college sport variable was created. If a participant responded "yes" to the college sport question in any of the time periods between T2-T6, this variable was coded as 1. The college sport variable was coded as 0 for those participants who answered "no" in all posttests (T2-T6). Again, since the pretest took place prior to the start of the students' first semester in college, this variable was coded as 0 for all participants in T1.

5.6. Salience of research topic

5.6.1. Victimization history

In the pretest, participants were asked: "Have you ever been forced or coerced by another person to do something sexually that you did not want to do?" where Yes = 1 and No = 0. Participants in T2 – T6 were asked: "Since the last survey period, were you forced or coerced by another person to do something sexually that you did not want to do?" A cumulative, lifetime sexual victimization history variable was created by assigning a value of 1 for anyone who answered "yes" to the victimization question in at least one of the time periods (T1-T6). If participants answered "no" to the victimization question in all time periods, the variable was coded to be 0.

5.7. Delinquent behavior

5.7.1. Perpetration history

In the pretest, participants were asked the following two questions from Lisak's Perpetrator History (PH) Scale (Lisak et al., 2000):

1) "*Have you ever had sexual intercourse with someone, even though they did not want to, because they were too intoxicated (on alcohol or other drugs) to resist your sexual advances (e.g., removing their clothes)?*" (Completed Perpetration) and
2) "*Have you ever been in a situation where you tried, but for various reasons did not succeed, in having sexual intercourse with someone by using or threatening to use physical force (twisting their arm, holding them down, etc.) if they did not cooperate?*" (Attempted Perpetration)

The response options for each question were coded as 0 = No and 1 = Yes. In later time periods the question was revised so that "*have you ever*" was changed to "*since the last survey period.*"

Based on participant responses to the completed perpetration question (#1 above), a cumulative *completed perpetration* variable was created for each participant by assigning a value of 1 if the respondent answered "yes" to the question in the current time period or any preceding time period. If participants answered "no" to the first perpetration question in all time periods leading up to and including the current time period, the variable was coded to be 0.

A second *attempted perpetration* variable was created in the same way. That is, a value of 1 was assigned if the respondent answered "yes" to the *attempted perpetration* question (#2 above) in the current time period or any preceding time period.

Finally, an *overall perpetration history* variable was created by taking the mean of the *completed perpetration* and *attempted perpetration* variables in each time period. Hence, at each time period (T1-T6) *perpetration history* is a cumulative, lifetime variable with three potential values 0 = *prior to the current time period, respondent never attempted to or completed perpetration of a sexual assault*, 0.5 = *prior to the current time period, respondent attempted or completed perpetration of a sexual assault*, and 1 = *prior to the current time period, respondent attempted and completed a sexual assault*. For example, if a participant indicated they attempted sexual violence in the past in time 2 and did not attempt or perpetrate sexual violence in between time 2 and time 3, they would receive a 0.5. If a participant indicated they attempted sexual violence in the past in the time 2 survey, and in time 2 they indicated they completed sexual violence since the time 2 survey, they would receive a 1.

5.8. Adherence to survey protocol

5.8.1. Passed "still reading" checks

In the surveys from all collection periods (T1-T6), there were two questions embedded in the middle of other scale items that read "*circle the number* '2' *for this question to show that you are still reading this survey*." If students circled "2" for both questions in any given survey, they were assigned a value of 1 = Passed "still reading" checks. If students did not circle 2 on one or both of the questions, they received a value of 0 = Failed at least one "still reading" check.

5.9. Research question

Do individuals who were matched into the longitudinal sample differ significantly from those who were unable to be matched?

5.10. Data analysis strategy

Binary logistic regressions were used to estimate the probability of being matched into the longitudinal dataset. Assuming that individual participants in a longitudinal study complete surveys in more than one time period, aggregate analyses of the completed surveys from all data collection periods would violate assumptions of independence. Therefore, robust standard errors were clustered by time period (Petersen, 2009). Three separate regression models were conducted.

Model 1 tested if demographics or other characteristics of participants had a significant effect on whether individuals were matched into the longitudinal dataset. Thus, in Model 1, the *matched* variable was regressed on the following variables: *survey collection period, ethnicity,* *gender, participation in college sport, participation in student government, participation in Greek life, perpetration history, victimization history,* and *passed "still reading" checks*. Surveys were excluded from the analysis if they were missing responses to any of the demographic questions. Because the ethnicity question was excluded in the first posttest (T2), for example, the Model 1 sample does not include surveys from T2 (refer to Fig. 2 for detailed sample information).

In Model 2, the *matched* variable was regressed on the *treatment* variable to test if 1) assignment to a treatment group had a significant effect on whether individuals were matched into the longitudinal dataset and 2) receipt of treatment has a significant effect on whether individuals were matched into the longitudinal dataset. Since random assignment occurred after T2, the Model 2 sample includes only posttest surveys (T2-T5). In addition, individual surveys were removed if they did not have an experimental assignment (see Fig. 2 for detailed sample information). Surveys completed by individuals who were assigned to the treatment group, but did not attend, were compared to the control group to assess whether assignment had an effect on whether surveys could be matched longitudinally.

Additionally, surveys completed by individuals who received the treatment were compared to those in the control group to assess whether participation in the additional doses of the prevention program affected whether surveys could be matched into the study. Model 2 also included *survey collection period, gender, participation in college sport, participation in student government, participation in Greek life, perpetration history, victimization history,* and *passed "still reading" checks* as control variables. For all models, individual surveys were removed from the analysis if they had missing on any of the demographic or control variables (see Fig. 2 for detailed sample information).

6. Results

Descriptive statistics for all variables included in the matching analyses are presented in Table 1. The matched/unmatched sample (e.g., all unique and reliable surveys across all time periods) included 10,135 surveys (See Fig. 2). Eighty-eight percent of these surveys were matched into the longitudinal dataset (see Table 1, column 2).

Table 2 displays the results from Model 1 (the effect of demographics (i.e., ethnicity, gender, participation in college sport, participation in student government, participation in Greek life, perpetration history, and victimization history), controlling for other factors, on being matched) and Model 2 (the effect of treatment assignment or receipt, controlling for other factors, on being matched).

6.1. Effect of demographic characteristics

Results from Model 1 suggest that surveys completed by individuals identifying as Asian had over double the odds of being matched longitudinally and surveys completed by individuals identifying as Latino/a had about one-quarter greater odds of being matched longitudinally compared to white students.[1] Results from Model 1 also indicated that surveys completed by individuals who participated in student government had about double the odds of being matched even when controlling for a variety of other variables. This latter finding remained consistent across Models 1 and 2.

6.2. Effect of treatment assignment and receipt

Results for Model 2 - the regression of *matched* on *treatment* - suggested that surveys completed by students who were assigned to the treatment group, but did not attend, i.e., students who did not comply with treatment protocol, had lower odds of being matched into the dataset than surveys completed by individuals in the control group when controlling for a variety of demographic variables. Alternatively, surveys completed by individuals who were assigned *and* participated in at least one additional dose of the intervention did not have significantly different odds of being matched into the longitudinal dataset than those completed by participants in the control group.

In Model 1, several variables including gender, fraternity participation, athletics participation, victimization history, perpetration history, passed "still reading" questions, and being in one of two racial/ethnic categories (Black and Other) were not significantly associated with being matched. In Model 2, however, within the subsample of surveys completed by individuals who were randomized to one of the experimental groups, surveys completed by female students and surveys completed by students who passed the "still reading" questions (adherence to survey protocol) had greater odds of being matched into the longitudinal dataset than surveys completed by participants who did not identify as female or who failed at least one of the embedded "still reading" checks. The remaining variables were not significant in Model 2.

7. Discussion

The findings from this study have important implications for longitudinal studies on sensitive topics that attempt to keep participation anonymous. First, the results of this study support Schnell et al.'s (2010) assertion that the use of self-generated identification codes is not perfect. As DiIorio et al. (2000) point out, it is not realistic to anticipate a 100 % matching rate. In the current study, 88 % of surveys were matched, which is higher than other studies with ranges of 45.8%–78.1% matching rate (DiIorio et al., 2000; Galanti et al., 2007; McGloin et al., 1996). Second, the results from the analysis regressing surveys matched into the longitudinal dataset on several demographic, engagement in campus life, salience of research topic, delinquent behavior, and adherence to survey protocol variables do not provide a lot of evidence of differences between characteristics of students whose surveys matched into the longitudinal dataset compared to those whose surveys were unable to be matched.

Results from the regressions suggest that students involved in student government, Asian students and Latino students, and students who adhered to the survey protocol had higher odds of being matched into the longitudinal study sample. On the other hand, surveys submitted by students who did not comply with instructions to attend additional interventions, had lower odds of being matched. While the results of the logistic regression suggest some student characteristics may be significantly associated with surveys being matched into a longitudinal sample, the small effect sizes and low pseudo r-squared values associated with the models suggest these student characteristics may not be the primary factors explaining variability in matched versus unmatched surveys.

Furthermore, surveys from students with victimization experience or perpetration history were not more or less likely to be matched. Other longitudinal campus victimization studies have found no significant differences in participation rates, self-selection, adherence, and/or attrition for participants with a history of victimization or perpetration (see, for example, Loh, Gidycz, Lobo, & Luthra, 2005; Rosenthal & Freyd, 2018). This suggests that these factors may not be critical in influencing whether or not surveys can be matched into longitudinal datasets. Further work in this area is needed to better understand the relationship, or lack thereof, between victimization and perpetration with participation in research on sexual violence, especially given concerns with revictimizing survivors (Edwards, Kearns, Calhoun, & Gidycz, 2009) as well as the challenges with measuring perpetration (Anderson, Silver, Ciampaglia, Vitale, & Delahanty, 2019). Ultimately, our results suggest that efforts to improve matching rates, therefore, may need to focus more on improving the accuracy of SGICs and optimizing post-hoc matching processes than interventions that

[1] Since ethnicity was not asked in T2, surveys from this data collection period were excluded from this Model.

Table 1
Matching Analysis Descriptive Statistics.

	Overall Analytic Sample	Collection Period Analysis	Ethnicity & Demographics Analysis	Treatment Analysis
N	10,958	10,135	8,536	5,794
Matched	84.8	87.7	88.0	85.7
Treatment				
Control	21.8	–	–	39.6
At Least 1 Additional Dose	23.9	–	–	43.6
Assigned Treatment/Did Not Attend	9.3	–	–	16.8
Time Period				
Pretest (T1)	40.1	41.0	48.5	NA
Posttest 1 (T2)	18.1	15.5	NA	23.9
Posttest 2 (T3)	11.8	12.4	14.7	21.7
Posttest 3 (T4)	10.6	11.0	13.1	19.3
Posttest 4 (T5)	10.0	10.2	12.1	17.9
Posttest 5 (T6)	9.5	9.9	11.7	17.3
Ethnicity				
White	38.4	–	47.1	–
Black	7.6	–	9.0	–
Latino/a	6.2	–	7.6	–
Asian	23.1	–	28.7	–
Other	6.2	–	7.7	–
Female	56.5	57.6	57.2	61.5
Participation in a fraternity or sorority	4.4	4.4	4.6	7.5
Participation in student government	4.6	4.6	4.7	7.9
Participation in a college sport	9.6	9.1	7.7	15.2
Sexual victimization history	6.7	7.0	7.8	4.5
Perpetration History	2.1	2.0	2.1	2.5
Passed “still reading” Checks	93.7	96.3	96.3	95.8

Note. Numbers are percentages.

Table 2
Binary Logistic Regression Models for Matched Analyses.

	Model 1		Model 2	
	Odds Ratios	Conf. Int. [95 %]	Odds Ratios	Conf. Int. [95 %]
N	8,536		5,794	
Time Period				
Posttest 1 (T2) as compared with baseline	–	–	–	–
Posttest 2 (T3) as compared with baseline	0.583***	0.567-0.599	0.974*	0.949-0.999
Posttest 3 (T4) as compared with baseline	0.588***	0.569-0.606	0.978	0.942-1.016
Posttest 4 (T5) as compared with baseline	0.545***	0.523-0.569	0.916***	0.885-0.947
Posttest 5 (T6) as compared with baseline	0.529***	0.503-0.557	0.892***	0.847-0.94
Female	1.140	0.852-1.525	1.247***	1.115-1.394
Participation in Fraternity	0.995	0.799-1.240	1.007	0.808-1.254
Participation in Student Government	2.008***	1.878-2.147	1.983***	1.793-2.193
Participation in College Sport	1.223	0.929-1.611	1.126	0.888-1.427
Victimization History	0.955	0.764-1.192	1.076	0.69-1.677
Perpetration History	0.915	0.623-1.343	0.972	0.686-1.378
Passed “still reading” Checks	1.162	0.873-1.548	1.326*	1.026-1.714
Ethnicity				
Black as compared to white	1.097	0.754-1.596		
Latino/a as compared to white	1.221***	1.132-1.317		
Asian as compared to white	2.157***	1.477-3.149		
Other as compared to white	1.182	0.837-1.667		
Experimental Group				
Treatment as compared to control			0.913	0.731-1.141
Assigned Treatment/Didn't Attend compared to control			0.750***	0.688-0.818
Pseudo R squared	0.024		0.008	

Note. Numbers are presented as Odds Ratios, except where otherwise noted. Robust Standard Errors were used. Significance based on Two-tailed *T*-test: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

target specific groups of students.

In this study 12 % of cases were unmatched and therefore missing from our final analytic sample. This unmatched rate, along with our attrition rate, fell below the predicted missingness and initial power analyses conducted for this study. Yet, there is no indication that the missingness due to matching was related to salient factors like gender or history of victimization or perpetration. Instead, failing to adhere to the research protocol seemed to be the most significant factor. Missingness due to unmatched surveys could have the potential to affect power and effect sizes in longitudinal studies, but there is little transparency about survey matching challenges in the current literature and/or how these challenges might affect power or effect sizes. In order to address this issue, we suggest researchers take several steps (below) to help improve matching in longitudinal studies (with SGIC accuracy, survey matching techniques, and demographic questions in surveys), researcher transparency of the extent of unmatched individuals, and how this type of missingness does or does not affect study results.

8. Lessons learned

8.1. SGIC accuracy

There are several steps researchers can take to reduce the amount of inconsistent or inaccurate identification codes. First, survey participants should be asked to provide responses to several (5–7) questions that would make up the code (Schnell et al., 2010). The best identification codes are stable, variable and proximate (Yurek et al., 2008). That is, the questions should result in the same response each time (e.g., month born), the question responses should vary among respondents (e.g., we did not ask for birth year in our study since all participants were incoming undergraduate students) and the questions should be about the respondent or someone close to the respondent (i.e., the closer proximity, the more likely they will know or remember the answer) (Yurek et al., 2008).

Our SGIC met most of these guidelines—in that we asked for month born, date born, and first three letters of the name of their mother—but our SGIC lacked enough variability to prevent duplicates (i.e., there were only 3 unique components instead of the recommended 5–7 components). Therefore, some participants had actual duplicate codes (e.g., twins or coincidental shared birthdays and first three letters of their mother's name). Another source of duplicates occurred because participants misinterpreted or did not read the instructions and provided the example identification code as their own. Or, it may be that some participants were concerned that their identity would not remain anonymous and may have intentionally provided inaccurate or inconsistent identification codes.

In this study, once the matching problems due to an insufficient SGIC were identified, the study team faced an ethical issue of asking the students to provide additional identifying information to aid in matching. Per our IRB protocol, we did not associate the identifying information with their survey responses, however some students may not have trusted that we would keep the identifying information separate from their answers. From these experiences, we have several suggestions to improve SGIC accuracy including pre-testing SGICs to determine the rate at which students can reliably recreate their code (see Ripper et al., 2017 for an example) or somehow gathering qualitative information about whether errors were intentional or unintentional.

8.2. Survey matching techniques

In addition to incorporating strategies to improve the accuracy of SGICs, findings from this study also suggest there is a need for more research focused on optimizing survey matching techniques, particularly for studies relying on SGICs. In this study we utilized a multi-stage approach to matching surveys that yielded higher matching rates than previous studies; however, there is a need for more research focused on multistage matching processes and pre-study strategies to increase the probability of surveys being matched. In this study, for example, we were able to utilize email addresses as a secondary mechanism to help match additional surveys into the longitudinal dataset when the SGIC was not consistent over time; however, special precautions have to be taken to ensure the anonymity of survey respondents if personal information such as email address are collected. Future research should focus on alternative strategies that ensure anonymity but maximize survey matching capabilities.

8.3. Sample demographic information

Another area that would have improved matching over time in the current study centered on which demographic questions were consistently included across all waves. The research team did not include race or ethnicity as a question in the second wave of the study, as this was seen as a characteristic that would not change over time, and it was presumed there would be no matching difficulties. In addition, although the sample was relatively racially and ethnically diverse, more than 50 % of the sample identified as White or Asian with smaller proportions identifying as Black (7.6 %), Latino/a (6.2 %), or Other (6.2 %). Related to gender, the current study only includes people who identified as Male or Female and although we offered an "Other" check box, we did not explicitly collect information about whether students identified as transgender.

Therefore, we advise researchers to consistently ask the same demographic questions in each wave and, if possible, pilot test demographic questions to test inclusivity and the extent of test-retest reliability among college students during a time of identity formation (e.g., a transgender student might identify as a man on the first (pre-college) survey but begin identifying as a woman once she arrives to college; or, it is possible that a multi-racial student may alter how they identify on a survey if college experiences affect racial identity).

8.4. Increase transparency

Based on our experiences, we suggest that researchers incorporate missingness estimates associated with survey matching challenges when conducting ad-hoc power analyses during the design phase of longitudinal studies. Also, while findings from this study suggest certain characteristics of students may not be the most salient factors influencing whether participants' surveys can be matched into longitudinal datasets over time, we suggest future longitudinal research conduct missing data analyses specifically exploring the nature of or factors associated with missing data resulting from unmatched surveys.

8.5. Disclose matching challenges

Finally, when presenting, publishing or otherwise disseminating the results of a longitudinal study, we recommend that researchers include information about matching challenges or unmatched individuals in their sample characteristics and limitations sections. That is, researchers should be explicit about the nature of their missing data – whether it is due to matching challenges or incomplete survey responses.

9. Conclusion

When designing a longitudinal study, researchers must weigh the importance of anonymity with the possibility that participants may not provide consistent identification codes. In this paper, we hope to assist other evaluators in preventing matching problems. However, if for evaluators find themselves in a similar situation, we have outlined a number of strategies to match students who participated in a longitudinal study on sexual violence prevention and demonstrated a method to assess the extent of bias introduced by excluding unmatched individuals from longitudinal analyses. We found that some bias was introduced due to a lack of consistent SGICs provided by study participants and that some types of study participants were more likely to stay in the study. Therefore, when certain students are more likely or less likely to provide correct and consistent SGICs, the final analytic sample may contain bias that will affect the external validity of the study.

Survey matching challenges are not often acknowledged or discussed in publications presenting findings from longitudinal studies. While most researchers have become accustomed to including an estimate of missing cases in ad-hoc power analyses to estimate necessary sample sizes for efficacy trials and in analyzing the nature of missing data due to non-response and attrition after completing longitudinal studies, few consider missingness due to survey matching challenges. In fact, most researchers do not report the percentage of unmatched cases nor indicate whether there may be implications for effect sizes. It is important to consider the impact of non-reporting of this information, especially if researchers find that unmatched samples differ in significant and meaningful ways, such as being more at risk for victimization or perpetration, or being less likely to benefit from participation in the intervention.

It is incumbent on researchers, especially those reporting findings of longitudinal studies, to analyze the characteristics of matched and unmatched samples and to report the extent of sample bias due to matching issues so that the findings can be interpreted with appropriate context. Although RCTs are considered a "gold standard," randomization does not completely eliminate the threat of selection bias due to other factors associated with missing data (such as an inability to match participants over time) (Hewitt, Kumaravel, Dumville, & Torgerson, 2010).

Funding source

This work was supported by the Centers for Disease Control and Prevention (grant number 5R01/CE001855-03).

CRediT authorship contribution statement

Jane E. Palmer: Conceptualization, Resources, Data curation, Software, Writing - original draft, Writing - review & editing. **Samantha C. Winter:** Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Visualization, Resources, Writing - original draft, Writing - review & editing. **Sarah McMahon:** Supervision, Project administration, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.evalprogplan.2020.101794.

References

Ahren, K., & Le Broque, R. (2005). Methodological issues in the effects of attrition: Simple solutions for social scientists. *Field Methods, 17*, 53–69.

Anderson, R. E., Silver, K. E., Ciampaglia, A. M., Vitale, A. M., & Delahanty, D. L. (2019). The frequency of sexual perpetration in college men: A systematic review of reported prevalence rates from 2000 to 2017. *Trauma, Violence & Abuse*. https://doi.org/10.1177/1524838019860619 Advanced online publication, https://journals-sagepub-com.proxyau.wrlc.org/doi/full/10.1177/1524838019860619.

Banyard, V. (2014). Improving college campus-based prevention of violence against women: A strategic plan for research built on multipronged practices and policies. *Trauma, Violence & Abuse, 15*(4), 339–351.

Beatty, J. R., Chase, S. K., & Ondersma, S. J. (2014). A randomized study of the effect of anonymity, quasi-anonymity, and Certificates of Confidentiality on postpartum women's disclosure of sensitive information. *Drug and Alcohol Dependence, 134*, 280–284.

Bouchard, K. (2016). Anonymity as a double-edged sword: Reflecting on the implications of online qualitative research in studying sensitive topics. *The Qualitative Report, 21*(1), 59–67.

Carifio, J., & Biron, R. (1978). Collecting sensitive data anonymously: The CDRGP technique. *Journal of Alcohol and Drug Education, 23*(2), 47–66.

Damrosch, S. P. (1986). Ensuring anonymity by use of subject-generated identification codes. *Research in Nursing & Health, 9*, 61–63.

DiIorio, C., Soet, J. E., Mater, D. V., Woodring, T. M., & Dudley, W. N. (2000). Focus on research methods: An evaluation of a self-generated identification code. *Research in Nursing & Health, 23*, 167–174.

Durant, L., Carey, M., & Schroder, K. (2002). Effects of anonymity, gender, and erotophilia on the quality of data obtained from self-reports of socially sensitive behaviors. *Journal of Behavioral Medicine, 25*(5), 439–467.

Edwards, K. M., Kearns, M. C., Calhoun, K. S., & Gidycz, C. A. (2009). College women's reactions to sexual assault research participation: Is it distressing? *Psychology of Women Quarterly, 33*(2), 225–234.

Fedina, L., Holmes, J. L., & Backes, B. L. (2018). Campus sexual assault: A systematic review of prevalence research from 2000 to 2015. *Trauma, Violence & Abuse, 19*(1), 76–93.

Galanti, M., Siliquini, R., Cuomo, L., Melero, J. C., Panella, M., & Faggiano, F. (2007). Testing anonymous link procedures for follow-up of adolescents in a school-based trial: The EU-DAP pilot study. *Preventive Medicine, 44*, 174–177.

Garvey Wilson, A. L., Hoge, C. W., McGurk, D., Thomas, J. L., Clark, J. C., & Castro, C. A. (2010). Application of a new method for linking anonymous survey data in a population of soldiers returning from Iraq. *Annals of Epidemiology, 20*, 931–938.

Grossman, J., & Mackenzie, F. J. (2005). The randomized controlled trial: Gold standard, or merely standard? *Perspectives in Biology and Medicine, 48*(4), 516–534.

Grube, J. W., Morgan, M., & Kearney, K. A. (1989). Using self-generated identification codes to match questionnaires in panel studies of adolescent substance use. *Addictive Behaviors, 14*, 159–171.

Hewitt, C. E., Kumaravel, B., Dumville, J. C., & Torgerson, D. J. (2010). Assessing the impact of attrition in randomized controlled trials. *Journal of Clinical Epidemiology, 63*, 1264–1270.

Jurczyk, P., Lu, J. L., Xiong, L., Cragan, J. D., & Correa, A. (2008). FRIL: A tool for comparative record linkage. *AMIA Annual Symposium Proceedings Archive,* 440–444.

Kristjansson, A. L., Sigfusdottir, I. D., Sigfusson, J., & Allegrante, J. P. (2013). Self-generated identification codes in longitudinal prevention research with adolescents: A pilot study of matched and unmatched subjects. *Prevention Science, 15*, 205–212. https://doi.org/10.1007/s11121-013-0372-z.

Lisak, D., Conklin, A., Hopper, J., Miller, P., Altschuler, L., & Smith, B. (2000). The Abuse-Perpetration Inventory: Development of an assessment instrument for research on the cycle of violence. *Family Violence and Sexual Assault Bulletin, 16*, 21–30.

Loh, C., Gidycz, C. A., Lobo, T. R., & Luthra, R. (2005). A prospective analysis of sexual assault perpetration: Risk factors related to perpetrator characteristics. *Journal of Interpersonal Violence, 20*(10), 1325–1348.

McAlister, A., & Gordon, N. P. (1986). Attrition bias in a cohort study of substance abuse onset and prevention. *Evaluation Review, 10*, 853–859.

McGloin, J., Holcomb, S., & Main, D. S. (1996). Matching anonymous pre-posttests using subject-generated information. *Evaluation Review, 20*, 724–736.

McMahon, S., Wood, L., Cusano, J., & Macri, L. M. (2019). Campus sexual assault: Future directions for research. *Sexual Abuse, 31*(3), 270–295.

Messman-Moore, T. L., Coates, A. A., Gaffey, K. J., & Johnson, C. F. (2008). Sexuality, substance use, and susceptibility to victimization: Risk for rape and sexual coercion in a prospective study of college women. *Journal of Interpersonal Violence, 23*, 1730–1746.

Messman-Moore, T. (2019). *Email with J. Palmer.* October 15.

Ong, A., & Weiss, D. (2000). The impact of anonymity on responses to sensitive questions. *Journal of Applied Social Psychology, 30*(8), 1691–1708.

Pérez, A., Ariza, C., Sánchez-Martínez, F., & Nebot, M. (2010). Cannabis consumption initiation among adolescents: A longitudinal study. *Addictive Behaviors, 35*, 129–134.

Petersen, M. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies, 22*, 435–480.

Ripper, L., Ciaravino, S., Jones, K., Jaime, M. C. D., & Miller, E. (2017). Use of a respondent-generated personal code for matching anonymous adolescent surveys in longitudinal studies. *Journal of Adolescent Health, 60*(6), 751–753.

Rosenthal, M. N., & Freyd, J. J. (2018). Sexual violence on campus: No evidence that studies are biased due to self-selection. *Dignity: A Journal on Sexual Exploitation and Violence, 3*(1), 7.

Schnell, R., Bachteler, T., & Reiher, J. (2010). Improving the use of self-generated identification codes. *Evaluation Review, 34*, 391–418.

Ullman, S. E. (2011). Longitudinal tracking methods in a study of adult women sexual assault survivors. *Violence Against Women, 17*(2), 189–200.

van de Looij-Jansen, P. M., Goldschmeding, J. E., & de Wilde, E. J. (2006). Comparison of anonymous versus confidential survey procedures: Effects on health indicators in Dutch adolescents. *Journal of Youth and Adolescence, 35*, 652–658.

Voth Schrag, R. (2017). Campus based sexual assault and dating violence: A review of study contexts and participants. *Affilia: Journal of Women and Social Work, 32*(1), 67–80.

Yurek, L. A., Vasey, J., & Sullivan Havens, D. (2008). The use of self-generated identification codes in longitudinal research. *Evaluation Review, 32*, 435–452.

Jane E. Palmer, Ph.D., M.S.W., is a Professorial Lecturer in the department of justice, law and criminology in American University's School of Public Affairs. She is also the Director of the Community-Based Research Scholars program at American. Her research interests include prevention of and policy responses to gender-based violence and measurement/methodological issues in research on sensitive topics. She holds a Ph.D. from the department of justice, law and society at American University. Her recent research has been published in *Violence Against Women, Journal of Interpersonal Violence, Review of Higher Education* and *NASPA Journal on Women in Higher Education*.

Samantha C. Winter, M.S. is an assistant professor at the Columbia School of Social Work. Dr. Winter's research focuses on violence against women and inequities in women's health and access to health-related services, including water, sanitation, and healthcare in informal settlements in sub-Saharan Africa. Dr. Winter's most recent work as a Dorothy Byrne Postdoctoral Fellow in Global Health focused on women's access to WASH and as well as women's physical and mental health outcomes—including experiences of violence against women—in informal settlements in Kenya. Dr. Winter holds PhD and Master of Social Work degrees from Rutgers University, an MS in Environmental Engineering and

Science from Stanford University, and a BS in Civil Engineering from Colorado State University.

Sarah McMahon, Ph.D., is the Director of the Center on Violence Against Women and Children and an Associate Professor of Social Work at Rutgers University. Her research focuses on violence against women, with an emphasis on using ecological frameworks to examine prevention and social change. Dr. McMahon has extensive experience in designing and implementing quantitative and qualitative studies with college students to measure their knowledge, attitudes and behaviors related to sexual violence, with a focus on bystander intervention and campus climate. She has led a number of research projects related to campus violence funded at the federal, state, and local level and has published her work in multiple outlets.