Effects of bystander programs on the prevention of sexual assault among adolescents and college students

Heather Hensman Kettrey, Emily Tanner-Smith

Submitted to the Coordinating Group of:

☒	Crime and Justice
☐	Education
	☐ Disability
☐	International Development
	☐ Nutrition
☐	Social Welfare
☐	Methods
☐	Knowledge Translation and Implementation
☐	Other:

Plans to co-register:

☒	No		
☐	Yes	☐ Cochrane	☐ Other
☐	Maybe		

Date Submitted: 29 August 2016
Date Revision Submitted: 31 March 2017
Approval Date: 9 May 2017
Publication Date: 10 May 2017

BACKGROUND

Sexual Assault among Adolescents and College Students

Sexual assault is a significant problem among adolescents and college students in the United States. Findings from the Campus Sexual Assault study estimated that 15.9% of college women had experienced attempted or completed sexual assault (i.e., unwanted sexual contact that could include sexual touching, oral sex, intercourse, anal sex, or penetration with a finger or object) prior to entering college and 19% had experienced attempted or completed sexual assault since entering college (Krebs, Lindquist, Warner, Fisher, & Martin, 2009). These rates are problematic, as sexual assault in adolescence and/or young adulthood is associated with numerous adverse outcomes, including repeated victimization, depressive symptomology, heavy drinking, and suicidal ideation (Cortens, Eckenrode, & Rothman, 2013; Cui, Ueno, Gordon, & Fincham, 2013; Halpern, Spriggs, Martin, & Kupper, 2009). Importantly, there is evidence indicating experiences of sexual assault during these two life phases are related, as victimization and perpetration during adolescence are, respectively, associated with increased risk of victimization and perpetration during young adulthood (Cui, Ueno, Gordon, & Fincham, 2013). Thus, early prevention efforts are of paramount importance.

Reviews of research on the effectiveness of programs designed to prevent sexual assault among adolescents and college students have noted both a dearth of high-quality studies, such as randomized controlled trials (RCTs), and minimal evidence that these prevention programs have meaningful effects on young people's behavior (DeGue, Valle, Holt, Massetti, Matjasko, & Tharp, 2014; De Koker, Mathews, Zuch, Bastien, & Mason-Jones, 2014). Concerning the latter point, evaluations of such programs tend to measure attitudinal outcomes (e.g., rape supportive attitudes, rape myth acceptance) more frequently than behavioral outcomes (e.g., perpetration or victimization) (Anderson & Whiston, 2005; Cornelius & Resseguie, 2007; DeGue et al., 2014). Additionally, findings from a meta-analysis of studies assessing outcomes of college sexual assault prevention programs suggested that effects are larger for attitudinal outcomes than for the actual incidence of sexual assault (Anderson & Whiston, 2005).

The Bystander Approach

Given this paucity of evidence of behavior change, it is imperative that researchers identify effective strategies for preventing sexual assault among adolescents and young adults. One promising strategy is the implementation of bystander programs, which encourage young people to intervene when witnessing incidents or warning signs of sexual assault. The strength of the bystander model lies in its emphasis on the role of peers in the prevention of violence. Peers are a salient influence on young people's intimate relationships (Adelman & Kil, 2007; Giordano, 2003). In some respects, this can be detrimental, as having friends involved in violent intimate relationships (i.e., characterized by sexual or physical violence)

is a risk factor for becoming both a perpetrator and victim of violence (Arriaga & Foshee, 2004; Foshee, Benefield, Ennett, Bauman, & Suchindran, 2004; Foshee, Linder, MacDougall, & Bangdiwala, 2001; Foshee, Reyes, & Ennett, 2010; McCauley et al., 2013). However, peers can also have a positive impact on intimate relationships.

Young victims and perpetrators of violence are often reluctant to divulge their experience or to seek help (especially from adults), but when they do seek help they often seek it from their peers (Ashley & Foshee, 2005; Black, Tolman, Callahan, Saunders, & Weisz, 2008; Molidor & Tolman, 1998; Weisz, Tolman, Callahan, Saunders, & Black, 2007). This indicates that victims trust their peers to provide a valuable source of support after an assault has occurred. Importantly, it also suggests that peers have the potential to play a pivotal role in the prevention of sexual assault by intervening when they witness its warning signs. In fact, in a contemporary "hookup culture" adolescents and young adults are more likely to meet and socialize in groups than they are to date in pairs and, thus, warning signs of assault are frequently exhibited in communal spaces (Bogle, 2007; 2008; Molidor & Tolman, 1998). Thus, the social nature of intimate relationships during these life stages can make peers pivotal actors in the prevention of sexual assault.

However, the potential for peer intervention can be undermined by a general "bystander effect" that diffuses responsibility for action in group settings (Darley & Latane, 1968). To intervene as a witness to sexual assault, individuals must notice the event (or its warning signs), define the event as warranting action/intervention, take responsibility for acting (i.e., feel a sense of personal duty), and demonstrate a sufficient level of self-efficacy (i.e., perceived competence to successfully intervene) (Latane & Darley, 1969). Studies have indicated that, as witnesses to sexual assault, young people often fail to meet these criteria (Banyard, 2008; Bennett, Banyard, & Garnhart, 2014; Burn, 2009; Casey & Ohler, 2012; Exner & Cummings, 2011; McCauley et al., 2013; McMahon, 2010; Noonan & Charles, 2009), with males being less likely than females to intervene (Banyard, 2008; Burn, 2009; Edwards, Rodenhizer-Stampfli, & Eckstein, 2015; Exner & Cummings, 2011; McMahon, 2010).

Thus, bystander programs seek to sensitize young people to warning signs of sexual assault, create attitudinal change that fosters bystander responsibility for intervening (e.g., creating empathy for victims), and build requisite skills/tactics for taking action (Banyard, 2011; Banyard, Plante, & Moynihan, 2004; Burn, 2009; McMahon & Banyard, 2012). Many of these programs are implemented with large groups of adolescents or college students in the format of a single training/education session (e.g., as part of college orientation). However, some programs use broader implementation strategies, such as advertising campaigns where signs are posted across college campuses to encourage students to act when witnessing signs of violence.

How the Intervention Might Work

By treating young people as potential allies in preventing sexual assault, bystander programs have the potential to be less threatening than traditional sexual assault prevention programs, which tend to approach young people as either potential perpetrators or victims of sexual violence (Burn, 2009; Messner, 2015; [Jackson] Katz, 1995). Instead of placing emphasis on how young people may modify their individual behavior to either respect the sexual boundaries of others or reduce their personal risk for being sexually assaulted, bystander programs aim to foster prerequisite knowledge and skills for intervening on behalf of victims. Thus, by treating young people as part of the solution to sexual assault, rather than part of the problem, bystander programs limit the risk of defensiveness or backlash among participants (e.g., decreased empathy for victims, increased rape myth acceptance) (Banyard et al., 2004; Katz, 1995).

As outlined by Burn (2009) bystander programs are designed to promote the following prerequisites for intervention: noticing an event, identifying a situation as warranting intervention, taking responsibility for acting, and deciding how to help. This often involves educating young people about what constitutes sexual assault, portraying victims as worthy of assistance, and building skills necessary to intervene (e.g., providing strategies for what to do and say). Although most bystander programs share the common goal of promoting such prerequisites for intervention, their specific program content and framing of sexual assault varies.

Research has indicated that, relative to males, females are overwhelmingly the victims of sexual assault (Foshee, 1996; Gressard, Swahn, & Tharp, 2015; Harned, 2001; Howard, Wang, & Yan, 2007). Thus, the earliest bystander programs tended to apply a gendered perspective to the prevention of sexual assault among adolescents and college students. For example, Katz (1995) developed the Mentors in Violence Prevention Program (MVP) with the goal of inspiring male college athletes to challenge sociocultural definitions of masculinity that equate men's strength with dominance over women. At the time of its inception, MVP was unique in its explicit focus on masculinity as well as its nonthreatening "bystander" approach that encouraged young men to intervene when witnessing acts (or warning signs) of violence against women. As Katz explained, MVP reduces young men's defensiveness to violence prevention efforts by focusing on men as potential *bystanders to violence*, rather than potential *perpetrators of violence*. In addition to reducing men's defensiveness to intervention efforts, this bystander approach emphasizes the point that "when men don't speak up or take action in the face of other men's abusive behavior toward women, that constitutes implicit consent of such behavior" (Katz, 1995, p. 168).

Since the inception of MVP a number of programs have emerged to address barriers to bystander intervention among adolescents and college students. While they all share the common goal of inspiring bystanders to act in ways that prevent sexual assault, these programs exhibit a great deal of variation in scope pertaining to their target bystander

populations (i.e., males and/or females, secondary school or college students), sex of victims, and gendered versus gender-neutral approach. For example, some programs use a gendered approach by (1) critiquing gender norms that can promote violence against women and (2) encouraging males to intervene on behalf of female victims (e.g., MVP, see Katz, 1995). Others use a gender-neutral approach to build a sense of community responsibility to intervene on behalf of both male and female victims of sexual assault (e.g., Bringing in the Bystander, see Banyard, Moynihan, & Crossman, 2009; Banyard, Moynihan, & Plante, 2007). One of the major differences between gendered and gender-neutral bystander programs is that the former places socio-cultural forces, such as gender norms, at the center of discussions of violence whereas the latter places the bystander, and his/her individual cognitive processes when encountering violence, at the center of discussions of violence ([Jackson] Katz, Hesiterkamp, & Fleming, 2011; Messner, 2015).

Comparing the effects of gendered and gender-neutral programs has the potential to identify important determinants of the success of bystander programs. Although there is no empirical examination of the different effects of these programs, there are theoretical reasons to believe that each has the potential to be successful under certain conditions. Specifically, gendered approaches to bystander education programs may be better suited to target socio-cultural facilitators of sexual assault against women and address different patterns of bystander behaviors exhibited by males and females (Banyard, 2008; Burn, 2009; Exner & Cummings, 2011; Katz et al., 2011; [Jennifer] Katz, 2015; [Jennifer] Katz, Colbert, & Colangelo, 2015; McCauley et al., 2013; McMahon, 2010; Messner, 2015). On the other hand, gender-neutral programs have the benefit of deflecting the criticism that prevention programs utilizing a gendered approach are inherently anti-male (Katz et al., 2011; Messner, 2015). Avoidance of such criticism is paramount to the success of sexual assault prevention programs. This is because adolescents and young adults who are coming of age in a "post-feminist" era are likely to reject gendered explanations of sexual assault and, instead, respond more positively to gender-neutral programs that use inclusive language that can be applied to a broad range of victims and perpetrators (Barreto & Ellemers, 2005; Kettrey, 2016; Swim, Aikin, Hall, & Hunter, 1995).

Why it is Important to do the Review

Illustrating the perceived benefits of bystander programs, the 2013 Campus Sexual Violence Elimination (SaVE) Act requires post-secondary educational institutions participating in Title IX financial aid programs to provide incoming college students with primary prevention and awareness programs addressing sexual violence. The Campus SaVE Act mandates that these programs include a component on bystander intervention. Currently, there is no comparable legislation regarding sexual assault among adolescents (e.g., mandating bystander programs in secondary schools). This is an unfortunate oversight, as adolescents who experience sexual assault are at an increased risk of repeated victimization in young adulthood (Cui et al., 2013). Thus, the implementation of bystander programs in

secondary schools not only has the potential to reduce sexual assault among adolescents, but may also have the long-term potential to reduce sexual assault on college campuses.

Findings from this systematic review will provide valuable evidence of the extent to which bystander programs, as mandated by the Campus SaVE Act, are effective in preventing sexual assault among college students. Additionally, by examining effects of these programs among adolescents, this review will provide educators and policy makers with information for determining whether such programs should be widely implemented in secondary schools.

Currently, there are no Campbell or Cochrane Collaboration Reviews evaluating the effects of bystander programs on sexual assault among adolescents and/or college students. Of modest relevance to the proposed review, the Campbell and Cochrane Collaboration libraries include meta-analyses of the effects of more general programs (not bystander programs) designed to prevent or reduce relationship/dating violence among adolescents and/or young adults (De La Rue, Polanin, Espelage, & Pigott, 2014; Fellmeth, Heffernan, Nurse, Habibula, & Sethi, 2013). Both of these reviews reported violence outcomes as aggregate measures that do not distinguish sexual violence from other forms of violence. Although they each found some evidence of significant effects on knowledge or attitudes pertinent to violence, neither found evidence of significant effects on young people's behavior (i.e., rates of perpetration or victimization).

Two reviews published outside of the Campbell and Cochrane Collaboration libraries are of closer relevance to this review. These include a meta-analysis of the effects of bystander programs on sexual assault on college campuses (Katz & Moore, 2013) and a narrative review of studies examining the effects of bystander programs on dating and sexual violence among adolescents and young adults (Storer, Casey, & Herrenkohl, 2015).

In what they called an "initial" meta-analysis of experimental and quasi-experimental studies (published through 2011), Katz & Moore (2013) found moderate effects of bystander programs on participants' self-efficacy and intentions to intervene, and small (but significant) effects on bystander behavior, rape-supportive attitudes, and rape proclivity (but not perpetration). Effects were generally stronger among younger samples and samples containing a higher percentage of males. The stronger effect for younger participants (i.e., younger college students) suggests such programs may be particularly effective with adolescents.

In a narrative review of studies examining the effects of bystander programs on dating violence and sexual assault among adolescents and young adults, Storer et al. (2015) highlighted positive effects on bystander self-efficacy and intentions, but noted less evidence of positive effects on actual bystander behavior or of effects on perpetration of violence.

While informative, each of these reviews has limitations. Katz & Moore's (2013) meta-analysis focused exclusively on sexual assault on college campuses and did not examine effects of such programs among adolescents. Although Storer et al. (2015) focused on studies

examining violence among both adolescents and young adults, their sample was limited in that it was exclusively comprised of peer-reviewed articles (i.e., the sample explicitly excluded theses, dissertations, and other gray literature) and the authors specified no research design criteria for inclusion (i.e., the sample included low-quality studies such as those utilizing single group pre- and post-test designs). Importantly, Storer et al. reported no meta-analytic findings. Thus, to our knowledge there are currently no existing meta-analyses examining the effects of bystander programs on attitudes and behaviors regarding sexual assault among both college students and adolescents. Additionally, Katz & Moore's early meta-analysis only included studies published/reported through 2o11 (two years prior to the 2013 Campus SaVE Act) and did not evaluate program content as a moderator.

Our review will examine the effects of bystander programs on attitudes (i.e., perceptions of violence/victims, self-efficacy to intervene, intentions to intervene) and behaviors (i.e., actual intervention behavior, perpetration) regarding sexual assault among adolescents and college students. Importantly, it will present meta-analytic findings to quantitatively assess the influence of moderators (e.g., gender composition of sample, mean age of sample, education level of sample, single- or mixed- sex implementation, gendered content of program, fraternity/sorority membership, athletic team membership) on the effects of bystander programs.

Given that the 2013 Campus SaVE Act mandates bystander education as a component of primary prevention of sexual assault on college campuses, findings will provide vital assessment of legislation outcomes. Specifically, findings will (1) provide an updated analysis of the effects of bystander programs, (2) examine previously unanalyzed moderators of the effectiveness of bystander programs (e.g., content), and (3) evaluate the effects of bystander programs among adolescents. This latter contribution has the potential to inform directions for future policy and/or legislation pertinent to prevention efforts in secondary schools. This is of vital importance in breaking patterns of sexual assault, considering that victimization and perpetration in adolescence are associated with victimization and perpetration in young adulthood (Cui et al., 2013).

OBJECTIVES

The overall objective of this systematic review and meta-analysis is to examine what effects bystander programs have on preventing sexual assault among adolescents and college students. More specifically, and given the study designs that will be included, this review will address three objectives:

The first objective is to assess the overall effects (including adverse effects), and the variability of the effects, of bystander programs on adolescents' and college students' attitudes and behaviors regarding sexual assault. This includes general attitudes toward

violence and victims, prerequisite skills and knowledge to intervene, self-efficacy to intervene, intentions/willingness to intervene when witnessing signs of sexual assault, actual intervention behavior, and perpetration of sexual assault. This information will allow practitioners and policy makers to determine whether bystander programs, which are mandated by the Campus SaVE Act, are effective in preventing sexual assault among adolescents and college students. Furthermore, the examination of adverse effects will provide decision-makers with useful information about any backlash effects of bystander programs (e.g., decreased victim empathy or increased rape myth acceptance). This is important, as the non-threatening approach of bystander programs is believed to present minimal risk of backlash effects (Banyard et al., 2004; Katz, 1995).

The second objective is to explore the comparative effectiveness of bystander programs for different profiles of participants (e.g., mean age of the sample, education level of the sample, proportion of males/females in the sample, proportion of fraternity/sorority members in the sample, proportion of athletic team members in the sample). This information will allow practitioners and policy makers to determine which groups of young people might benefit most from bystander programs.

The third objective is to explore the comparative effectiveness of different bystander programs in terms of gendered content and approach (e.g., conceptualizing sexual assault as a gendered or gender-neutral problem, mixed- or single-sex group implementation). This information will allow practitioners and policy makers to identify characteristics of the most effective bystander programs.

METHODOLOGY

Criteria for including and excluding studies

Types of study designs

To be eligible for inclusion in the review, studies must use an experimental or controlled quasi-experimental research design to compare an intervention group (e.g., students assigned to complete a bystander program) with a comparison group (e.g., students not assigned to complete a bystander program). We limit our review to such studies because these typically present lower risk of bias relative to other research designs (e.g., natural experiments, single group designs). More specifically, we plan to include the following designs:

1. Randomized controlled trials: Studies in which individuals, classrooms, schools, or other groups are randomly assigned to intervention and comparison conditions.

2. Quasi-randomized controlled trials: Studies where assignment to conditions is quasi-random, for example, by birth date, date of week, student identification number, month, or some other alternation method.

3. Controlled quasi-experimental designs: Studies where participants are not assigned to conditions randomly or quasi-randomly (e.g., participants self-select into groups). Given the potential selection biases inherent in these controlled quasi-experimental design, we will only include those that also meet one of the following criteria:

    a. Regression discontinuity designs: Studies that use a cutoff on a forcing variable to assign participants to intervention and comparison groups, and assess program impacts around the cutoff of the forcing variable.

    b. Studies that use propensity score or other matching procedures to create a matched sample of participants in the intervention and comparison groups. To be eligible for inclusion, these studies must also provide enough statistical information to permit estimation of pretest effect sizes for the matched groups.

    c. For studies where participants in the intervention and comparison groups are not matched, enough statistical information must be reported that will permit us to estimate pretest effect sizes for at least one outcome measure.

Consistent with Campbell Collaboration policies and procedures, studies using experimental and quasi-experimental research designs will be synthesized separately in the meta-analyses. This is because experimental studies present the lowest risk of bias (Higgins & Green 2011). Furthermore, we plan to collect extensive data on the risk of bias and study quality of all eligible studies, which we will attend to carefully when interpreting the findings from the systematic review and meta-analyses (described in greater detail below).

Types of participants

The review will focus on studies that examine outcomes of bystander programs that target sexual assault and are implemented with adolescents and/or college students in educational settings. Eligible participants include adolescents enrolled in grades 7 through 12 and college students enrolled in any type of undergraduate postsecondary educational institution.

Eligible participant populations will include studies that report on general samples of adolescents and/or college students as well as studies using specialized samples such as those primarily consisting of college athletes, fraternity/sorority members, and single-sex samples. Study samples primarily consisting of post-graduate students will be ineligible for inclusion; the mean age of samples may be no greater then 25 to be included in the review.

Types of interventions

Eligible intervention programs are those that approach participants as allies in preventing and/or alleviating sexual assault among adolescents and/or college students. Some part of the program must focus on ways that cultivate a willingness for a person to respond to others who are at risk for sexual assault. All delivery formats are eligible for inclusion (e.g., in-person training sessions, video programs, web-based training, advertising/poster campaigns). There are no treatment duration criteria for inclusion.

Studies that report bystander outcomes but do not meet the aforementioned intervention criteria are not eligible for inclusion. Additionally, studies that assess outcomes of programs that aim to facilitate pro-social bystander behavior, but that do not explicitly include a component addressing sexual assault (e.g., programs to prevent bullying) are not eligible.

Eligible comparison groups must receive no intervention services targeting bystander attitudes/behavior or sexual assault. Thus, treatment-treatment studies that compare individuals assigned to a bystander program with individuals assigned to a general sexual assault prevention program are not eligible for inclusion. However, eligible comparison groups may receive a sham or attention treatment that is expected to have no effect on bystander outcomes or attitudes/behaviors regarding sexual assault.

Types of outcome measures

The review will include studies that measure the effects of bystander programs on at least one of the following primary outcome domains:

1. General attitudes toward sexual assault and victims (e.g., victim empathy, rape myth acceptance).

2. Prerequisite skills and knowledge for bystander intervention as defined by Burn (2009) (e.g., noticing sexual assault or its warning signs, identifying a situation as appropriate for intervention, taking responsibility for acting/intervening, knowing strategies for helping/intervening).

3. Self-efficacy with regard to bystander intervention (e.g., respondents' confidence in their ability to intervene).

4. Intentions to intervene when witnessing instances or warning signs of sexual assault.

5. Actual intervention behavior when witnessing instances or warning signs of sexual assault.

6. Perpetration of sexual assault (i.e., rates of perpetration among individuals assigned to the treatment or control group of a study).

Any outcome falling in these domains will be eligible for inclusion. This includes outcomes measured with any form of assessment (e.g., self-report, official/administrative report,

observation, etc.) that can be summarized by any type of quantitative score (e.g., percentage, continuous variable, count variable, categorical variable, etc.). In the event that a particular study includes multiple measures of a single construct (e.g., two different scales measuring rape myth acceptance) we will code each eligible outcome and synthesize these using the methods described later in this protocol (see "Criteria for determination of independent findings" below).

Depending on directionality, these outcomes capture both intended and adverse effects (e.g., increases or decreases in victim empathy, pro-social bystander behavior, etc.) that are important to adolescents, college students, and decision-makers alike.

Duration of follow-up

Studies reporting follow-ups of any duration are eligible for inclusion. When studies report more than one follow-up wave, each wave will be coded and identified by its reported duration. As described in more detail below, follow-ups of similar durations will be analyzed together.

Types of settings

The review will focus on studies that examine outcomes of bystander programs that target sexual assault and are implemented with adolescents and/or college students in educational settings. Eligible educational settings include secondary schools (i.e., grades 7-12) and colleges or universities. Studies that assess bystander programs implemented with adolescents and young adults outside of educational institutions (e.g., in community settings, military settings) are ineligible for inclusion in the review.

Search strategy

We will identify candidate studies through searches of electronic databases, relevant academic journals, and gray literature sources. We will also contact leading authors and experts on bystander programs to identify any current/ongoing research that might be eligible for the review. Additionally, we will screen the bibliographies of eligible studies and relevant reviews to identify additional candidate studies. We will conduct forward citation searches (searches for reports citing eligible studies) using the website Google Scholar, as this database produces similar results to other search engines (e.g., Web of Science; Tanner-Smith & Polanin, 2015) and is also more likely to locate existing grey literature.

Once candidate studies have been identified each will be entered into the meta-analysis database as a separate record. Two reviewers will then independently screen each study title and abstract and record their eligibility recommendation (i.e., ineligible or eligible for full-text screening) into the pertinent database record. Disagreements between reviewers will be resolved by consensus and the final abstract screening decision will be recorded in the database. Potentially eligible studies will then be retrieved in full text and these full texts will

be reviewed for eligibility, again using two independent reviewers who will record their eligibility recommendation (and, when applicable, rationale for an ineligibility recommendation) into the pertinent database record. Disagreements between reviewers will again be resolved via discussion and consensus and the final eligibility decision will be recorded in the database. If we cannot determine eligibility due to missing information in a report, we will contact study authors for this information.

Throughout the search and screening process we will maintain a document that includes the number of unique candidate studies identified through various sources (e.g., electronic database searches, academic journal searches, and gray literature searches). We will use the information in this record to create a PRISMA flow chart that reports the screening process (Moher et al., 2015). Additionally, we will use the final screening decisions recorded in the meta-analysis database to create a table that lists studies excluded during the full-text screening phase along with the rationale for each exclusion decision.

Electronic Databases

The prevention of sexual assault among college students and adolescents is a topic that spans multiple disciplines (e.g., sociology, psychology, education, public health). Thus, we will search a variety of databases that are relevant to these fields. Search terms will vary by database, but will generally include two blocks of terms and appropriate Boolean or proximity operators, if allowed: blocks will include terms that address the intervention and outcomes. The following are electronic databases (hosts) we plan to search:

- Cochrane Central Register of Controlled Trials (CENTRAL)
- Cochrane Database of Abstracts of Reviews of Effects (DARE)
- Education Resources Information Center (ERIC, via ProQuest)
- Education Database (via ProQuest)
- International Bibliography of the Social Sciences (IBSS, via ProQuest)
- PsycINFO (via ProQuest)
- PsychARTICLES (via Proquest)
- PubMed
- Social Services Abstracts (via Proquest)
- Sociological Abstracts (via Proquest)

The strategy for searching electronic databases will involve the use of search terms specific to the types of interventions and outcomes eligible for inclusion. Search terms for types of interventions will include general terms for bystander programs as well as names of specific bystander programs (e.g., Mentors in Violence Prevention). Search terms for types of outcomes will include terms that are specific to measures of sexual violence (e.g., sexual assault) as well as more general terms that have the potential to identify studies that measure physical and/or sexual violence. Due to the overwhelming focus of bystander programs on adolescents and college students (aside from a few implementations with military samples) search terms will not limit initial results by the age or general

characteristics of the target population.

The search terms and strategy for PsycINFO via ProQuest are as follows (to be modified for other databases):

(AB,TI("bystander")) AND (AB,TI("education" OR "program" OR "training" OR "intervention" OR "behavior" OR "attitude" OR "intention" OR "efficacy" OR "prosocial" OR "pro-social" OR "empowered" OR "Bringing in the Bystander" OR "Green Dot" OR "Step Up" OR "Mentors in Violence Prevention" OR "MVP" OR "Know Your Power" OR "Hollaback" OR "Circle of 6" OR "That's Not Cool" OR "Red Flag Campaign" OR "Where Do You Stand" OR "White Ribbon Campaign" OR "Men Can Stop Rape" OR "The Men's Program" OR "The Women's Program" OR "The Men's Project" OR "Coaching Boys into Men" OR "Campus Violence Prevention Program" OR "Real Men Respect" OR "Speak Up Speak Out" OR "How to Help a Sexual Assault Survivor" OR "InterACT")) AND (AB,TI("sexual assault" OR "rape" OR "violence" OR "victimization"))

Searches of Academic Journals

We will search the tables of contents of current issues of journals where a large proportion of eligible studies are published. Preliminary searches of the literature suggest the following journals may be promising sources of eligible studies: *Journal of Adolescent Health, Journal of Community Psychology, Journal of Family Violence, Journal of Interpersonal Violence, Psychology of Violence, Violence Against Women, and Violence and Victims.*

Gray Literature Sources

We will conduct gray literature searches to identify unpublished studies that meet inclusion criteria. This will include searching electronic databases that catalog dissertations and theses, searching conference proceedings, and searching websites with content relevant to sexual assault and/or violence against women. The following are gray literature sources we plan to search:

- ProQuest Dissertations and Theses Global
- Clinical Trials Register (https://clinicaltrials.gov)
- End Violence Against Women International - conference proceedings
  (http://www.evawintl.org/conferences.aspx)
- National Sexual Violence Resource Center website (nsvrc.org)
- National Online Resource Center on Violence Against Women website (VAWnet.org)
- Us Department of Justice Office on Violence Against Women website
  (www.justice.gov/ovw)
- Center for Changing our Campus Culture website
  (www.changingourcampusculture.org)

Description of methods used in primary research

Based on preliminary searches of the literature, we anticipate that most included studies will employ randomized control trial (RCT) designs in which participants are randomly assigned to an intervention or comparison group – either individually or in clusters. Examples of clusters might include classrooms, groups of incoming college students attending orientation, fraternities/sororities, athletic teams, etc.

In a well-cited example of an RCT with individual-level assignment Banyard, Moynihan, and Plante (2007) recruited undergraduates through flyers posted around campus. They then randomly assigned participants to one of three groups: (1) a no-treatment control that only completed pre- and post- intervention surveys, (2) a treatment group that was assigned to complete one 90-minute bystander program session, or (3) a treatment group that was assigned to complete three 90-minute bystander program sessions. Note that, for the purposes of this review, we will handle such multiple-arm studies by contrasting each treatment group with the comparison group in a series of sensitivity analyses (described in greater detail below). That is, we do not plan to compare multiple treatment groups to one another.

In an example of an RCT with cluster assignment, Miller et al. (2012) randomly assigned 16 high schools to either (1) a treatment condition in which coaches were trained to lead athletes in eleven brief (10-15 minute) discussions on respect and violence prevention or (2) a comparison condition in which coaches were instructed to interact with athletes in their usual fashion.

Although we expect most included studies to be RCTs, we also anticipate that some studies conducted after the institution of the 2013 Campus SaVE Act will use controlled quasi-experimental designs. The Campus SaVE Act mandates that all post-secondary educational institutions participating in Title IV financial aid programs implement primary prevention/awareness programs that address sexual violence (and that they include a bystander component). Thus, we suspect that many studies conducted since 2013 will summarize findings from routine programs (e.g., programs incorporated into college orientation for incoming students) rather than programs implemented strictly for research purposes. Due to both logistical and ethical issues, these routine programs are unlikely to involve randomization of participants. Instead, we expect many of these studies will involve non-random assignment of participants, but report will pre-intervention measures of outcomes that may be used to measure equivalence between the treatment and comparison groups.

Criteria for determination of independent findings

Throughout the screening and coding processes, we will flag and set aside reports for determination of independence if they meet one or more of the following criteria:

1. They share one or more common author with another report.

2. They summarize findings from similar datasets (e.g., data collected on the same college campus).

3. They report the same grant number or funding source.

If we determine that the results of two or more reports are related, we will assign the reports the same study number and will code them concurrently as a single "study." If it is unclear whether two or more reports are independent we will contact the authors for clarification.

Within each independent study, we also anticipate that most studies will report on multiple outcome measures at multiple follow-up points. We will group coded effect sizes together according to the construct that they represent (i.e., attitudes toward sexual assault and victims, prerequisite skills and knowledge to intervene, self-efficacy to intervene, intentions to intervene, actual intervention behavior, perpetration of sexual assault) and the follow-up assessment period (e.g., 0-3 months, 3-6 months, and 6+ month follow-up). All quantitative syntheses (meta-analyses) will then synthesize findings separately by outcome construct and follow-up time points, to ensure the statistical independence of the effect sizes included in any given analysis.

We will handle within-study effect size dependency (e.g., more than one effect size reported per outcome construct/follow-up time point within a single study) differently depending on how many studies are eligible for the review. Specifically, robust variance estimation, a method for handling dependent effect sizes, is only appropriate in cases with larger samples. We use a cut-off sample size of 10, as recommended by Tanner-Smith and Tipton (2014) for determining the appropriateness of robust variance estimation. In the event that there are less than 10 studies that report multiple outcomes within the same construct category (e.g., two measures of rape myth acceptance within each study), we will only include one outcome per study for that construct. We will select the most similar outcomes for synthesis within a construct category. In the event that there are ten or more studies that report multiple outcomes within the same construct we will use robust variance estimation to handle statistical dependencies in the synthesis (Hedges, Tipton, & Johnson, 2010; Tanner-Smith & Tipton, 2014; Tipton, 2013).

Details of study coding categories

Two reviewers will independently double-code all included studies, using a piloted codebook (see Appendix for a draft codebook). Coding disagreements will be resolved via discussion and consensus. If data needed to calculate an effect size are missing from a report, we will contact the primary study authors for this information.

The primary categories for coding are as follows: participant demographics and characteristics (e.g., age, gender, education level, race/ethnicity, athletic team membership, fraternity/sorority membership); intervention setting (e.g., state, country, secondary or post-secondary institution, mixed- or single-sex group); study characteristics (e.g., attrition,

duration of follow-up, study design, participant dose, sample N); outcome construct (e.g., type, description of measure); and outcome results (e.g., timing at measurement, baseline and follow-up means and standard deviations or proportions).

We will assess risk of bias in included studies using the Cochrane risk of bias tools for randomized studies (Higgins & Green, 2011) and non-randomized studies (Sterne et al., 2016). The Cochrane risk of bias tool for randomized studies assesses risk of bias in the following domains: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias. These domains are each rated as low, unclear, or high risk of bias.

The Cochrane risk of bias tool for non-randomized studies, ROBINS-I, requires that at the protocol stage of the review, two sets of items are determined: (1) confounding areas that are expected to be relevant to all or most studies in the review and (2) co-interventions that could be different between intervention groups with the potential to differentially impact outcomes. For this review, we anticipate the following confounding factors, which will be coded in the risk of bias assessments: gender, fraternity/sorority membership, athletic team membership, pre-intervention attitudes (e.g., victim empathy, rape myth acceptance), pre-intervention bystander measures (e.g., efficacy, intentions, behavior), and prior sexual assault victimization. We anticipate the following co-interventions to have a potential impact on outcomes, so they will also be coded in the risk of bias assessments: general sexual assault prevention programs, dating violence prevention programs, and general bystander programs (not explicitly targeting sexual violence).

Statistical procedures and conventions

Effect Sizes. Outcomes for included studies will likely be reported as both binary measures (e.g., proportion of participants who report bystander intervention behavior) and continuous measures (e.g., scores on rape myth acceptance scales). For binary measures, we will report effect sizes in the form of odds ratios (OR). For continuous measures, we will report effect sizes as standardized mean differences (SMD), adjusted for small sample size (Hedges' *g*). We will code OR effect sizes such that positive values (i.e., greater than 1) indicate greater odds of a beneficial outcome for the treatment group (relative to the comparison group), whereas negative values (i.e., less than 1) indicate lower odds of success for the treatment group (i.e., an adverse effect). We will code SMD effect sizes such that positive values (i.e., greater than 0) indicate a beneficial outcome for the treatment group. In the event that similar outcome constructs are represented as both binary and continuous measures across or within studies (e.g., proportion of respondents who report bystander intervention behavior, mean number of bystander intervention behaviors reported) we will assume that these measures reflect different underlying constructs. Thus, we will synthesize the OR and SMD effect sizes separately.

Additionally, to minimize any potential bias in the meta-analysis results due to effect size outliers, prior to conducting the meta-analysis, we will Winsorize effect sizes or study sample sizes that lie three or more standard deviations from the mean (Lipsey & Wilson, 2001).

Model. We will conduct meta-analyses using random-effects inverse variance weights and will report 95% confidence intervals along with all mean effect sizes. We will conduct and report all meta-analyses separately for RCTs and non-RCTs. Additionally, we will conduct all syntheses separately by outcome domain (i.e., attitudes toward sexual assault and victims, prerequisite skills and knowledge to intervene, bystander self-efficacy, bystander intentions, bystander behavior, perpetration) and follow-up timing. We will display all syntheses using forest plots.

Assessment of Heterogeneity and Moderator Analysis. We will assess and report heterogeneity using the χ^2 statistic and it's corresponding p value, *I^2, and τ^2.* If at least 10 studies are included in any given meta-analysis, we will then use mixed-effect meta-regression models to conduct moderator analyses. Proposed moderators include the following: (1) gendered content of the program; (2) mixed- or single-sex group implementation; (3) gender composition of the sample; (4) education level of the sample (i.e., secondary school or college students); (5) mean age of the sample; (6) proportion of fraternity/sorority members in the sample; (7) and proportion of athletic team members in the sample.

Sensitivity Analyses. We will use the contour enhanced funnel plot (Palmer, Peters, Sutton, & Moreno, 2008), Egger's regression test (Egger, Smith, Schneider, & Minder, 1997), and a trim and fill analysis (Duval & Tweedie, 2000) to evaluate the potential of publication bias or small study bias in the included studies.

We will also conduct sensitivity analyses to examine whether study-level attrition and high risk of bias (for each domain assessed with the risk of bias tools) are associated with effect size magnitude, using mixed-effects meta-regression models. Additionally, we will conduct sensitivity analyses that remove any Winsorized effect sizes or sample sizes from the meta-analysis, to assess whether this method for handling outliers may have substantively altered the review findings.

Data Management and Software

We will use the following software throughout the review process: FileMaker Pro (literature screening and data extraction), Microsoft Excel, Microsoft Word, and *R* (synthesis, generation of figures including forest plots).

Treatment of qualitative research

Qualitative research is not eligible for inclusion in this review.

REFERENCES

Adelman, M. & Kil, S. H. (2007). Dating conflicts: Rethinking dating violence and youth conflict. *Violence Against Women, 13,* 1296-1318. doi: 10.1177/1077801207310800

Anderson, L. A., & Whiston, S. C. (2005). Sexual assault education programs: A meta-analytic examination of their effectiveness. *Psychology of Women Quarterly, 29,* 374-388.

Arriaga, X. B. & Foshee, V. A. (2004). Adolescent dating violence: Do adolescents follow in their friends', or their parents', footsteps? *Journal of Interpersonal Violence, 19,* 162-184. doi: 10.1177/0886260503260247

Ashley, O. S., & Foshee, V. A. (2005). Adolescent help-seeking for dating violence: Prevalence, sociodemographic correlates, and sources of help. *Journal of Adolescent Health, 36,* 25-31. http://dx.doi.org/10.1016/j.jadohealth.2003.12.014

Banyard, V. L. (2008). Measurement and correlates of prosocial bystander behavior: The case of interpersonal violence. *Violence and Victims, 23,* 83-97. doi: http://dx.doi.org/10.1891/0886-6708.23.1.83

Banyard, V. L. (2011). Who will help prevent sexual violence? Creating an ecological model of bystander intervention. *Psychology of Violence, 1,* 216-229. http://psycnet.apa.org/doi/10.1037/a0023739

Banyard, V. L., Moynihan, M. M., & Crossman, M. T. (2009). Reducing sexual violence on campus: The role of student leaders as empowered bystanders. *Journal of College Student Development, 50,* 446-457. doi: 10.1353/csd.0.0083

Banyard, V. L., Moynihan, M. M., & Plante, E. G. (2007). Sexual violence prevention through bystander education: An experimental evaluation. *Journal of Community Psychology, 35,* 463-481. doi: 10.1002/jcop.20159

Banyard, V. L., Plante, E. G., & Moynihan, M. M. (2004). Bystander education: Bringing a broader community perspective to sexual violence prevention. *Journal of Community Psychology,* 32, 61-79. doi: 10.1002/jcop.10078

Barreto, M. & Ellemers. N. (2005). The perils of political correctness: Men's and women's responses to old-fashioned and modern sexist views. *Social Psychology Quarterly, 68,* 75-88. doi: 10.1177/019027250506800106

Bennett, S., Banyard, V. L., & Garnhart, L. (2014). To act or not to act, That is the question? Barriers and facilitators of bystander intervention. *Journal of Interpersonal Violence, 29,* 476-496. doi: 10.1177/0886260513505210

Black, B. M., Tolman, R. M., Callahan, M., Saunders, D. G., & Weisz, A. N. (2008). When will adolescents tell someone about dating violence victimization? *Violence Against Women, 14,* 741-758. doi: 10.1177/1077801208320248

Bogle, K. (2007). The shift from dating to hooking up in college: What scholars have missed. *Sociology Compass, 1/2,* 775-788. doi: 10.1111/j.1751-9020.2007.00031.x

Bogle, K. (2008). *Hooking Up: Sex, Dating, and Relationships on Campus.* New York: New York University Press.

Burn, S. M. (2009). A situational model of sexual assault prevention through bystander intervention. *Sex Roles, 60,* 779-792. doi: 10.1007/s11199-008-9581-5

Casey, E. A. & Ohler, K. (2012). Being a positive bystander: Male antiviolence allies' experiences of 'Stepping Up'. *Journal of Interpersonal Violence, 27,* 62-83. doi: 10.1177/0886260511416479

Cornelius, T. L. & Resseguie, N. (2007). Primary and secondary prevention programs for dating violence: A review of the literature. *Aggression and Violent Behavior, 12,* 364-375. http://dx.doi.org/10.1016/j.avb.2006.09.006

Cui, M., Ueno, K., Gordon, M., & Fincham, F. D. (2013). The continuation of intimate partner violence from adolescence to young adulthood. *Journal of Marriage and Family, 75,* 300-313. doi: 10.1111/jomf.12016

Darley, J. M. & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 4,* 377-383. http://psycnet.apa.org/doi/10.1037/h0025589

DeGue, S., Valle, L. A., Holt, M. K., Massetti, G. M., Matjasko, J. L., & Tharp, A. T. (2014). A systematic review of primary prevention strategies for sexual violence perpetration. *Aggression and Violent Behavior, 19,* 346-362. http://dx.doi.org/10.1016/j.avb.2014.05.004

De Koker, P., Mathews, C., Zuch, M., Bastien, S., & Mason-Jones, A. J. (2014). A systematic review of interventions for preventing adolescent intimate partner violence. *Journal of Adolescent Health, 54,* 3-13. http://dx.doi.org/10.1016/j.jadohealth.2013.08.008

De La Rue, L., Polanin, J. R., Espelage, D. L., & Pigott, T. D. (2014). School-based interventions to reduce dating and sexual violence: A systematic review. *Campbell Systematic Reviews,* 10(7).

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel- plot–based method of testing and adjusting for publication bias in meta- analysis. *Biometrics, 56*(2), 455-463. doi: 10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G.D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal 315*, 629-634. doi: http://dx.doi.org/10.1136/bmj.315.7109.629

Exner, D. & Cummings, N. (2011). Implications for sexual assault prevention: College students as prosocial bystanders. *Journal of American College Health, 59,* 655-657. doi: 10.1080/07448481.2010.515633

Fellmeth, G. L. T., Heffernan, C., Nurse, J., Habibula, S., & Sethi, D. (2013). Educational and skills-based interventions for preventing relationahip and dating violence in adolescents and young adults (Review). *Cochrane Database of Systematic Reviews, Issue 6. Art. No.: CD004534.* doi: 10.1002/14651858.CD004534.pub3

Foshee, V. A. (1996). Gender differences in adolescent dating abuse prevalence, types, and injuries. *Health Education Research, 11,* 275-286. doi: 10.1093/her/11.3.275

Foshee, V. A., Benefield, T. S., Ennett, S. T., Bauman, K. E., & Suchindran, C. (2004). Longitudinal predictors of serious physical and sexual dating violence victimization during adolescence. *Preventive Medicine, 39,* 1007-1016. http://dx.doi.org/10.1016/j.ypmed.2004.04.014

Foshee, V. A., Linder, F., MacDougall, J. E., & Bangdiwala, S. (2001). Gender differences in the longitudinal predictors of adolescent dating violence. *Preventive Medicine, 32,* 128-141. http://dx.doi.org/10.1006/pmed.2000.0793

Foshee, V. A., Reyes, H. L. M., & Ennett, S. T. (2010). Examination of sex and race differences in longitudinal predictors of the initiation of adolescent dating violence perpetration. *Journal of Aggression, Maltreatment, and Trauma, 19,* 492-516. doi: 10.1080/10926771.2010.495032

Giordano, P. C. (2003). Relationships in adolescence. *Annual Review of Sociology, 29,* 257-281. doi: 10.1146/annurev.soc.29.010202.100047

Gressard, L. A., Swahn, M. H., & Tharp, A. T. (2015). A first look at gender inequality as a societal risk factor for dating violence. *American Journal of Preventive Medicine, 49,* 448-457. http://dx.doi.org/10.1016/j.amepre.2015.05.017

Halpern, C. T., Spriggs, A. L., Martin, S. L., & Kupper, L. L. (2009). Patterns of intimate partner violence victimization from adolescence to young adulthood in a nationally representative sample. *Journal of Adolescent Health, 45,* 508-516. http://dx.doi.org/10.1016/j.jadohealth.2009.03.011

Harned, M. S. (2001). Abused women or abused men? An examination of the context and outcomes of dating violence. *Violence and Victims, 16,* 269-285.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1,* 39-65. doi: 10.1002/jrsm.5

Higgins J. P. T., & Green, S. (Eds) (2011). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available from www.cochrane-handbook.org

Howard, D. E., Wang, M. Q., Yan, F. (2007). Prevalence and psychosocial correlates of forced sexual intercourse among U.S. high school adolescents. *Adolescence, 42,* 629-643.

Katz, J. [Jackson] (1995). Reconstructing masculinity in the locker room: The Mentors in Violence Prevention project. *Harvard Educational Review, 65,* 163-174. http://dx.doi.org/10.17763/haer.65.2.55533188520136u1

Katz, J. [Jackson]. Heisterkamp, H. A., & Fleming, W. M. (2011). The social justice roots of the Mentors in Violence prevention model and its application in a high school setting. *Violence Against Women, 17,* 684-702. doi: 10.1177/1077801211409725

Katz, J. [Jennifer]. (2015). Effects of group status and victim sex on male bystanders' responses to a potential party rape. *Journal of Aggression, Maltreatment, & Trauma, 24,* 588-602. doi:10.1080/10926771.2015.1029184

Katz, J. [Jennifer], Colbert, S., & Colangelo, L. (2015). Effects of group status and victim sex on female bystanders' responses to a potential party rape. *Violence and Victims, 30,* 265-278. http://dx.doi.org/10.1891/0886-6708.VV-D-13-00099

Katz, J. [Jennifer] & Moore, J. (2013). Bystander education training for campus sexual assault prevention: An initial meta-analysis. *Violence and Victims, 28,* 1054-1067. http://dx.doi.org/10.1891/0886-6708.VV-D-12-00113

Kettrey, H. H. (2016). What's gender got to do with it? Sexual double standards and power in heterosexual college hookups. *The Journal of Sex Research, 53,* 754-765. doi: 10.1080/00224499.2016.1145181

Krebs, C. P., Lindquist, C. H., Warner, T. D., Fisher, B. S., & Martin, S. L. (2009). College women's experiences with physically forced, alcohol- or other drug-enabled, and drug-facilitated sexual assault before and since entering college. *Journal of American College Health, 57,* 639-649. doi:10.3200/JACH.57.6.639-649

Latane, B. & Darley, J. N. (1969). Bystander 'apathy.' *American Scientist, 57,* 244-268. http://www.jstor.org/stable/27828530

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analyses.* Thousand Oaks, CA: Sage Publications, Inc.

McCauley, H. L., Tancredi, D. J., Silverman, J. G., Decker, M. R., Austin, S. B., McCormick, M. C., Virata, M. C., & Miller, E. (2013). Gender-equitable attitudes, bystander behavior, and recent abuse perpetration against heterosexual dating partners of male high school athletes. *American Journal of Public Health, 103,* 1882-1887.

McMahon, S. (2010). Rape myth beliefs and bystander attitudes among incoming college students. *Journal of American College Health, 59,* 3-11. doi: 10.1080/07448481.2010.483715

McMahon S. & Banyard, V. L. (2012). When can I help? A conceptual framework for the prevention of sexual violence through bystander intervention. *Trauma, Violence, & Abuse, 13,* 3-14. doi: 10.1177/1524838011426015

Messner, M. A. (2015). Bad men, good men, bystanders: Who is the rapist? *Gender & Society.* Advance online publication. doi: 10.1177/0891243215608781

Miller, E., Tancredi, D. J., McCauley, H. J., Decker, M. R., Virata, M. C. D., Anderson, H. A., . . . & Silverman, J. G. (2012). "Coaching boys into men": A cluster randomized controlled trial of a dating violence prevention program. *Journal of Adolescent Health, 51,* 431-438. doi: http://dx.doi.org/10.1016/j.jadohealth.2012.01.018

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1. http://doi.org/10.1186/2046-4053-4-1

Molidor, C. & Tolman, R. M. (1998). Gender and contextual factors in adolescent dating violence. *Violence Against Women, 4,* 180-194. doi: 10.1177/1077801298004002004

Noonan, R. K. & Charles, D. (2009). Developing teen dating violence prevention strategies: Formative research with middle school youth. *Violence Against Women, 15,* 1087-1105. doi: 10.1177/1077801209340761

Palmer, T., Peters, J., Sutton, A.J., & Moreno, S. (2008). Contour enhanced funnel plots for meta-analysis. *The Stata Journal, 8*, 242-254. doi: http://dx.doi.org/10.1016/j.jclinepi.2007.11.010

Sterne, J. A. C, Higgins, J. P. T., & Reeves, B. C. on behalf of the development group for ROBINS-I. (2016, March). A tool for assessing Risk Of Bias In Non-randomized Studies of Interventions, Version 7. Available from http://www.riskofbias.info [accessed July 29, 2016].

Storer, H. L., Casey, E., & Herrenkohl, T. (2015). Efficacy of bystander programs to prevent dating abuse among youth and young adults: A review of the literature. *Trauma, Violence, & Abuse.* Advance online publication. doi: 10.1177/1524838015584361

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology, 68,* 199-214. http://psycnet.apa.org/doi/10.1037/0022-3514.68.2.199

Tanner-Smith, E. E., & Polanin, J. R. (2015). A Retrospective Analysis of Dissemination Biases in the Brief Alcohol Intervention Literature. *Psychology of Addictive Behaviors, 29*(1), 49–62. http://doi.org/10.1037/adb0000014

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods, 5,* 13-30. doi: 10.1002/jrsm.1091

Tipton, E. (2013). Robust variance estimation in meta-regression with binary dependent effects. *Research Synthesis Methods, 4,* 169-187. doi: 10.1002/jrsm.1070

Weisz, A. N., Tolman, R. M., Callahan, M. R., Saunders, D. G., & Black, B. M. (2007). Informal helpers' responses when adolescents tell them about dating violence or romantic relationship problems. *Journal of Adolescence, 30,* 853-868. http://dx.doi.org/10.1016/j.adolescence.2006.09.004

REVIEW AUTHORS

Lead review author:

The lead author is the person who develops and co-ordinates the review team, discusses and assigns roles for individual members of the review team, liaises with the editorial base and takes responsibility for the on-going updates of the review.

Name:	Heather Hensman Kettrey, PhD
Title:	Research Associate
Affiliation:	Peabody Research Institute, Vanderbilt University
Address:	230 Appleton Place, BOX 0181 GPC
City, State, Province or County:	Nashville, TN
Postal Code:	37203-5721
Country:	USA
Phone:	615-875-1094
Mobile:	
Email:	Heather.h.kettrey@vanderbilt.edu

Co-author(s): (There should be at least one co-author)

Name:	Emily Tanner-Smith, PhD
Title:	Associate Research Professor
Affiliation:	Peabody Research Institute, Vanderbilt University
Address:	230 Appleton Place, BOX 0181 GPC
City, State, Province or County:	Nashville, TN
Postal Code:	37203-5721
Country:	USA
Phone:	615-322-6304
Mobile:	
Email:	e.tanner-smith@vanderbilt.edu

ROLES AND RESPONSIBLIITIES

Members of the research team responsible for core areas of the review are as follows:

- Content: Heather Hensman Kettrey

- Systematic review methods: Heather Hensman Kettrey, Emily Tanner-Smith

- Statistical analysis: Heather Hensman Kettrey, Emily Tanner-Smith

- Information retrieval: Heather Hensman Kettrey

The research team at Vanderbilt University's Peabody Research Institute (PRI) that will undertake this project includes the lead report author, co-author, and a research assistant. Dr. Kettrey, the lead review author, will coordinate the review team and assume responsibility for the implementation of the project throughout its duration. Specific tasks include compiling the sample of research reports, creating the database, coding studies, coding program training materials (i.e., program content variables), analyzing data, and preparing the Campbell Review. Dr. Tanner-Smith, review co-author, will provide methodological guidance and mentorship to Dr. Kettrey throughout all phases of data collection and analysis (i.e., sampling, database construction, coding, data analysis, and preparing the Campbell Review).

The bibliographic information found in a thorough search for meta-analysis requires careful screening, first of abstracts, then retrieval and further screening of candidate full-text reports. The studies found eligible for the meta-analysis must then be carefully read and coded into an electronic database. To manage this labor-intensive portion of the work, one research assistant (TBD) will devote time to the project. This person will be selected from among the staff members of PRI who are familiar with meta-analysis, thus, minimizing the necessary training effort.

SOURCES OF SUPPORT

This review is supported by a grant from the Campbell Collaboration (CSR1.60).

DECLARATIONS OF INTEREST

The authors have no conflicts of interest to report.

PRELIMINARY TIMEFRAME

July 2016 (Completed)

- Submitted Title Registration to Campbell Collaboration (Draft submitted with grant application, 3/29/16; Revision submitted 7/1/16)

August 2016 (Completed)

- Submitted Protocol to Campbell Collaboration (Draft submitted 8/29/16)

September –October 2016 (Completed)

- Drafted Coding Manual

- Constructed Data Management System

November 2016 (Completed)

- Conducted Literature Search

- Screened Abstracts

December 2016 – January 2017 (Completed)

- Retrieved Literature

- Screened Candidate Studies for Eligibility

- Began Data Extraction and Coding

February 2017 – March 2017 (Completed)

- Revised Protocol and Coding Manual per Campbell Reviewers' Comments (Received reviews 2/21/17)

- Submitted Revised Protocol to Campbell Collaboration (Submitted on 3/31/17)

April 2017 – June 2017

- Update Literature Search and Screening per Revised Protocol

- Complete Data Extraction and Coding (Includes re-coding previously completed studies per revised protocol)

July 2017

- Data Analysis

August 2017

- Prepare Final Review

September 2017

- Submit Completed Review to Campbell Collaboration by 9/29/17 per Campbell
  Collaboration Grant CSR1.60.

October 2017, or as needed

- Prepare and Submit Review Revisions to Campbell Collaboration

PLANS FOR UPDATING THE REVIEW

Heather Hensman Kettrey anticipates updating the review at least once every five years, pending continual accrual of sufficient research on the topic. If she is unable to assume this responsibility she will transfer responsibility to another qualified researcher – with approval from the Campbell Collaboration.

AUTHOR DECLARATION

Authors' responsibilities

By completing this form, you accept responsibility for preparing, maintaining and updating the review in accordance with Campbell Collaboration policy. The Campbell Collaboration will provide as much support as possible to assist with the preparation of the review.

A draft review must be submitted to the relevant Coordinating Group within two years of protocol publication. If drafts are not submitted before the agreed deadlines, or if we are unable to contact you for an extended period, the relevant Coordinating Group has the right to de-register the title or transfer the title to alternative authors. The Coordinating Group also has the right to de-register or transfer the title if it does not meet the standards of the Coordinating Group and/or the Campbell Collaboration.

You accept responsibility for maintaining the review in light of new evidence, comments and criticisms, and other developments, and updating the review at least once every five years, or, if requested, transferring responsibility for maintaining the review to others as agreed with the Coordinating Group.

Publication in the Campbell Library

The support of the Coordinating Group in preparing your review is conditional upon your agreement to publish the protocol, finished review, and subsequent updates in the Campbell

Library. The Campbell Collaboration places no restrictions on publication of the findings of a Campbell systematic review in a more abbreviated form as a journal article either before or after the publication of the monograph version in *Campbell Systematic Reviews*. Some journals, however, have restrictions that preclude publication of findings that have been, or will be, reported elsewhere and authors considering publication in such a journal should be aware of possible conflict with publication of the monograph version in *Campbell Systematic Reviews*. Publication in a journal after publication or in press status in *Campbell Systematic Reviews* should acknowledge the Campbell version and include a citation to it. Note that systematic reviews published in *Campbell Systematic Reviews* and co-registered with the Cochrane Collaboration may have additional requirements or restrictions for co-publication. Review authors accept responsibility for meeting any co-publication requirements.

I understand the commitment required to undertake a Campbell review, and agree to publish in the Campbell Library. Signed on behalf of the authors:

Form completed by: Heather Hensman Kettrey   **Date:** 31 March 2017

BYSTANDER INTERVENTION META-ANALYSIS CODING MANUAL

Last Updated March 31, 2017

TABLE OF CONTENTS

ELIGIBILITY CRITERIA

INTERVENTION

1. Eligible intervention programs are those that approach participants as allies in preventing and/or alleviating intimate partner violence (IPV) and/or sexual assault among adolescents and/or college students. (NOTE: For the purposes of this meta-analysis, IPV includes physical or sexual violence perpetrated by an intimate partner. Sexual assault refers to sexual violence that may be perpetrated by an intimate partner, acquaintance, or stranger). Some part of the program must focus on ways that cultivate a willingness for a person to respond to others who are at risk for IPV or sexual assault. All delivery formats are eligible for inclusion (e.g., in-person training sessions, video programs, web-based training, advertising/poster campaigns). There are no treatment duration criteria for inclusion. The following interventions are ineligible:

    a. Studies that report bystander outcomes but do not meet the aforementioned intervention criteria are not eligible for inclusion. For example, studies that assess bystander outcomes of general IPV/sexual assault prevention programs that contain NO bystander content.
    b. Studies that assess outcomes of programs that aim to facilitate pro-social bystander behavior, but that do not explicitly include a component addressing IPV and/or sexual assault (e.g., programs that use a bystander approach to prevent bullying) are not eligible.

COMPARISON

2. Eligible comparison groups must receive no intervention services targeting bystander attitudes/behavior, IPV, or sexual assault. Thus, treatment-treatment studies that compare individuals assigned to a bystander program with individuals assigned to a general IPV or sexual assault prevention program are not eligible for inclusion. However, eligible comparison groups may receive a sham or attention treatment that is expected to have no effect on bystander outcomes or attitudes/behaviors regarding IPV or sexual assault (e.g., nutrition or health education).

PARTICIPANTS AND SETTING

3. The intervention must be implemented with adolescents and/or college students in educational settings. Eligible participants include adolescents enrolled in grades 7 through 12 and college students enrolled in any type of undergraduate postsecondary educational institution. Studies that assess bystander programs implemented with adolescents and young adults outside of educational institutions (e.g., in community settings, military settings) are ineligible for inclusion in the review.

    a. Eligible studies include those that report on general samples of adolescents and/or college students as well as studies using specialized samples such as those primarily consisting of college athletes, fraternity/sorority members, and single-sex samples.
    b. Study samples primarily consisting of post-graduate students will be ineligible for inclusion; the mean age of samples may be no greater then 25 to be included in the review.

RESEARCH DESIGN

4. Eligible studies must use an experimental or controlled quasi-experimental research design to compare an intervention group (e.g., students assigned to complete a bystander program) with a comparison group (e.g., students not assigned to complete a bystander program). The following are eligible designs:

   a. Randomized controlled trials: Studies in which individuals, classrooms, schools, or other groups are randomly assigned to intervention and comparison conditions.

   b. Quasi-randomized controlled trials: Studies where assignment to conditions is quasi-random, for example, by birth date, day of week, student identification number, month, or some other alternation method.

   c. Controlled quasi-experimental designs: Studies where participants are not assigned to conditions randomly or quasi-randomly (e.g., participants self-select into groups). Given the potential selection biases inherent in these controlled quasi-experimental designs, we will only include those that also meet one of the following criteria:

      i. Regression discontinuity designs: Studies that use a cutoff on a forcing variable to assign participants to intervention and comparison groups, and assess program impacts around the cutoff of the forcing variable.

      ii. Studies that use propensity score or other matching procedures to create a matched sample of participants in the intervention and comparison groups. To be eligible for inclusion, these studies must also provide enough statistical information to permit estimation of pretest effect sizes for the matched groups (NOTE: Pretest effect sizes refer to pre-intervention measures of an outcome variable. Pre-intervention measures that are not reported as outcomes are group equivalence measures).

      iii. For studies where participants in the intervention and comparison groups are not matched, enough statistical information must be reported that will permit us to estimate pretest effect sizes for at least one outcome measure. Pretest must be administered to the treatment and comparison group at the same point in time.

SAMPLE SIZES

5. Intervention and control groups must each contain at least 10 individuals at the time of assignment to study conditions.

OUTCOME MEASURES

6. Eligible studies must report the effects of bystander programs on at least one of the following primary outcomes:

   a. General attitudes toward IPV and/or sexual assault and victims (e.g., victim empathy, rape myth acceptance).

   b. Self-efficacy with regard to bystander intervention (e.g., measures of confidence in respondents' ability to intervene).

   c. Intentions to intervene when witnessing instances or warning signs of IPV and/or sexual assault.

   d. Actual intervention behavior when witnessing instances or warning signs of IPV and/or sexual assault.

   e. Perpetration of IPV and/or sexual assault.

NOTES: (1) Studies that report outcomes by breakout groups (e.g., report findings separately for males/females, fraternity/nonfraternity, etc.) are not eligible unless findings are reported such that outcomes for breakout groups can be combined. (2) Studies must assess at least one

outcome pertinent to sexual assault to be included in the main Campbell review and meta-analysis. Studies that only assess physical violence outcomes (e.g., evaluations of IPV bystander programs that measure physical violence outcomes, but not sexual violence outcomes) will be coded and set aside for separate analysis. These must be marked as **"Physical Violence Only"** on the eligibility screen in FileMaker. There are no other restrictions on eligibility. Studies that meet all eligibility criteria but do not provide sufficient information for calculating effect sizes are still eligible for inclusion and will be coded (i.e., the study will be coded on all items except the effect sizes).

STUDY LEVEL

[studyid]
Study identification number. The "unit" you will code here consists of a study, i.e., one research investigation of a defined subject sample or subsamples compared to each other, and the treatments, measures, and statistical analyses applied to them. Sometimes there are several different reports about a single study. In such cases, the coding should be done from the full set of relevant reports, using whichever report is best for each item to be coded; BE SURE YOU HAVE THE FULL SET OF RELEVANT REPORTS BEFORE BEGINNING TO CODE. Sometimes a single report describes more than one study sample, e.g., evaluations at three separate sites. In these cases, each study sample will have a unique study identification number and each study should be coded separately as if it had been described in a separate report.

Each study has its own study identification number, or StudyID (e.g., 619). Each report also has an identification number (e.g., 619.01), which you will find in the FileMaker bibliography. The ReportID has two parts; the part before the decimal is the StudyID, and the part after the decimal is used to distinguish the reports within a study. (These two types of ID numbers, along with bibliographic information, are assigned and tracked using the bibliography.) When coding, use the study ID (e.g., 619) to refer to the study as a whole, and use the appropriate report ID (e.g., 619.01) when referring to an individual report.

[coder]
Coder's initials

HEADER VARIABLES
[pub.year]
Year of publication for the primary report (four digits). If more than one report exists, choose the date for the report that provides the effect sizes. If effect sizes come from more than one report, choose the earliest date.

[publication]
Type of publication for the primary report. If you are using more than one type of publication to code your study, choose the publication that supplies the effect sizes (in cases where more than one report provides effect sizes, choose a "peer reviewed" choice over another option, or choose the report that provides the most effect sizes).
1. Journal Article
2. Book or book chapter
3. Dissertation
4. MA/MS Thesis
5. Private report
6. Government report
7. Conference paper
8. Other
9. Unclear

[peer]
Is this a peer-reviewed publication (primary report only)?
1. Yes

2. No
3. Unclear

[funded]
Do the authors declare any funding?
1. Yes
2. No
3. Unclear

[researcher_role]
Did one or more of the researchers assessing outcomes of the intervention program (i.e., authors of the research report) play a significant role in developing the program?
1. Yes – One or more researchers developed the intervention program.
2. No – Researchers are independent evaluators.
3. Unclear/Cannot tell.

The following are the developers of some potentially eligible bystander (or peer/social network-based) programs. Some of the programs below were developed by individuals (e.g., academics) and others were developed by organizations. In the latter case, any research report authored by a representative of the developer/organization should be coded as "1." Less popular programs that do not appear on this list may require a careful reading of the text (e.g., the authors will likely cite the program in the evaluation report) and/or some additional research.

- **Bringing in the Bystander:** Developed by Victoria Banyard, Mary Moynihan, Elizabeth Plante (University of New Hampshire).
- **Circle of 6 (potentially ineligible):** Developed by Nancy Schwartzman, Thomas Cabus. See list of advisory board members (http://www.circleof6app.com/about/#member2).
- **Coaching Boys into Men:** Developed by Futures without Violence. See list of board and staff members (https://www.futureswithoutviolence.org/about-us/board-and-staff/).
- **Green Dot:** Developed by Dorthy Edwards.
- **Hollaback (potentially ineligible):** Developed by Emily May.
- **InterACT:** Marc Rich may or may not be the developer, but he has a conflict of interest, as he directs this program/troupe.
- **Know Your Power:** This is the campaign companion to Bringing in the Bystander. Developed by Sharyn Potter, Jane Stapleton, Mary Moynihan (University of New Hampshire).
- **Men Can Stop Rape:** Developed by John Stoltenberg. See MCSR staff members (http://www.mencanstoprape.org/Our-Staff/) and board members (http://www.mencanstoprape.org/Our-Board/).
- **Mentors in Violence Prevention (MVP):** Developed by Jackson Katz.
- **Red Flag Campaign:** Developed by Virginia Sexual and Domestic Violence Action Alliance. See list of governing body members (http://www.vsdvalliance.org/#/aboutcontact-us/governing-body).
- **Speak Up Speak Out (potentially ineligible):** Developed by Maryland Coalition Against Sexual Assault.
- **Step Up!:** Developed by Becky Bell.
- **That's Not Cool:** Developed by Futures without Violence. See list of board and staff members (https://www.futureswithoutviolence.org/about-us/board-and-staff/).
- **The Men's Program (How to Help a Sexual Assault Survivor):** Developed by John Foubert (Oklahoma State University).
- **The Women's Program**: Developed by John Foubert (Oklahoma State University).

- **Where Do You Stand:** This is an ad campaign developed by Men Can Stop Rape. MCSR was developed by John Stoltenberg. See MCSR staff members (http://www.mencanstoprape.org/Our-Staff/) and board members (http://www.mencanstoprape.org/Our-Board/).
- **White Ribbon Campaign:** Developed by Michael Kaufman, Jack Layton.

GROUP IDENTIFICATION AND SELECTION
At this stage, you will need to identify the groups in the study for which effect size statistics can be computed. Note that for any group comparison coding, the two groups involved must be from the same experiment or quasi-experiment; that is, they must have been involved in the same randomization, matching, etc. from the same design. If two or more experiments or quasi-experiments are presented in the same report, each must be handled separately.

[txa-d]   Write in Name
Treatment Groups   1-4 _____

[cta-d]   Write in Name
Comparison Groups   1-4 _____

STUDY CONTEXT
[country]
Country in which the study was conducted:
1. United States
2. Canada
3. Mexico
4. Australia
5. European Nation
6. Other

[state]
State in which the study was conducted.
1. Alabama
2. ....
3. ...
51. District of Columbia
52. Single state (unspecified)
53. Multiple states (unspecified)
54. NA – Conducted outside US

[study.year]
Year(s) when study was conducted. Code NA if cannot tell.

[ed.setting]
Type of educational setting in which the study was conducted:
1. Secondary school (grades 7-12)
2. College or university
3. Unclear

STUDY DESIGN AND METHODS
[design]

Method of assignment to groups. This item focuses on the initial method of assignment to groups regardless of subsequent degradations due to attrition, refusal, etc. prior to treatment onset. These latter situations are coded elsewhere.

Random or near-random:
1. Random assignment at the individual level: Individual participants are randomly assigned to conditions. In some cases random assignment may be done after individuals have been matched or blocked.
2. Random assignment by group: Intact groups such as classrooms are assigned.
3. Regression discontinuity design: Individuals are assigned to groups based on a cut-off score of a forcing variable (i.e., pretest or risk factor measure). Note that this design will be rare.
4. Quasi-randomized procedure presumed to produce comparable groups: This applies to groups which have individuals assigned by some naturally occurring process that is apparently random, e.g. alternation, date of birth, medical record number. The key here is that the procedure used to select groups is not strictly random, but the method of allocation should not create nonequivalence between groups.

Non-random, but matched or statistically controlled:
Note: Matching refers to the process by which individuals are selected for conditions (e.g., treatment and comparison) in a manner that ensures that the individuals in one group are matched on certain relevant characteristics in the other group. Comparing the characteristics of the groups AFTER they have been assigned to experimental conditions does NOT constitute matching.
5. Matched individually, through sampling, on one or more pretest measures (i.e., attitudes toward IPV/sexual assault or victims, bystander self-efficacy, bystander intentions, bystander behavior, IPV/sexual assault perpetration).
6. Participants in the intervention and comparison groups are not matched, but enough statistical information is provided to permit the estimation of pretest effect sizes for at least one outcome measure (i.e., attitudes toward IPV/sexual assault or victims, bystander self-efficacy, bystander intentions, bystander behavior, IPV/sexual assault perpetration).
7. Matched at a larger group level; that is, intact groups were matched on their means for one or more pretest measures (i.e., attitudes toward IPV/sexual assault or victims, bystander self-efficacy, bystander intentions, bystander behavior, IPV/sexual assault perpetration). In these cases the mean pretest measures of the groups are similar, but each subject in one group has not been individually matched to a subject in the other group on age.

[matched.var.list]
Please list all of the variables used in the matching and/or statistical controls.

[m]
For cluster randomized trials, please enter the average cluster size (i.e., average number of participants in each cluster). Code -9 for cannot tell. Code -8 for Not applicable.

[attrf_o]
What is the overall attrition rate (across all groups) in the study between the time of assignment to groups to the first follow-up? This item refers to overall attrition in the study; more detailed attrition calculations will be estimated using the assigned and observed sample sizes coded in the effect size section.

[attrl_o]

What is the overall attrition rate (across all groups) in the study between the time of assignment to groups to the last follow-up? Again, this item refers to overall attrition in the study; more detailed attrition calculations will be estimated using the assigned and observed sample sizes coded in the effect size section.

[itt]

Did the authors use an intent-to-treat (ITT) analysis? Intent-to-treat analysis refers to situations where researchers 'analyze as randomized', meaning that all individuals who were randomized to the intervention/control groups are included in the final outcome analysis, regardless of whether they actually attended/completed the intervention. Note, that it is possible for a study to conduct an ITT analysis even if they have attrition, as long as they had intended to include any non-completers in their final model.

1. Yes – explicit ITT analysis
2. No – completer analysis, TOT analysis
3. Cannot tell

[missdata]

How did the authors handle missing data in their analysis?

1. Listwise deletion
2. Pairwise deletion
3. Mean or mode imputation
4. Single regression imputation
5. Dummy variable approach (imputed value at zero with dummy variable)
6. Multiple imputation
7. Full information maximum likelihood (FIML)
8. Other method
9. Not applicable – no missing data
10. Cannot tell

RISK OF BIAS: NON-EXPERIMENTAL STUDIES

Although these risk of bias variables will largely be used in analyses of <u>non-experimental studies</u>, please code them <u>for all eligible studies</u>, regardless of research design.

[rob_ne_confounding]

Describe any risk of bias for non-experimental studies based on pre-specified confounding factors (i.e., gender, fraternity/sorority membership, athletic team membership, pre-intervention attitudes, pre-intervention bystander measures, and prior sexual assault victimization). That is, please note if a considerable percentage of participants exhibited any of these confounding variables (e.g., if the study was implemented with a sample of fraternity/sorority members).

[rob_ne_cointerventions]

Describe any risk of bias for non-experimental studies based on pre-specified co-interventions (i.e, general sexual assault prevention programs, IPV prevention programs, and general bystander programs). That is, please note if there is reason to believe participants had previous or concurrent exposure to any potentially confounding co-interventions.

RISK OF BIAS: EXPERIMENTAL STUDIES

Although these risk of bias variables will largely be used in analyses of <u>experimental studies</u>, please code them <u>for all eligible studies</u>, regardless of research design.

[rob_sg]

Random Sequence Generation: What is the risk of *sequence generation* due to inadequate generation of a randomized sequence?

1. Low risk. The investigators describe a random component in the sequences generation process such as;
   a. Referring to a random number table;
   b. Using a computer random number generator;
   c. Coin tossing;
   d. Shuffling cards or envelopes;
   e. Throwing dice;
   f. Drawing of lots;
   g. Minimization (may be implemented without a random element, and this is considered equivalent to being random).
2. High risk. The investigators describe a non-random component in the sequence generation process. Usually, the description would involve some systematic non-random approach, for example:
   a. Sequence generated by odd/even birth dates;
   b. Sequence generated by some rule based on date (or day) of admission;
   c. Sequence generated by some rule based on some sort of (e.g., hospital/clinic) record number.

   Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgment or some method of non-random categorization of participants, for example:
   a. Allocation by judgment of the clinician (or teacher or practitioner);
   b. Allocation by preference of the participant;
   c. Allocation based on results of a laboratory test or series of tests;
   d. Allocation by availability of the intervention.

   By definition, any quasi-experimental design where participants self-select into conditions is at high risk of bias.

3. Unclear risk of bias. Insufficient information is provided about the sequence generation process to permit judgement of low or high risk.

[rob_sg_text]
Provide a description of the information used to code the risk of bias due to sequence generation.

[rob_allocation]
Allocation Concealment: What is the risk of bias due to ***allocation concealment*** (inadequate concealment of allocations prior to assignment)? Note that most studies will not report information on allocation concealment.
1. Low risk. Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:
    a. Central allocation (including telephone, web-based, and pharmacy-controlled randomization);
    b. Sequentially numbered drug containers of identical appearance;
    c. Sequentially numbered, opaque, sealed envelopes
2. High risk. Participants or investigators enrolling participants could possibly foresee assignments and thus introduce bias, such as allocation based on:
    a. Using an open random allocation schedule (e.g., a list of random numbers);
    b. Assignment envelopes were used without appropriate safeguards (i.e., if envelopes were unsealed or non-opaque or not sequentially numbered);
    c. Alternation or rotation;
    d. Date of birth;
    e. Case record number;
    f. Any other explicitly unconcealed procedure.
3. Unclear risk. Insufficient information to permit judgment of 'low risk' or 'high risk.' This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque or sealed.

[rob_allocation_text]
Provide a description of the information used to code the risk of bias due to allocation concealment.

[rob_blinding_part]
Blinding of Participants and Personnel: What is the risk of bias due to ***knowledge*** of the allocated interventions ***by participants and personnel***?
1. Low risk. Any one of the following:
    a. No blinding or incomplete blinding, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding;
    b. Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken.
2. High risk. Any one of the following:
    a. No blinding or incomplete blinding, and the outcome measurement is likely to be influenced by lack of blinding;
    b. Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome measurement is likely to be influenced by lack of blinding.
3. Unclear risk. Insufficient information to permit judgment of 'low risk' or 'high risk.'

[rob_blinding_part_text]

Provide a description of the information used to code the risk of bias due to insufficient blinding of participants and personnel.

[rob_blinding_assess]
Blinding of Outcome Assessment: What is the risk of bias due to ***knowledge*** of the allocated interventions ***by outcome assessors***?
1. Low risk. Any one of the following:
    a. No blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding;
    b. Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken
2. High risk. Any one of the following:
    a. No blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding;
    b. Blinding of outcome assessment, but likely that the blinding could have been broken, and the outcome measurement is likely to be influenced by lack of blinding.
    c. All outcomes are self-reported outcomes (which inherently are NOT blinded)
3. Unclear risk. Insufficient information to permit judgment of 'low risk' or 'high risk.'

[rob_blinding_assess_text]
Provide a description of the information used to code the risk of bias due to insufficient blinding of outcome assessment.

[rob_incomplete]
Incomplete Outcome Data: What is the risk of attrition bias due to amount, nature, or handling of ***incomplete outcome data***?
1. Low risk. Any one of the following:
    a. No missing outcome data;
    b. Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias);
    c. Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups;
    d. For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate;
    e. For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size;
    f. Missing data have been imputed using appropriate methods.
2. High risk. Any one of the following:
    a. Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups;
    b. For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate;
    c. For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size;
    d. "As-treated" analysis done with substantial departure of the intervention received from that assigned at randomization;
    e. Potentially inappropriate application of single imputation.
3. Unclear risk. Insufficient information to permit judgment of 'low risk' or 'high risk.'

[rob_incomplete_text]
Provide a description of the information used to code the risk of bias due to incomplete outcome data.

[rob_report]
Selective Reporting: What is the risk of bias due to ***selective outcome reporting***?
1. Low Risk of Bias. Any one of the following:
    a. The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way;
    b. The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon)
2. High Risk of Bias. Any one of the following:
    a. Not all of the study's pre-specified primary outcomes have been reported;
    b. One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified;
    c. One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect);
    d. One or more outcomes of interest in the review are reported incompletely so that they cannot be entered into a meta-analysis;
    e. The study report fails to include results for a key outcome that would be expected to have been reported for such a study.
3. Unclear/Cannot Tell. Insufficient information to permit judgment of 'low risk' or 'high risk.'

[rob_report_text]
Provide a description of the information used to code the risk of bias due to selective reporting.

[rob_other]
Other Bias: What is the risk of bias due to ***problems not covered elsewhere***?
1. Low risk. The study appears to be free of other sources of bias.
2. High risk. There is at least one important risk of bias. For example, the study:
    a. Had a potential source of bias related to the specific study designed used; or
    b. Has been claimed to be fraudulent; or
    c. Had some other problem.
3. Unclear/cannot tell. Insufficient information to assess whether an important risk of bias exists.

[rob_other_text]
Provide a description of the information used to code the risk of bias due to problems not covered elsewhere.

Create one record in this database for each of the intervention and/or comparison groups you selected earlier for coding. For example, studies with a single intervention group and a single comparison group will have two records in this section of the database.

GROUP IDENTIFICATION
[groupid]
Number each group consecutively within a study, starting with 1.

[tvc]
Select the type of group you are coding.
1. Intervention group
2. Control/comparison group

[type]
What type of services does this group receive?
1. Focal/primary bystander intervention program. There may be several focal programs in a study, as when two different types of programs are compared, both of which are expected to be effective.
2. Active treatment that is not a bystander program or IPV/sexual violence prevention program. This is a group that receives a sham treatment (e.g., watches a video on nature, receives nutrition information, diet intervention) intended to take the same duration as the focal intervention program, but does not involve any active bystander or IPV/sexual violence prevention/education components.
3. Inactive comparison condition. This is a group that receives no prevention program and gets only assessments.
4. Active business as usual. This is a group that receives "usual" active treatment (e.g., school-based sex education, college orientation that contains some sex education) that may be effective in preventing sexual violence but is not the focal treatment of the study. This treatment must be limited to services that the group would receive whether or not the research study was implemented (e.g., mandated college orientation). Programs that focus specifically on bystander behaviors or IPV/sexual assault are not eligible controls; however, sex education programs that may contain limited information on violence are eligible.
5. Other.

[name]
Program name. Write in the program name or label for this group. Please use the exact name the study authors report (many of the programs will be branded – Bringing in the Bystander, Mentors in Violence Prevention).

[descrip]
Program description. Write in a brief description of the treatment this group receives. As much as possible, quote or give a close paraphrase of the relevant descriptive text in the study report; always include page numbers from the report when appropriate. It is acceptable to copy and paste directly from the article as long as you include the information in quotations and provide a page number for the quotation.

PARTICIPANT CHARACTERISTICS

[permale]
Enter the percent of participants who identify as male in this group. Use -9 for cannot tell.

[perfemale]
Enter the percent of participants who identify as female in this group. Use -9 for cannot tell.

[pertrans]
Enter the percent of participants who identify as transgender in this group. Use -9 for cannot tell.

[perwhite]
Enter the percent of White participants in this group. Use -9 for cannot tell.

[pernonwhite]
Enter the percent of Non-White participants in this group. Use -9 for cannot tell.

[perblack]
Enter the percent of Black participants in this group. Use -9 for cannot tell.

[perhisp]
Enter the percent of Hispanic participants in this group. Use -9 for cannot tell.

[perasian]
Enter the percent of Asian participants in this group. Use -9 for cannot tell.

[age]
Enter the average age of the group using number of years. Use -9 for cannot tell.

[agerange]
Enter the age range of the group using "XX-XX" format. Use -9 for cannot tell.

[ed_level]
What is the grade/educational level for the majority of the group?
    1. 7th-8th grade
    2. 9th-10th grade
    3. 11th-12th grade
    4. First year college
    5. Second year college
    6. Third year college
    7. Fourth year college
    8. Other/Cannot tell

[perfrat]
Enter the percent of participants in this group who reported being a member of a fraternity. Use -9 for cannot tell.

[persor]
Enter the percent of participants in this group who reported being a member of a sorority. Use -9 for cannot tell.

[pergreek]
Enter the percent of participants in this group who reported being a member of a fraternity or sorority. Note that this item is of value when the study authors do not break apart fraternity and sorority membership. Use -9 for cannot tell.

[perathlete]
Enter the percent of participants in this group who reported being a member of a school athletic team. Use -9 for cannot tell.

[rape_myth]
Enter the mean rape myth acceptance score for participants in this group at baseline. Use -9 for cannot tell.

[empathy]
Enter the mean victim empathy score for participants in this group at baseline. Use -9 for cannot tell.

[per.sexual.victim]
Enter the percent of participants in this group who reported being a victim of sexual assault (by any type of perpetrator). Use -9 for cannot tell.

[per.physical.victim]
Enter the percent of participants in this group who reported being a victim of physical assault perpetrated by a partner (i.e., hookup, date, relationship). Use -9 for cannot tell.

[perknow]
Enter the percent of participants in this group who report knowing someone who has been a victim of IPV and/or sexual assault. Use -9 for cannot tell.

[efficacy_mean]
Enter the mean bystander efficacy score for participants in this group at baseline. Use -9 for cannot tell.

[perefficacy]
Enter the percent of participants in this group who reported confidence in their ability to intervene when witnessing violence or its warning signs (measure at baseline). Use -9 for cannot tell.

[intentions_mean]
Enter the mean bystander intention score for participants in this group at baseline. Use -9 for cannot tell.

[perintentions]
Enter the percent of participants in this group who reported intentions to intervene when witnessing violence or its warning signs (as measured at baseline). Use -9 for cannot tell.

[bystanderbx_mean]
Enter the mean number of times participants in this group reported engaging in pro-social bystander behavior (i.e., intervening when witnessing any type of violence or its warning signs) at baseline. Use -9 for cannot tell.

[perbystanderbx]

Enter the percent of participants in this group who reported engaging in pros-social bystander behavior (i.e., intervening when witnessing any type of violence or its warning signs) at baseline. Use -9 for cannot tell.

INTERVENTION GROUP CHARACTERISTICS
[bystander_tx]
To what extent did the intervention focus on bystander behavior?
1. Bystander intervention was the main focus of the program.
2. General sexual assault prevention program with some bystander content. Note that sexual assault programs should not include content on physical violence (these will likely constitute IPV prevention programs).
3. General IPV prevention program with some bystander content. Note that IPV prevention programs can focus on physical violence exclusively or on both physical and sexual violence. Programs focusing exclusively on sexual violence should be coded as sexual assault prevention programs.
4. Other/Cannot tell. Please bring these to the attention of the PI, as they may not be eligible.

[mixedsex]
What was the sex composition of the intervention group (in delivery setting)?
1. All females.
2. All males.
3. Single-Sex groups for mixed-sex sample.
4. Mixed sex.
5. Cannot tell.
6. NA – Individual Implementation.

[sex_perp]
How did the program conceptualize the sex of perpetrators of IPV/sexual assault?
1. Male perpetrators only.
2. Female perpetrators only.
3. Gender-neutral. Generic/ambiguous portrayal of sex or explicitly highlighting an equal possibility of male and female perpetrators.
4. Mostly male perpetrators.
5. Mostly female perpetrators.

[sex_vic]
How did the program conceptualize the sex of victims of IPV/sexual assault?
1. Male victims only.
2. Female victims only.
3. Gender-neutral. Generic/ambiguous portrayal of sex or explicitly highlighting an equal possibility of male and female victims.
4. Mostly male victims.
5. Mostly female victims.

[gender_txt]
Describe any intervention program content that addressed gender norms, rape myths, or other cultural norms as a root of IPV and/or sexual assault (or as an influence on bystander behavior). Where possible, please indicate whether gender/cultural norm content applies to sexual or physical violence. If a specific (branded) program is implemented (and a treatment manual or

other program materials may be available) please clearly indicate this, as we may wish to obtain these materials for further coding.

[format]
In what format was the program typically delivered?
1. In-person. This involves some type of presentation or instruction delivered by one or more facilitators.
2. Video. This includes videos that are viewed on televisions or large screens as well as on computers (by groups or individual participants).
3. Web/computer delivered. This applies to non-interactive presentations that are viewed online as well as interactive computer-based programs. It excludes programs that primarily consist of participants viewing videos on computers (these should be coded as videos).
4. Ad/poster campaign. This applies to programs that are designed to promote awareness and contain minimal content (i.e., posters or flyers distributed across campus, brief emails sent to a listserv).
5. Other/cannot tell.

[tx_context]
In what context was the program typically delivered?
1. Individual with facilitator
2. Small groups (<10) with facilitator
3. Large group or whole classroom with facilitator
4. Individual alone (e.g., self-guided)
5. Other/cannot tell

[facilitator]
Who typically delivered the program?
1. Teachers or school administrators
2. Athletic coaches
3. Peer educators
4. Medical professionals (e.g., doctors, nurses, clinic staff)
5. Campus resource center staff (e.g., women's center, sexual assault center, LGBTQ student center)
6. Non-campus community agency (e.g., domestic violence shelter staff/volunteers)
7. Mixed (no predominant type)
8. NA – Self-guided/implemented
9. Other/cannot tell

[monitored]
Monitoring of program implementation. Was the implementation of the program monitored by the author/researcher or program personnel to assess whether it was delivered as intended?
1. Yes, but no indication of feedback to facilitators. Do not infer that monitoring happened. Select "yes" ONLY if the report specifically indicates that implementation was monitored.
2. Yes, with indication that facilitators received feedback. Do not infer that feedback was provided. Select "yes" ONLY if the report specifically indicates that feedback was provided.
3. No indication that service delivery was monitored.

[impprob]
Implementation quality. Based on evidence or author acknowledgment, was there any uncontrolled variation or degradation in implementation or delivery of the program, e.g., high dropouts, erratic attendance, low compliance, program not delivered as intended, wide

differences between settings or facilitators, etc. Note that this question has to do with variation in program delivery, not research contact. That is, there is no "dropout" if all participants complete treatment, even if some fail to complete the outcome measures.

1. Yes
2. Possible
3. No, apparently implemented as intended

[impfid_txt]
Implementation fidelity. Provide a description of any other implementation fidelity measures or assessments, including page numbers where appropriate.

[frequency_contact]
Approximate (or exact) frequency of contact between participant and facilitator or program activity. This refers only to the element of program that is different from what the control group receives.

1. Single-day program
2. Less than weekly
3. One to two times a week
4. Three to four times a week
5. Daily contact
6. Cannot tell

[txwks]
Duration of program in WEEKS. Approximate (or exact) mean number of weeks for the period over which participants received the program, from first to last intervention contact, excluding follow-ups designated as such. Divide days by 7; multiply months by 4.3; multiply years by 52; round to a whole number. Estimate for this item if necessary and if you can come up with a reasonable order of magnitude number (e.g., take the midpoint of a range if it is all that's provided). Code -9 if cannot tell.

[txhours]
Duration of program as received in CONTACT HOURS. Approximate (or exact) mean number of contact hours for the period over which the participants received the program, from first program contact to last contact, excluding follow-ups designated as such. If dosage rates are provided multiply intended duration by dosage rate. Code -9 if cannot tell.

[facilit]
Facilitator training, preparation, or qualifications. Describe any information provided about the intervention facilitators' training, level of preparation, or qualifications required for delivery. Provide page numbers.

[incent]
Incentives for recruitment or participation. Describe any incentives for participant recruitment and/or participation. Provide specific information about incentives (including dollar amounts), when available. Provide page numbers.

Step 1. Study and DV Identification

Create one record for each dependent variable that you will be coding. For example, if the study measures bystander intentions and bystander behavior, you will have two dependent variable records. This is different from the number of times a dependent variable is measured in a study. For example, if the study measures bystander intentions before and after the intervention, you will have only one record here – for the bystander intention measure (but you will have two effect sizes for this outcome measure: one at pretest and one at posttest).

[varid]

Variable number. This number is an identification number for the dependent variable you are coding. Each dependent variable is numbered consecutively, within the study you are coding so that each has a unique VarNo for that study. If there is only one dependent measure for this study, you will create only one record in this worksheet, and the variable number will be 1. If there are three dependent measures, they will be numbered 1, 2, and 3.

[dv.outcome.label]

Enter a label for the variable you are coding. When possible, use the exact label that the study report authors use.

[dvdes]

Write in a brief description of the measure you are coding. This should include the authors' label for this variable (e.g., rape myth acceptance, bystander efficacy, etc.), the instrument, the direction of scoring (e.g., lower scores are better), and information about what is being measured. Quote or closely paraphrase the description that is provided in the original report.

For group equivalence variables make sure the label describes successes in terms of bystander attitudes and behaviors (e.g., females, younger participants, participants in lower education/grade levels, participants who have been IPV/sexual assault victims, participants who know an IPV/sexual assault victim, participants with lower rape myth acceptance or greater victim empathy). When coding race always default to white v. non-white. Code non-white as a success.

[dvmicro]

What type of measure are you coding?

Bystander Intervention Behavior
1. Intervention – sexual violence – female victim
2. Intervention – sexual violence – male victim
3. Intervention – sexual violence – unknown victim sex
4. Intervention – physical violence – female victim
5. Intervention – physical violence – male victim
6. Intervention – physical violence – unknown victim sex
7. Intervention – unclear violence type – female victim
8. Intervention – unclear violence type – male victim
9. Intervention – unclear violence type – unknown victim sex

Bystander Attitudes
10. Bystander efficacy
11. Bystander intentions

Violence Perpetration

12. Sexual Assault Perpetration – female victim
13. Sexual Assault Perpetration – male victim
14. Sexual Assault Perpetration – unknown victim sex
15. Physical IPV Perpetration – female victim
16. Physical IPV Perpetration – male victim
17. Physical IPV Perpetration – unknown victim sex
18. IPV Perpetration – unclear violence type – female victim
19. IPV Perpetration – unclear violence type – male victim
20. IPV Perpetration – unclear violence type – unknown victim sex

Violence Attitudes

21. Victim empathy
22. Rape myth acceptance

Victimization

23. Sexual Assault Victimization
24. Physical IPV Victimization
25. Know a victim of IPV or sexual assault

Other Predictor Variables

26. Age
27. Sex
28. Race/ethnicity
29. Grade/education level
30. Other predictor of bystander behavior or IPV/sexual assault

Prerequisite Skills for Intervening

31. Noticing a sexual assault or its warning signs
32. Identifying a situation as appropriate for intervention
33. Taking responsibility for acting/intervening
34. Knowing strategies for helping/intervening

[dvtype]

Type of data collection used for outcome measure.

1. Pencil & paper questionnaire
2. Online/computer assisted questionnaire
3. In-person interview
4. Phone interview
5. Other
6. Cannot tell

[dvdays]

Number of Days: Enter the number of days over which outcome was counted. Enter -8 for lifetime measures (e.g., "ever" experienced sexual assault). Enter -9 if cannot tell. Multiply months by 30 (e.g., enter 3 months as 90 days).

[icc]

For cluster randomized trials, please enter the intraclass correlation coefficient (ICC) for each outcome variable coded. Code -8 for not applicable and -9 for cannot tell.

Although this is the final section of coding, it is a good idea to identify at least one codable effect size before you start coding a study, because studies that appear eligible frequently end up presenting data that cannot be coded into an effect size.

[reportid]
Report ID for this effect size. Indicate the report number (e.g., 22.02) for the report in which you found the information for this effect size. This is important so that we can find the source information for the effect sizes later on, if necessary, and is especially important for studies with multiple reports.

[page]
Page number for this effect size. Indicate the page number of the report identified above on which you found the effect size data. If you used data from two different pages, you can type in both, but use a comma or dash between the page numbers.

[estype]
Type of effect size you are coding.
1. Pretest and Posttest
2. Group equivalence

There are 3 types of effect sizes that can be coded: pretest, posttest, and group equivalence (or baseline similarity) effect sizes. They are defined as follows:
- **Group equivalence effect size**. Group equivalence effect sizes are used to code the equivalence of two groups prior to treatment delivery on (a) sex, (b) age, (c) race/ethnicity, or (d) predictors of pro-social bystander behavior. When multiple racial/ethnic group compositions are reported please report only white/non-white proportions (if not available, select another racial/ethnic group). When available, always code sex, age, and race/ethnicity. When multiple other predictors are reported select the three deemed to be most relevant (NOTE that behaviors are more relevant than attitudes/intentions).
- **Pretest effect size**. This effect size measures the difference between an intervention and comparison group <u>before</u> the intervention (or at the beginning of the intervention) <u>on the same variable</u> used as an outcome measure. Note: Use pretest data for different analytic samples if available. (e.g., separate pretest data for different follow-up waves).
- **Posttest/follow-up effect size.** This effect size measures the difference between two groups <u>after intervention receipt</u> on some outcome variable. Some posttests can occur during the intervention (after intake), immediately after the intervention ends, or any subsequent follow-up period after the intervention ends.

GROUP SELECTION
[esgroup1] [esgroup2]
Select the groups being compared in this effect size. Always select the focal prevention program to be 'group 1.'

DEPENDENT VARIABLE SELECTION
[varid]
Select the dependent variable for this effect size.

[estiming]

Timing of measurement. Approximate (or exact) number of weeks **after the end of the intervention** when measurement occurred. Divide days by 7; multiply months by 4.3. Enter -9 if cannot tell, but try to make an estimate if possible.

EFFECT SIZE CALCULATION AND DATA ENTRY
It is now time to identify the data you will use to calculate the effect size and to calculate the effect size yourself if necessary.

You need to determine what effect size format you will use for each effect size calculation. There are two general formats you can use, each with its own section in FileMaker:
1. Compute ES from means, sds, variances, test statistics, etc.
2. Compute ES from frequencies, proportions, contingency tables, odds, odds ratios, etc.

Also note that within each of the above effect size formats, effect sizes can be calculated from a variety of statistical estimates; to determine which data you should use for effect size calculation, please refer to the following guidelines in order of preference:

1. Compute ES from regression coefficients with statistical controls for pretest measures and other potential confounding measures at baseline.
2. Compute ES from univariate descriptive statistics (means, sds, frequencies, proportions).
3. Compute ES from test statistics (t, F, Chi square).
4. If significance tests statistics are unavailable or unusable but p-values and degrees of freedom (df) are available, determine the corresponding value of the test statistic (e.g., t, chi-square) and compute ES as if that value had been reported. If you encounter these types of data, please see the PI.

Note that if the authors present both covariate adjusted and unadjusted means, you should use the covariate adjusted ones. If adjusted standard deviations are presented, however, they should not be used.

[esfavor]
Which group is favored?

For intervention-control comparisons, the intervention group is favored when it does "better" than the comparison group. The comparison group is favored when it does "better" than the intervention group.

For group equivalence comparisons, the intervention group is favored when it has a higher probability of pro-social bystander attitudes/behaviors than the comparison group (e.g., females, younger participants, participants in lower education/grade levels, participants who have been IPV/sexual assault victims, participants who know an IPV/sexual assault victim, participants with lower rape myth acceptance or greater victim empathy). For the time being, code non-white as a success.

Remember that you cannot rely on simple numerical values to determine which group is favored. For example, a researcher might assess rape myth acceptance, and report this as a score. A lower score is better than a higher score, so in this case a lower number, rather than a higher one, indicates a more favorable outcome.

Sometimes it may be difficult to tell which group is better off because a study uses multi-item measures in which it is unclear whether a high score or a low score is more favorable. In these situations, a thorough reading of the text from the results and discussions sections usually can

bring to light the direction of effect – e.g., the authors will often state verbally which group did better on the measure you are coding, even when it is not clear in the data table.

Note that if you cannot determine which group has done better, you will not be able to calculate a numeric effect size. (You will still be able to create an effect size record—just not a numeric effect size.)

Select the group that has done "better":
1. Intervention
2. Comparison
3. Neither, exactly equal
4. Cannot tell

[esdata]
Effect size derived from what type of statistics?
1. N successful/unsuccessful (frequencies)
2. Proportion successful/unsuccessful (percentage successful or not)
3. Means and SDs; means and variances; means and standard errors
4. Independent t-test
5. Chi-square statistic (1 degree of freedom)
6. Effect sizes as reported directly in the study
7. Other statistical approximation

[esadj]
For this effect size, did you use adjusted data (e.g., covariate adjusted means) or unadjusted data? If both unadjusted and adjusted data are presented, you should use the adjusted data for the group means or mean difference, but use unadjusted standard deviations or variances. Adjusted data are most frequently presented as part of an analysis of covariance (ANCOVA). The covariate is often either the pretest or some personal characteristic such as socioeconomic status.
1. Unadjusted data
2. Pretest adjusted data (or other baseline measure of an outcome variable construct)
3. Data adjusted on some variable other than the pretest (e.g., socioeconomic status)
4. Data adjusted on pretest plus some other variables

[estxasn]
Assigned N for the intervention group

[exctasn]
Assigned N for the comparison group

[estxobn]
Observed N for the intervention group

[exctobn]
Observed N for the comparison group

[estxmean]
Mean for intervention group

[esctmean]
Mean for comparison group

[estxsd]
Standard deviation for intervention group

[esctsd]
Standard deviation for comparison group

[escella]
N successful for intervention group

[escellc]
N successful for comparison group

[escellb]
N failed for intervention group

[escelld]
N failed for comparison group

[esindt]
Independent t-value

[eschisq]
χ^2 (df=1)

[esauth]
Standardized mean difference (SMD) reported by authors

[esor]
Odds ratio (OR) reported by authors

[eshdse]
Standard error (for SMD or OR) reported by authors

Final Effect Size Determination
[es_fmsmd]
Effect size value- standardized mean difference

[es_fmor]
Effect size value- odds ratio

Remember that you cannot rely on simple numerical values to determine which group has done better. For intervention-control comparisons, a positive effect size should indicate that the intervention group did "better" on the outcome measure than the comparison group, while a negative effect size indicates that the comparison group did "better" than the intervention group, and a zero effect size means that the two groups are exactly equal on the measure.

You must make sure that the sign of the effect size matches the way we think about direction, such that the effect size is positive when the intervention group (or posttest) is better and negative when the comparison group (or pretest) is better.

Effect sizes can range anywhere from around –3 to +3. However, you will most commonly see effect sizes in the –1 to +1 range.

Note: If the authors report an effect size, include that in your coding and use it for the final effect size value if no other information is reported. However, if the authors also include enough information to calculate the effect size, always calculate your own and report it in addition to that reported in the study.

[esprob]
Any problems coding this effect size?