

SERVICES FOR IPV VICTIMS: ENCOURAGING STRONGER RESEARCH METHODS TO PRODUCE MORE VALID RESULTS

BY MELISSA RORIE, BETHANY BACKES AND JASPREET CHAHAL

Using rigorous research methods like randomized controlled trials can help increase confidence in evaluation findings.



Violence against women remains prevalent in the United States. In 2010, six out of every 1,000 women¹ experienced intimate partner violence (IPV; sometimes referred to as “domestic violence”), and two out of every 1,000 women were raped or sexually assaulted.²

Victims need careful and often intensive services to help address their physical, emotional and financial suffering. To that end, the federal government funds a variety of programs that provide victims with transitional housing, legal aid, counseling, job training and other assistance. But in recent years, funding for these programs has declined. For example, in 2013, the Violence Against Women Act (VAWA) — which covers not only victims’ services but also research, batterer interventions and criminal justice capacity building — received \$388 million, down from \$412.5 million in 2012. According to the National Network to End Domestic Violence, shrinking budgets

have caused service providers to close victims’ shelters, reduce programming hours and cut program staff.³

Given the strain on funding, it is critical that the programs that do receive money effectively increase safety, increase victims’ knowledge of and confidence in legal options, and help them recover from victimization. Yet few programs to date have been evaluated using rigorous research designs and systematic study. Fewer than half of 18 seminal studies on IPV victims’ services employed a randomized controlled trial (RCT), generally thought

to be the strongest method for establishing whether a program is effective.⁴ As we move forward, more efforts should be made to use rigorous research methods like RCTs when evaluating victims' services programs.

Prior Research

NIJ has supported efforts to understand violence against women, including preventing this type of violence and helping women to recover. NIJ has a long history of funding research on:

- Protection orders
- Policy and legislation
- Victim services and advocacy
- Risk factors for homicide and serious injury
- Evaluations of grants funded under VAWA

Program evaluations help inform policy and strategies to improve the lives of victims. To date, evaluations of victims' services in the IPV field generally have focused on advocacy programs that connect women to legal, financial and emotional support. These studies tend to show positive results: Women participating in advocacy programs are more likely to be involved in legal proceedings and have better success in obtaining resources and support.⁵

Another important branch of IPV research has focused on counseling services; studies have generally found that receiving counseling helps women build better self-esteem, assertiveness, social support, coping skills and self-efficacy.⁶ There have also been evaluations on multidisciplinary responses to sexual assault, specifically Sexual Assault Nurse Examiner (SANE) programs. These studies indicate that SANE programs result in more cases reaching the final stages of prosecution; for example, such programs increase the likelihood of a guilty plea and conviction.⁷ Finally, research has shown that shelters are one of the most important and effective services for abused women.⁸ Although these findings are all promising, their "certainty" may be diminished if evaluations of the programs' effectiveness failed to use an experimental research design.

Overall, the use of strong methods to evaluate victims' services programs is wanting. In a 2009 review that examined 18 intervention studies for victims of IPV, only three of the studies qualified as rigorous based on three criteria:

- Random assignment of participants to a treatment group (i.e., individuals receiving the intervention) or a control group (i.e., individuals receiving "business as usual")
- At least 20 cases included in each group
- An outcome of IPV recidivism or violence severity⁹

Most studies to date do not account for inherent differences between participants who receive services and those who do not. This means that any benefits attributed to receiving treatment may not be due to the program itself but instead may be due to a characteristic that treatment participants share — for example, being highly motivated to receive services or having more social support encouraging them to seek services. When we evaluate programs using methods that account for this kind of "selection bias" — for example, through random assignment to groups — we might find that the programs are not as effective as once thought.

When looking at the most frequently cited studies in the IPV literature, we find that with the exception of a select few, most of the studies used a quasi-experimental design (QED) at best — that is, a design that compares a treatment group to a control group but in which researchers cannot control who is assigned to which group (more information on QED is provided below). These studies do not explain with certainty whether the treatment provided directly relates to the positive outcomes. These studies cannot say that the comparison groups are truly equivalent prior to treatment, so we are left guessing whether the effect seen is due to something other than the interventions. (See a table comparing 18 IPV studies at NIJ.gov, keywords: IPV victim services.)

QED methods are incredibly beneficial and often demonstrate strong correlational links between interventions and positive outcomes. However, service providers can better argue that a program works if its evaluation uses RCT methods, because they

better account for differences in groups and thereby increase confidence that the outcome is due to the program.

Understanding Research Designs

The desire to improve services for victims should serve as the driving force behind the choice of a research design. In other words, we should craft research that will provide the most accurate and useable information. However, the choice of a research design must also consider the fiscal resources available to agencies and researchers as well as their capacity to carry out the research design.

Generally, RCTs are thought to be the best method for increasing the accuracy of study results because they require a comparison between a treatment group and a control group. Participants are randomly assigned to the treatment group, thus allowing researchers to balance out certain pre-existing characteristics — for example, motivation to seek treatment — or reduce their impact so that the researchers can evaluate the program more clearly.

Here's an example: An IPV services agency begins a counseling program for battered women. The agency conducts an evaluation that compares outcomes for women who *volunteered* for counseling with outcomes for women who declined counseling and finds that counseling improves client self-esteem and quality of life. In this example, it is easy to see that increased self-esteem and quality of life might be due to the client's desire to better herself and seek social support — and not to the program itself. On the other hand, if the agency *recruits* women who are interested in counseling and then randomly assigns them to either counseling or treatment-as-usual conditions, the finding that the counseling group experienced better outcomes is more clearly related to the program and not necessarily to motivation.

RCTs are not without their own limitations, however. Many practitioners and researchers alike argue that RCTs are too simplistic and verge on artificiality — that is, they may not account for the way that certain individual characteristics or environments make the treatment more or less effective.¹⁰ An important question, then, is how well a program implemented

The desire to improve services for victims should serve as the driving force behind the choice of a research design.

under “optimal conditions” — that is, under careful scrutiny during an RCT — will work when implemented in the real world. Researchers should demonstrate not only that the program is effective during the research phase but also that ground-level service providers can implement the program with fidelity.¹¹

It also may be problematic to assume that randomization creates unbiased treatment and control groups. Even in these types of experiments, it is important to know who the individuals are who make up the sample and whether differences exist between the groups despite randomization. Similarly, randomization does not correct for volunteer bias in the sample as a whole. In other words, even if there are no differences among people in the treatment and control groups, there may be differences among people participating in the study and those who do not.¹²

Some situations might render an RCT design inappropriate. For example, another research design might be better at encouraging participation and, therefore, might produce results that more accurately represent the population of interest as a whole.¹³ This is particularly important in the IPV field, as these victims might be particularly reluctant to have any form of information collected, no matter how many assurances of confidentiality are provided.

In other cases, program staff might be highly resistant to randomization despite researchers' best efforts to allay concerns, or the cost and resources involved in a full randomization might be prohibitive (see sidebar, “Practitioner and Researcher Concerns About RCTs”). In these cases, alternative methods such as QEDs might be available that allow researchers

to manipulate the data and sampling to simulate treatment and control groups. Some argue that QEDs might even be more desirable than RCTs because researchers must be more deliberate about how comparisons are made.¹⁴ In this article, we argue that RCTs are the best method for establishing causality, but it is worth noting that QEDs can make a valuable contribution when random assignment is not feasible.

One example of a QED is the use of “nonequivalent control groups.” In this method, researchers compare participants who receive treatment to those who do not, but they do not randomly assign participants to the groups. Instead, the researchers attempt to ensure that the groups are as similar as possible before the intervention begins. This can be done in IPV studies, for example, by comparing individuals who, prior to the evaluation, had similar risk for revictimization.

Another strategy is a “time-series” design. Researchers collect data for several time points prior to the treatment being introduced and for a long follow-up period to determine how incidences of victimization changed after the intervention was put

into place. Having data on longer periods of time helps ensure that the intervention’s effect is real and that any changes in victimization are not due to temporary fluctuations.

No matter which research method is used, researchers must always explain the reasons for their decision, be aware of any limitations that might arise from the use of that method, and understand how these limitations affect the results and the interpretation of the results. When RCTs are not feasible due to ethical concerns, costs, buy-in or other reasons, QEDs serve as an alternative that can increase the credibility of the results when compared with most nonexperimental evaluations.

Of course, like RCTs, QEDs have their own limitations and might not always be appropriate for the research questions. Table 1 compares how well RCTs and QEDs establish that the intervention directly causes changes in relevant outcomes. Both methods can determine that an association exists between the program and the outcomes as well as ensure that the intervention precedes the outcome of interest (i.e., the designs

Table 1. Comparing Randomized Controlled Trials With Quasi-Experimental Designs

	Randomized Controlled Trials	Quasi-Experimental Designs
Can the research method establish an association between the intervention and the outcomes?	Yes. One can determine whether an individual’s treatment status is related to changes in the outcome.	Yes. One can determine whether an individual’s treatment status is related to changes in the outcome.
Can the research method establish that the treatment preceded the desired outcome?	Yes. The outcome is measured after the treatment is implemented.	Yes. Generally, there are assessments prior to and after treatment implementation.
Can the research method rule out other explanations?	Yes. Randomization helps ensure that the treatment and control groups are equivalent on both observed and unobserved variables.	No. Although matching ensures balance on observed variables, differences might exist that are not measured. With time-series designs, there might be historical events that impact the respondents, or respondents might change their attitudes or behaviors naturally over time.

establish accurate time-ordering). However, QEDs are less successful at ruling out alternative explanations for the results.

Encouraging the Use of More Rigorous Designs

The best program evaluations will use the most rigorous method possible to successfully gather evidence about a program's effect. When deciding which method is appropriate, researchers and practitioners should consider the benefits and the limitations of RCTs. However, even given the method's limitations, it is important that we do not dismiss RCTs without strong reasons for doing so. Randomization and the use of a control group generally enhance the accuracy of research findings. Furthermore, funding agencies increasingly prioritize rigorous methods, as these agencies attempt to show that public money is being used efficiently.

So how do we mitigate the concerns of both practitioners and researchers and encourage the use of RCTs?

- (1) **Establish a strong researcher-practitioner relationship:** Practitioners often have strong opinions about whether a program works and whether a client needs the program (as well as potential safety concerns should the client not receive the intervention). Researchers must be ready to explain why an RCT is the best method by which to study a program and should work with practitioners to resolve any concerns about safety or nontreatment. Some research suggests that conducting an RCT soon after a program's development — that is, before practitioners establish expectations about a program's success — might be beneficial.¹⁵

Practitioner staff might want to override an assignment of an individual to the control group, perhaps because of potential safety concerns. In these cases, establishing a procedure where such overrides can take place with the approval of upper-level management might be necessary.¹⁶ Alternatively, the pool of an agency's clients who

are chosen for randomization might be those who are less at risk than the clientele of the agency in general — many RCTs in the IPV realm draw only from misdemeanor cases.¹⁷

In addition, it is often difficult to recruit clients to enroll and remain in the program for its duration; working closely with program staff can make this process much easier. Practitioners and researchers should work together to develop methods to increase safety (for example, by ensuring that no participant can be identified in the data) and to encourage clients to remain involved in the program while the evaluation is being conducted.

- (2) **Link funding to research designs:** A constant concern for agencies and organizations is the availability of resources. RCTs are time-intensive and at times can cost more money than alternative research methods for a variety of reasons, including the amount of training required. External funders, such as NIJ, must emphasize the desirability of RCTs and widely disseminate examples of such rigorous studies. When applying for funds, programs and researchers should make sure that they carefully calculate the resources necessary to implement the design accurately. And if RCTs show that a program has no effect, then funding should focus on alternative program development and evaluation.¹⁸

By using more rigorous research methods like RCTs, we can increase confidence in evaluation findings. In doing so, service providers will be better positioned to find funding for programs and to encourage clients to seek out life-enhancing resources.

About the Authors

Melissa Rorie is an assistant professor of criminal justice at the University of Nevada, Las Vegas. Bethany Backes is a social science analyst at NIJ. Jaspreet Chahal is a doctoral student at George Mason University and a graduate research assistant at NIJ.

Practitioner and Researcher Concerns About RCTs

Although the use of randomized controlled trials (RCTs) can improve the validity of research results, it is not without complications. Practitioners and researchers commonly cite six concerns about RCTs:

Practitioner Concerns

- (1) **Victim safety:** When it comes to implementing experimental designs and collecting data in intimate partner violence (IPV) research, the victim's anonymity must be protected. The most dangerous time for IPV victims is when they decide to leave the relationship or reach out for services.¹ Providers take great care to ensure victims are out of harm's way — for example, shelters are unmarked, victims must be a certain distance from the shelter in order to meet anyone, and the Health Insurance Portability and Accountability Act protects any medical treatment they seek. It is imperative that victims' anonymity be preserved to maintain their safekeeping. Practitioners' concern for victim safety is justified and should be highly considered when using experimental designs.
- (2) **Equity in services:** In RCTs, the control group is not deprived of services. In most designs, the control group receives the agency's standard services, whereas the treatment group receives the agency's standard services *and* the enhanced intervention being tested.

Still, is it ethical to deny additional services to potential recipients based on the luck of the draw? Will victims receiving only basic treatment face increased harm compared with the treatment (enhanced services) group? Conversely, is it ethical to deliver additional services without knowing whether they are helpful or even detrimental to victims? Is it ethical to deliver services that often require significant time and emotional investment from the victims without knowing whether the services work? Some researchers argue that RCTs might actually be more ethical than other methods because of the improved ability to determine whether a program is helpful and whether there are any adverse effects.²

- (3) **Agency burden of evaluation:** Many victims' services agencies are small and based on a highly collaborative, mission-driven model. Being mandated to evaluate its programs requires an agency to spend staff time and funding on research instead of on direct service. Without adequate funding to ease these additional pressures, this demand on resources might be seen as harmful to victims seeking help.³ Further, experimental designs are not only the most rigorous method of evaluation but also can be more expensive than other methods. As such, RCTs may be more costly than small programs can afford.

There are also the potential consequences of doing an evaluation. If an evaluation indicates that the services an organization provides are not effective, then what? Has the nonprofit organization used its resources for "nothing" because the evaluation showed no effect? Where does the organization go from here? Program staff may fear that unfavorable findings will put them out of business. They also may argue that quantitative results do not accurately depict their work or the successes they see on a daily basis. The agency may measure success using anecdotal evidence, like victims' stories, and may argue that the program's success cannot be measured with quantitative data, regardless of the strength of the methods.

In cases in which the treatment does not demonstrate the desired effect, researchers and practitioners should collaborate to understand why the program is not working and what benefits the program is having. They should then modify the program based on the findings. Practitioners and researchers must work together to establish within-agency routines that allow for continuous program performance evaluations after the researchers withdraw.

Researcher Concerns

(1) **Implementation:** The greatest concern for researchers is that practitioners — who know their clients well and are in a good position to make recommendations to them — will use anecdotal information to alter the random assignment of the experiment or trial. When practitioners feel certain that the program is effective or when they fear the effect that unfavorable or nonsignificant findings will have on support or funding, concerns with implementing the research design become more salient. This is a problem with all evaluations, not just RCTs.⁴ But with RCTs in particular, practitioners might not fully understand the underlying assumptions and methods and, therefore, might not know what role they are supposed to take in the evaluation.⁵

It is important for researchers to work with practitioners and provide feedback throughout the evaluation. Assuring practitioners that they will have an opportunity to voice their interpretations about the results might help alleviate their fears about the consequences of the evaluation.⁶ The Enhanced Nurse Family Partnership Study in Los Angeles offers a prime example of a successful collaboration between researchers and practitioners. In this project, researchers and the Multnomah County Health Department worked together to integrate an IPV prevention program into the department's long-standing nurse-family partnership program and to test it using RCT methods.⁷

(2) **Confidentiality issues:** If a researcher does not take sufficient care to ensure data confidentiality and fails to comply with institutional review board (IRB) review, data might be subpoenaed and a victim's privacy compromised.⁸ Because of this concern, IRBs place most experiments that evaluate victims of IPV under special scrutiny. It is critical that researchers follow IRB recommendations, as well as human subjects and confidentiality regulations dictated by their funding agencies, and that they work to address any confidentiality and safety issues when designing the study.⁹

(3) **Cost to researchers:** Researchers and practitioners alike find that designing and implementing an RCT is incredibly time consuming. Practitioners are crunched for time as they struggle with overwhelming caseloads, and researchers might prefer to do something less arduous so that they can publish their results quickly. One study asked researchers why they would choose randomized methods over less rigorous designs: One of the most important motivators researchers cited was funding agency priorities. Subsidizing the randomization process might be an important way to increase the use of such designs.¹⁰

These concerns need to be addressed before deciding that an RCT is the best method to evaluate a specific program. Collaboration and funding might help alleviate some discomfort with this design. However, if concerns for victim safety and privacy remain high, then using alternative methods might be warranted.

Practitioner and Researcher Concerns About RCTs *(continued)*

Notes

1. Campbell, Jacquelyn C., Daniel Webster, Jane Koziol-McLain, Carolyn Block, Doris Campbell, Mary Ann Curry, Faye Gary, Nancy Glass, Judith McFarlane, Carolyn Sachs, Phyllis Sharps, Yvonne Ulrich, Susan A. Wilt, Jennifer Manganello, Xiao Xu, Janet Schollenberger, Victoria Frye, and Kathryn Laughon, "Risk Factors for Femicide in Abusive Relationships: Results From a Multisite Case Control Study," *American Journal of Public Health* 93 (7) (2003): 1089-1097, available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1447915>.
2. Weisburd, David, "Ethical Practice and Evaluation of Interventions in Crime and Justice: The Moral Imperative for Randomized Trials," *Evaluation Review* 27 (3) (2003): 336-354.
3. Bennett, Larry, Stephanie Riger, Paul Schewe, April Howard, and Sharon Wasco, "Effectiveness of Hotline, Advocacy, Counseling, and Shelter Services for Victims of Domestic Violence: A Statewide Evaluation," *Journal of Interpersonal Violence* 19 (7) (2004): 815-829, available at <http://tigger.uic.edu/~lwbenn/lwb/JIVDVOutcomes.pdf>.
4. This, itself, might affect the results: If the treatment is more effective for victims who are more at risk, then using a pool of less vulnerable victims will likely dull the treatment impact seen in the study. Such concerns should, of course, be weighed against the need to protect the safety of the victim. See Labriola, Melissa, Michael Rempel, and Robert Carl Davis, "Testing the Effectiveness of Batterer Programs and Judicial Monitoring: Results From a Randomized Trial at the Bronx Misdemeanor Domestic Violence Court," Final report to the National Institute of Justice, grant number 2001-WT-BX-0506, November 2005, NCJ 245144, available at <http://www.courtinnovation.org/sites/default/files/battererprogramseffectiveness.pdf>.
5. Feder, Lynette, Phyllis Holditch Niolon, Jacquelyn Campbell, Jan Wallinder, Robin Nelson, and Hattie Larrouy, "The Need for Experimental Methodology in Intimate Partner Violence: Finding Programs That Effectively Prevent IPV," *Violence Against Women* 17 (3) (2011): 340-358.
6. Labriola, Rempel, and Davis, "Testing the Effectiveness of Batterer Programs and Judicial Monitoring: Results From a Randomized Trial at the Bronx Misdemeanor Domestic Violence Court."
7. Olds, David L., Charles R. Henderson, Jr., Robert Tatelbaum, and Robert Chamberlin, "Improving the Life-Course Development of Socially Disadvantaged Mothers: A Randomized Trial of Nurse Home Visitation," *American Journal of Public Health* 78 (11) (1988): 1436-1445.
8. Sherman, Lawrence W., and Richard A. Berk, "The Specific Deterrent Effects of Arrest for Domestic Assault," *American Sociological Review* 49 (2) (1984): 261-272; Sullivan, Cris M., "Evaluating Domestic Violence Support Service Programs: Waste of Time, Necessary Evil, or Opportunity for Growth?" *Aggression and Violent Behavior* 16 (4) (2011): 354-360.
9. Labriola, Rempel, and Davis, "Testing the Effectiveness of Batterer Programs and Judicial Monitoring: Results From a Randomized Trial at the Bronx Misdemeanor Domestic Violence Court."
10. Lum, Cynthia, and Sue-Ming Yang, "Why Do Evaluation Researchers in Crime and Justice Choose Non-Experimental Methods?" *Journal of Experimental Criminology* 1 (2) (2005): 191-213.

NCJ 247880

Cite this article as: Rorie, Melissa, Bethany Backes and Jaspreet Chahal, "Services for IPV Victims: Encouraging Stronger Research Methods to Produce

More Valid Results," *NIJ Journal* 274 (2014): 24-33, <http://nij.gov/journals/274/Pages/ipv-victim-services.aspx>.

Notes

1. "Women" includes females aged 12 and older.

2. Catalano, Shannan M., *Intimate Partner Violence, 1993-2010*, Special Report, Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics, November 2012, NCJ 239203, available at <http://www.bjs.gov/index.cfm?ty=pbdetail&iid=4536>; Planty, Michael, Lynn Langton, Christopher Krebs, Marcus Berzofsky, and Hope Smiley-McDonald, *Female Victims of Sexual Violence, 1994-2010*, Special Report, Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics, March 2013, NCJ 240655, available at <http://www.bjs.gov/content/pub/pdf/fvsv9410.pdf>.
3. National Network to End Domestic Violence, "Funding and Appropriations [Internet]," 2014, available at <http://www.nnedv.org/policy/issues/funding.html>.
4. Coalition for Evidence-Based Policy, *Hierarchy of Study Designs for Evaluating the Effectiveness of a STEM Education Project or Practice*, Washington, D.C.: Author, 2007, available at <http://coalition4evidence.org/wp-content/uploads/2009/05/study-design-hierarchy-6-4-09.pdf>; Sherman, Lawrence W., Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter, and Shawn Bushway, *Preventing Crime: What Works, What Doesn't, What's Promising*, Final report to the National Institute of Justice, grant number 96-Mu-Mu-0019, 2008, available at <https://www.ncjrs.gov/works>.
5. Bennett, Larry, Stephanie Riger, Paul Schewe, April Howard, and Sharon Wasco, "Effectiveness of Hotline, Advocacy, Counseling, and Shelter Services for Victims of Domestic Violence: A Statewide Evaluation," *Journal of Interpersonal Violence* 19 (7) (2004): 815-829, available at <http://tigger.uic.edu/~lwbenn/lwb/JIVDVOOutcomes.pdf>; Sullivan, Cris M., and Deborah I. Bybee, "Reducing Violence Using Community-Based Advocacy for Women With Abusive Partners," *Journal of Consulting and Clinical Psychology* 67 (1) (1999): 43-53; Wathen, C. Nadine, and Harriet L. MacMillan, "Interventions for Violence Against Women: Scientific Review," *JAMA: The Journal of the American Medical Association* 289 (5) (2003): 589-600.
6. National Network to End Domestic Violence, "Funding and Appropriations [Internet]"; Mancoske, Ronald J., Dale Standifer, and Cathleen Cauley, "The Effectiveness of Brief Counseling Services for Battered Women," *Research on Social Work Practice* 4 (1) (1994): 53-63.
7. Campbell, Rebecca, Debra Patterson, and Lauren F. Lichty, "The Effectiveness of Sexual Assault Nurse Examiner (SANE) Programs: A Review of Psychological, Medical, Legal, and Community Outcomes," *Trauma, Violence, & Abuse* 6 (4) (2005): 313-329; Campbell, Rebecca, Debra Patterson, and Deborah Bybee, "Prosecution of Adult Sexual Assault Cases: A Longitudinal Analysis of the Impact of a Sexual Assault Nurse Examiner Program," *Violence Against Women* 18 (2) (2012): 223-244.
8. National Network to End Domestic Violence, "Funding and Appropriations [Internet]"; Lyon, Eleanor, Shannon Lane, and Anne Menard, *Meeting Survivors' Needs: A Multi-State Study of Domestic Violence Shelter Experiences, Final Report*, Washington, D.C.: U.S. Department of Justice, National Institute of Justice, February 2008, NCJ 225025, available at <https://www.ncjrs.gov/pdffiles1/nij/grants/225025.pdf>; Sullivan, Cris M., "Evaluating Domestic Violence Support Service Programs: Waste of Time, Necessary Evil, or Opportunity for Growth?" *Aggression and Violent Behavior* 16 (4) (2011): 354-360.
9. Stover, Carla Smith, Amy Lynn Meadows, and Joan Kaufman, "Interventions for Intimate Partner Violence: Review and Implications for Evidence-Based Practice," *Professional Psychology: Research and Practice* 40 (3) (2009): 223-233.
10. Dobash, R. Emerson, and Russell P. Dobash, "Evaluating Criminal Justice Interventions for Domestic Violence," *Crime & Delinquency* 46 (2) (2000): 252-270.
11. Flay, Brian R., Anthony Biglan, Robert F. Boruch, Felipe González Castro, Denise Gottfredson, Sheppard Kellam, Eve K. Mościcki, Steven Schinke, Jeffrey C. Valentine, and Peter Ji, "Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination," *Prevention Science* 6 (3) (2005): 151-175.
12. Ross, Sue, Adrian Grant, Carl Counsell, William Gillespie, Ian Russell, and Robin Prescott, "Barriers to Participation in Randomised Controlled Trials: A Systematic Review," *Journal of Clinical Epidemiology* 52 (12) (1999): 1143-1156; Rosnow, Ralph, and Robert Rosenthal, "Taming of the Volunteer Problem: On Coping With Artifacts by Benign Neglect," *Journal of Personality and Social Psychology* 30 (1) (1974): 188-190.
13. Sampson, Robert J., "Gold Standard Myths: Observations on the Experimental Turn in Quantitative Criminology," *Journal of Quantitative Criminology* 26 (4) (2010): 489-500.
14. Lyon, Lane, and Menard, *Meeting Survivors' Needs: A Multi-State Study of Domestic Violence Shelter Experiences, Final Report*.
15. Davis, Robert C., and Bernard Auchter, "National Institute of Justice Funding of Experimental Studies of Violence Against Women: A Critical Look at Implementation Issues and Policy Implications," *Journal of Experimental Criminology* 6 (4) (2010): 377-395.
16. Ross, Grant, Counsell, Gillespie, Russell, and Prescott, "Barriers to Participation in Randomised Controlled Trials: A Systematic Review."
17. This, itself, might affect the results: If the treatment is more effective for victims who are more at risk, then using a pool of less vulnerable victims will likely dull the treatment impact seen in the study. Such concerns should, of course, be weighed against the need to protect the safety of the victim. See Sherman, Lawrence W., and Richard A. Berk, "The Specific Deterrent Effects of Arrest for Domestic Assault," *American Sociological Review* 49 (2) (1984): 261-272; Labriola, Melissa, Michael Rempel, and Robert

Carl Davis, "Testing the Effectiveness of Batterer Programs and Judicial Monitoring: Results From a Randomized Trial at the Bronx Misdemeanor Domestic Violence Court," Final report to the National Institute of Justice, grant number 2001-WT-BX-0506, November 2005, NCJ 245144, available at <http://www.courtinnovation.org/sites/default/files/battererprogramseffectiveness.pdf>. Also, although any randomized research will help increase the understanding of the benefits of a program, evaluating the program's impact only on lower-risk clients means the results of the study would apply only to that particular type of client, and follow-up work would need to determine whether the treatment is effective with higher-risk clients.

18. Ross, Grant, Counsell, Gillespie, Russell, and Prescott, "Barriers to Participation in Randomised Controlled Trials: A Systematic Review."