



Georgetown University Law Center
Scholarship @ GEORGETOWN LAW

2017

Beyond the RCT: Integrating Rigor and Relevance to Evaluate the Outcomes of Domestic Violence Programs

Deborah Epstein

Georgetown University Law Center, epstein@law.georgetown.edu

Lisa A. Goodman

Boston College

Cris M. Sullivan

Michigan State University

This paper can be downloaded free of charge from:

<https://scholarship.law.georgetown.edu/facpub/2047>

<https://ssrn.com/abstract=3148261>

Deborah Epstein, Lisa A. Goodman & Cris M. Sullivan, *Beyond the RCT: Integrating Rigor and Relevance to Evaluate the Outcomes of Domestic Violence Programs*, 39 *Am. J. Evaluation* 58-70 (2018)

This open-access article is brought to you by the Georgetown Law Library. Posted with permission of the author. Follow this and additional works at: <https://scholarship.law.georgetown.edu/facpub>



Beyond the RCT: Integrating Rigor and Relevance to Evaluate the Outcomes of Domestic Violence Programs

American Journal of Evaluation
1-13

© The Author(s) 2017

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/1098214017721008

journals.sagepub.com/home/aje



Lisa A. Goodman¹, Deborah Epstein², and Cris M. Sullivan³

Abstract

Programs for domestic violence (DV) victims and their families have grown exponentially over the last four decades. The evidence demonstrating the extent of their effectiveness, however, often has been criticized as stemming from studies lacking scientific rigor. A core reason for this critique is the widespread belief that credible evidence can derive only from research grounded in randomized control trials (RCTs). Although the RCT method has its strengths, we argue that it is rarely an optimal—or even a possible—approach for evaluating multifaceted DV programs. This article reviews the reasons that RCT is a poor fit for such programs and argues that a more inclusive conceptualization of credible evidence is critical to expanding our knowledge base about how DV programs affect survivors' safety and well-being.

Keywords

randomized control trials, program evaluation, outcomes, domestic violence, intimate partner violence

For over four decades, domestic violence¹ (DV) has been the focus of serious attention from activists, legislators, researchers, and policy makers, as well as public and private funders. Violence within an intimate relationship is no longer considered a private problem in which the state should not interfere; today, our society recognizes such violence as an enormous social problem requiring a broad-based societal response (Messing, Ward-Lasher, Thaller, & Bagwell-Gray, 2015). This attitudinal sea change has resulted in exponential growth of local and national hotlines, dedicated shelters, support groups, advocacy and counseling programs, and other community-based services for DV survivors and their families. These services have goals ranging from providing immediate

¹ School of Education, Boston College, Chestnut Hill, MA, USA

² Georgetown University Law Center, Washington, DC, USA

³ Michigan State University, East Lansing, MI, USA

Corresponding Author:

Lisa A. Goodman, School of Education, Boston College, Campion 310, 140 Commonwealth Ave. Chestnut Hill, MA 02467, USA.

Email: goodmalc@bc.edu

safety to enhancing survivors' longer term social and emotional well-being (Goodman & Epstein, 2008; Sullivan, 2016). Despite the enormous investment of funding, time, effort, and creativity that have gone into DV programs, evidence demonstrating their effectiveness has often been criticized as insufficiently scientifically rigorous (e.g., Rorie, Backes, & Chahal, 2014). Why is this the case?

A core reason for this critique is a persistent belief that credible evidence of program effectiveness can derive only from research grounded in the randomized control trial ("RCT") method, where intervention recipients are compared to those in a randomized control group (Smyth & Schorr, 2009). Although the RCT method was originally developed to evaluate medical interventions, its influence has expanded well beyond its originally intended scope. Over the last several decades, RCT has come to be considered the "gold standard" in determining "what works," not only in assessing medical interventions but also in determining the effectiveness of complex social programs, including those serving survivors of DV (California Evidence-Based Clearinghouse for Child Welfare, 2016; Chatterji, 2007; Schorr & Farrow, 2014). One effect of this view has been that funders and policy makers are often reluctant to support non-RCT studies or to rely on them to guide decision-making (Chatterji, 2007; Rorie et al., 2014; Schorr & Farrow, 2014).

The RCT method has numerous strengths when applied in the right circumstances. However, it is rarely an optimal—or even a feasible—approach for evaluating multifaceted social programs. Identifying more appropriate and meaningful alternatives to the RCT is not simply an academic exercise. It is crucial to use the most appropriate methods to uncover evidence that is valid, reliable, and meaningful to those working to support DV survivors. An overemphasis on RCTs deprives DV programs, policy makers, and funders of critical information needed to identify best practices and influence innovative approaches.

In this article, we first outline why the RCT design is generally a poor fit for evaluating DV programs, which, like many other social service programs, provide complex, individually tailored, flexible interventions, and are embedded in large, highly variable public systems and community contexts. We then argue for a broader conceptualization of credible and rigorous evidence in the assessment of DV program effectiveness. Only by drawing on a range of research methodologies and utilizing those that fit program realities can we expand our existing knowledge base and our corresponding ability to combat the significant problem of DV.

The Misplaced Preference for RCTs in Evaluating Complex, Multifaceted Social Service Programs

RCTs have long been considered optimal for testing drugs and other medical interventions because the method allows for a high level of confidence in identifying the causal relationship between independent and dependent variables ("internal validity"). Randomly assigning participants to either a clearly defined experimental group (which receives, e.g., a new medication) or a control group (which receives a placebo or an established medication), and holding all else constant, provides a high level of confidence that it is the drug itself—not the doctor–patient relationship, a placebo effect, or good luck—that is responsible for any subsequent differences between the groups (Campbell & Stanley, 1963; Wendt, 2006).

But national and local policy makers have encouraged, and even required, application of the RCT method far beyond the context of discrete, targeted medical interventions, in a wide variety of disciplines involving social program evaluation (Chatterji, 2007). Federal funders have followed suit, expressing a strong preference for research that follows the classic experimental design. For example, a 2014 National Institute of Justice report urged DV researchers to use an RCT design to determine program effectiveness whenever possible, citing a limited confidence in evidence obtained through nonexperimental research designs (Rorie et al., 2014). Similarly, the Centers for

Disease Control has noted that for an intervention to be considered “well supported,” it must have been evaluated using an RCT (Puddy & Wilkins, 2011).

This federal preference for RCTs has an enormous impact on DV program research and practice. Grants administered by federal agencies account for a substantial proportion of available research funding in the field, thus sharply limiting funding for non-RCT designs. In addition, interventions that *are* evaluated through non-RCT methods are far less likely to be valued and widely disseminated through important federal resources. For example, the Substance Abuse and Mental Health Services Administration (2016) publishes a National Registry of Evidence-based Programs and Practices (NREPP). NREPP, which “promotes the adoption of scientifically established behavioral health interventions,” by reviewing and reporting on the evidentiary support for mental health and substance abuse interventions, will only consider for inclusion evaluations based on experimental or highly controlled quasi-experimental designs. This restriction, common among many evidence registries (e.g., Cochrane Reviews), contributes to the veneration of studies that have employed RCTs, and the undervaluation of evidence gathered through other methods. As many researchers, as well as funders and policy makers, rely on these repositories to identify the best available evidence, the absence of evaluations using nonexperimental designs hampers the ability of DV researchers and program developers to gain easy access to important non-RCT-based studies.

The RCT method is a mismatch for evaluating DV programs for a number of reasons. These reasons are grounded in the fact that these, like many other social programs, are wholly different in nature than discrete, targeted medical interventions. As others have pointed out, many:

[p]romising social programs are sprawling efforts with multiple components requiring constant mid-course corrections, the active involvement of committed human beings and flexible adaptation to local circumstances [T]he very nature of successful programs makes them almost impossible to evaluate like one would a new drug. (Schorr & Yankelovitch, 2010)

Renowned evaluation scholar Michael Scriven may have put it best when he noted: “The RCT design is a theoretical construct of considerable interest, but it has essentially zero practical application to the field of human affairs” (2008, p. 12).

As discussed in detail below, DV programs are purposefully designed to provide a large number of complementary services and supports to meet the diverse needs of different DV survivors. Services within one single agency, for example, can include shelter, advocacy, counseling, support groups, children’s programming, and community building and community organizing. These forms of support are often intertwined and overlapping, may be provided by multiple advocates, and may be accessed simultaneously through a single program. Moreover, programs exist in complex and variable public system and community contexts. These contextual factors are not simply variables to be controlled for purposes of experimental design. Instead, they determine how program services develop, how survivors make use of programs, and what outcomes are possible (Blamey & MacKenzie, 2007). As a result, a program’s full, complex impact on the lives of its clients cannot be determined by studies whose focus is limited to individual, disaggregated services in controlled contexts. Put simply, the RCT research design is inadequate to the task (Kidwell & Hyde, 2015; Silverstein & Maher, 2008).²

DV Programs Focus on Participants With Complex Needs Who Rarely Meet Strict Research Eligibility Criteria

Survivors who come to DV programs are often, if not usually, residentially unstable, economically impoverished, and physically endangered, and have been subject to multiple forms of trauma and discrimination (Davies & Lyon, 2014; Goodman & Epstein, 2008). Some have come to distrust formal service systems that have caused harmed to them or people they know. Others are seeking

help for the first time. Some have mental health disorders, struggle with substance abuse, or have disabilities. Such factors can profoundly shape clients' experience of violence as well as their choice of and response to supports and services provided (Collins, 1998; Goodman, Smyth, Borges, & Singer, 2009). Most advocates are deeply committed to serving survivors regardless of the complexity of their challenges or the point they have reached in their journey toward safety and stability. In addition, some federal funding sources either strongly urge or require advocates to adopt this approach (e.g., U.S. Department of Health and Human Services Family and Youth Services Bureau, 2016; U.S. Department of Justice Office on Violence Against Women, 2015).

The very complexity of survivors' needs and backgrounds render RCTs highly problematic for evaluating the programs they access. For example, the RCT approach requires that research participants be as similar to each other as possible, to minimize factors that can confound results (e.g., variation across mental health issues), even when doing so requires screening out many people who need and could benefit from the intervention being evaluated. This limitation on the RCTs' capacity to achieve external validity—that is, the degree to which results can be generalized to different people (Cook & Campbell, 1979)—makes it especially ill-suited to the evaluation of DV programs that aim to support survivors facing multiple challenges who are most likely to utilize a number of different services offered within an agency. Yet, the impact of programs on these survivors, who often are those in greatest need, is precisely where research is most desperately needed.

DV Programs Eschew “One Size Fits All” Interventions and Outcomes

DV program advocates are not working toward one or two discrete interventions that will work for all DV survivors; rather, they are committed to providing an array of supports to meet different survivors' varying needs. Advocates try to “meet survivors where they are,” doing their best to respond to each individual's particular goals and contexts, which are, in turn, shaped not only by a survivor's individual challenges but also by identity, resources, culture, strengths, and community. It is imperative to maintain flexibility throughout the engagement, as an individual's needs evolve in the face of an often fluid situation (Davies & Lyon, 2014; Kulkarni, Bell, & Rhodes, 2012; Macy, Giattina, Sangster, Crosby, & Montijo, 2009).

For example, an older immigrant woman who lives in a small, tight-knit cultural community, speaks little English, and has limited skills that would allow her to work out of the home, might fear being ostracized and isolated if she leaves her abusive husband. Her goal might, therefore, be to find a way to remain in her relationship but stay safe. The types of services she receives from a DV program could include support groups, advocacy related to immigration, safety planning specific to her situation, and brief peer counseling. In contrast, a woman who is geographically flexible, interested in ending the relationship, and has supportive family in multiple locations and transferable job skills, might have a primary goal of relocating to another city. The types of services she receives would likely include housing advocacy, employment advocacy, safety planning specific to her situation, and perhaps temporary shelter. Further, a survivor might receive different services depending on whether she seeks help from a mainstream program—typically staffed predominantly by professionals and focused primarily on the provision of a particular set of services—versus a program designed to serve a particular cultural or racial group—typically focused primarily on community-building among survivors, between survivors and staff, and within the larger community (Goodman, Banyard, Woulfe, Ashe, & Mattern, 2016; Serrata et al., 2017).

Individual survivors' goals and needs also may change during their participation in a DV program. For example, a woman leaving an abusive husband may initially seek help moving to a new community. The program staff member who is working with her—usually called an advocate—will partner with her to find affordable housing and help her explore how the move might affect other aspects of her life. As the survivor reflects on her options and the potential challenges her advocate

raises, she may decide that a different course makes more sense, such as trying to repair relationships with her parents and siblings, so they can support her safety and assist her in remaining in her current community. The advocate would view this shift as a survivor's progress toward better goal discernment. In response, the advocate would refocus the work, perhaps arranging a meeting with the survivor's family to begin to heal damaged relationships. Working with this survivor would entail an evolving combination of housing support, counseling, and network strengthening. As this example illustrates, most DV programs serve survivors best by adopting a flexible, constantly changing approach.

The RCT methodology was developed to assess the effect of a standardized intervention on narrowly defined and easily measurable outcomes. These studies require the adoption of a standardized, consistent approach, as the goal is to evaluate an intervention that is highly replicable. The RCT method was not designed to examine the individually tailored, flexible, nonstandardized interventions that are typically provided by DV programs.

DV Programs Are Designed to Be Responsive to Diverse Community Contexts

DV programs are intentionally responsive to variable local community needs, values, strengths, and limitations. The services that program advocates provide in a particular community are substantially shaped by the specifics of local politics and cultural practices, transportation systems, civil and criminal justice systems, and the availability and quality of public and affordable private housing, mental health and substance abuse treatment programs, educational opportunities, and services for children and families.

For example, program strategies may look quite different in a community where police officers routinely arrest both parties in a DV dispute than in a community where police offer tailored responses to victims and their families. In the former, program advocates may be oriented to helping survivors avoid police calls, while in the latter, advocates may work collaboratively with the police to encourage survivors to consider formal reporting. Neither program's focus would work in the other community. Similarly, in a community with a sanctuary city policy, a DV program may encourage undocumented survivors to engage with local formal service organizations to get the support they need, whereas in a city where local government officials are cooperating with Immigration and Customs Enforcement, the program might encourage survivors to seek informal support from friends and family.

The RCT method ensures a high degree of confidence in its results by measuring interventions only in highly controlled conditions. Because this approach runs directly counter to the ways that DV programs are designed, an RCT-based study of one program would have limited general applicability: A program based in another substantially different community context could not implement the RCT-evaluated model and expect similar success (Hess & Henig, 2008).

DV Programs Often Have No Clear Start or End Dates

DV programs are purposefully designed to enable survivors to access services as needed, whether that is continuously or sporadically. Although a particular service offered within a DV program—such as a time-limited support group—may have a clear and predetermined time frame, the overall program must provide its clients with maximum flexibility. This is a crucial aspect of program practice, not only because survivors' needs often change over time but also because survivors often must be geographically mobile to maintain their own safety, may be precariously housed or sporadically homeless or may wish to remain hidden, making continuous use of a single program impractical (Center for Policy Research and National Resource Center on Domestic Violence, 2017).

Further, the length of time a survivor makes use of a particular service may depend largely on external variables out of the control of the program: In many communities, for example, the availability of low-income housing stock drives length-of-stay in residential programs. If a public housing unit becomes available, a survivor may leap to take it, whether she is benefiting from shelter services or not. On the other hand, a survivor may be ready to leave but have nowhere to go in communities where there is a shortage of public or subsidized housing. Either way, the presence or absence of affordable housing in the community can thus shape a survivor's engagement with the program.

The RCT research design requires sustained participation in an intervention that is continuous, with a clear beginning, middle, and end point. Only this way can the researcher know that the participant is getting the full "dose" of a clearly specified intervention. This method cannot be implemented in the context of complex, individualized DV programs.

For all of the foregoing reasons,³ by imposing a strong preference for RCT design, funders and policy makers severely constrain the ability of innovative (and even mainstream) DV programs to demonstrate their effectiveness, improve survivor services, and advance the fight to end DV.

Thinking Beyond RCTs to Evaluate DV Programs

Scientific rigor should be defined both by the integrity of the data produced and its value in accurately answering meaningful questions (Whitesell, 2016). What are the meaningful questions related to DV program effectiveness, and what methods can best address them?

Over the many decades during which we have worked with DV programs, we have served as consultants, advocates, lawyers, board members, and researchers. In each of these roles, our researcher, advocate, and survivor colleagues have shared their thoughts about the kinds of knowledge they would find valuable in evaluating the effectiveness of the services they provide. Below we list three types of questions that have arisen frequently in these discussions. Each can be explored through non-RCT research methods that balance the demands of both rigor and relevance.

How Can We Better Understand What Works When Our Clients Use Our Services in Such Different Ways?

Program advocates are—of course—eager to understand as much as possible about what approach works best to serve their client base. But as discussed above, survivors require individually tailored advocacy responses that are based on their own individual goals and circumstances; they avail themselves of various services simultaneously, serially, or with long-time gaps in-between; and they seek help from programs in communities with widely varying systems and resources. While evaluations of discrete services are useful, they do not capture the complexity of many survivors' service use and therefore tell an incomplete story. The essential question of what works can instead be addressed with a naturalistic, longitudinal methodology—that is, gathering data from participants in their everyday settings over time—to identify patterns of service use that appear to yield positive outcomes (Sites, Masaracchia, & Davis, 2017). Such an approach would ideally be conducted on a large scale to ensure adequate statistical power to examine subgroup differences and to account for a wide range of contextual variables. For example, if survivors could be interviewed about their particular historical situations and needs at the point of program entry, and then followed and interviewed about their use of services repeatedly over time, researchers could obtain a great deal of information from which patterns of success could be ascertained. By conducting several simultaneous studies, using the same methods, in different parts of the country, researchers could mine a wealth of data to significantly increase our understanding about what works, for whom, under what circumstances.

This type of evaluation is currently underway in Washington State, where the third author is examining the impact of the domestic violence housing first (DVHF) model on survivors' safety, housing stability, and well-being. The DVHF approach is inherently flexible and responsive to individual needs; in addition, the model is likely to differ depending on a survivor's location in rural versus urban settings. Accordingly, the evaluation involves interviewing clients at four agencies (two in urban locations, two in rural ones) about a variety of personal as well as contextual factors. Interviews are conducted when survivors first approach an organization for assistance, and then every 6 months over at least an 18-month period. Extensive information is also being gathered about each of the agencies delivering the services. The design was chosen to maximize external validity and will carefully document the details about what survivors want and receive over time, not just from the agency they were recruited from but from other community sources as well, and will examine how a variety of factors work together to impact survivors' safety, housing stability, and well-being over time.

To What Extent Does the “How,” Rather Than the “What,” of Programs Contribute to Participant Success?

Based on foundational knowledge about the disempowering effects of DV, program advocates have come to general agreement about how services should be delivered across program components. For example, many DV programs adhere to a “survivor-defined approach”—partnering with individual clients to develop their own goals—to ensure that clients are receiving the supports they are looking for (Davies & Lyon, 2014; Kulkarni, Herman-Smith, & Ross, 2015). Advocates also engage in “empowering practice” with survivors, helping ensure they have the knowledge, skills, and self-confidence to gain control over their lives again (Cattaneo & Goodman, 2015; Goodman & Epstein, 2008; McGirr & Sullivan, 2017; Sullivan et al., 2008). A growing consensus also exists that a trauma-informed approach, one that fully integrates understanding of trauma into all program policies, procedures, and practices, will contribute to a range of positive outcomes (Wilson, Fauci, & Goodman, 2015). Moreover, DV programs are committed to providing culturally relevant services (Goodman, Sullivan, et al., 2016). Therefore, although intervention services themselves are not standardized, uniform approaches to the work do exist, undergirding what can appear to be widely divergent services, supports, and contexts.

The effectiveness of these generally agreed-upon approaches can only be evaluated through methods that can answer questions such as “Does strong adherence to specific service delivery approaches—regardless of the intervention and context—result in positive outcomes for a diverse population of clients?” Such research would provide programs with crucial information about how to further develop their work; a recent national report recommends that such research be prioritized (Center for Policy Research and National Resource Center on Domestic Violence, 2017).

One particularly promising method for evaluating the how of service delivery is the *common factors approach*, developed by clinical and counseling psychologists to assess the “active ingredients” of psychotherapy (Wampold, 2010). Instead of assessing the effectiveness of specific types of therapy, the common factors perspective asks the question, “What are the common factors across different therapeutic orientations that account for change?” (Beutler, 2014; Laska, Gurman, & Wampold, 2014; Wampold, 2001). Common factors researchers use naturalistic, longitudinal designs and rely on rigorous statistical methods and controls to increase confidence that it is the common factor—rather than client, setting, or therapeutic model—that constitutes the cause of the observed effect. Controlling for a range of external variables, for example, researchers consistently have shown that the working alliance between therapist and client predicts positive outcomes across a variety of therapeutic models (Hatcher & Gillaspay, 2006; Horvath, Del Re, Flückiger, & Symond, 2011).

Two evaluations already have demonstrated promising results using this approach within DV agencies. One recent cross-sectional study showed a relationship between survivor-defined practice, as perceived by survivors themselves, and their sense of empowerment in relation to safety (Goodman, Thomas, et al., 2016). Another longitudinal study demonstrated that, among survivors in four separate DV shelters, the degree to which clients received trauma-informed, culturally relevant services related to significant improvement in their self-efficacy and safety-related empowerment over the course of their shelter stay (Sullivan, Goodman, Virden, Strom, & Ramirez, in press). Larger, more rigorous longitudinal studies are needed across different settings to enhance confidence in these findings.

What Can Survivors Tell Us About How They Understand Success, What Contributes to It, and Why?

Another critical set of questions centers on how survivors perceive program effectiveness in helping them achieve meaningful outcomes. The most basic of these questions might be: “How do survivors themselves characterize meaningful outcomes?” Survivors often have complex ways of understanding “success” that do not necessarily map onto the standardized outcome measures available to researchers (Goodman et al., 2015; Melbin, Jordan, & Smyth, 2014; Song, 2012; Sullivan et al., 2008). A detailed understanding of survivors’ dynamic ways of conceptualizing success is a critical prerequisite to meaningfully examining how they might best reach their desired goals. For example, a recent mixed methods study of 301 survivors from a variety of backgrounds and programs found that, for over half, improvements in the domain of safety meant losses in other domains, such as sense of community, economic security, or custody of children (Thomas, Goodman, & Putnins, 2015). Almost half of participants (49.1%) had not expected to experience these losses. These findings provide a critical backdrop for addressing a second question: “What can survivors tell us, systematically, about what works to facilitate their success, and why?” Survivors have the most profound understanding of their own strengths and needs; the supports and services that have made the greatest difference, given their individual goals and contexts; and how cultural, community, and systemic factors have facilitated or impeded program effectiveness. Their firsthand knowledge could, and should, be mined for broader use and constant reexamination. Questions about survivor perceptions and experiences lend themselves to qualitative approaches where survivors’ voices are central, such as interviews, focus groups, community-based participatory research, and participatory action (Martinez, Callejas, & Hernandez, 2010; Schorr & Farrow, 2014).

What Can Advocates Tell Us About What Works and What Kinds of Innovations Are Necessary for Particular Communities?

Advocates, like survivors, have unique perspectives to contribute. Their perspectives may be particularly valuable to address questions such as: “What kinds of new and innovative program approaches appear to have promise?” Program advocates are deeply familiar with the particular communities with whom they work and often have powerful insights into approaches that would be most effective, given community members’ distinctive cultural backgrounds, strengths, and needs. This knowledge often leads to new, innovative practices (Serrata et al., 2017). Sometimes, these novel ideas are utilized only by the individual who conceptualizes them; they may not be shared with other program advocates, much less described and analyzed for use in the broader movement.

Knowledge obtained through systematic exploration of advocate perspectives can be conceptualized as “community-defined evidence.” This term, originally developed to characterize innovative behavioral health practices in Latino/Hispanic communities (Community Defined Evidence Project, 2009), is also useful to describe the unique information that advocates in DV programs can provide. Community-defined evidence comprises a set of practices that have been used and determined to

yield positive results by consensus over time, though they may be unacknowledged or undiscovered outside the bounds of a particular community (Martinez, Callejas, & Hernandez, 2010). Community-defined evidence can be captured by a variety of specific research approaches, including qualitative methods of various types. For example, one focus group study conducted with advocates across a single rural state explored advocates' perspectives on the benefits and costs of practices designed to help survivors who have been isolated by abusive partners regain formal and informal social support (Goodman, Banyard, Woulfe, Ash, & Mattern, 2016). Advocates delineated a set of practices they invented on the fly with their clients to repair relationships disrupted by abuse; they reported high confidence that these practices dramatically improved the well-being of isolated survivors, especially survivors of color and immigrants within a largely white state. The focus group method, by allowing for systematic collection of these insights, could result in the adoption of new practices by advocates in other organizations or locations who otherwise might not have come up with these innovative strategies.

Conclusion

If the RCT is the only valued approach to documenting program effectiveness, yet it is a mismatch for DV programs, then both researchers and programs are in a bind. They can continue to do experimental evaluations of discrete program elements with a subset of clients, providing data that have limited value in terms of furthering an understanding of the field and in terms of programmatic utility. Or, they can use alternative methods to shed more light on complex program work, but take the risk that results will be undervalued. Over time, an overemphasis on RCTs may result in a widening disparity between research and practice: Promising programs for which RCTs are appropriate will continue to be researched, highlighted, and disseminated, and will therefore improve. Promising programs not amenable to RCTs, in contrast, will receive less research funding, will be less highlighted, and will have far fewer opportunities to advance (Whitesell, 2016). Much needed creativity will be stifled and the fight to eradicate DV and support survivors will be substantially undermined.

It is time to adopt a broader, more inclusive understanding of acceptable evaluative data in the DV program context as well as in other organizations serving people with complex challenges. Some of this work is already underway: In 2016, for example, the U.S. Department of Health and Human Services Office of Planning, Research and Evaluation (OPRE) hosted a meeting of national experts to explore the most appropriate methods for evaluating community-based interventions; and in that same year, the OPRE collaborated with the Center for Policy Research and National Resource Center on Domestic Violence to host a meeting on building evidence for DV services and interventions that highlighted a number of non-RCT designs. We hope this article sparks further discussion and debate as we work to find creative ways to appropriately evaluate the important work of DV programs as they work to help survivors transform their lives.

Acknowledgments

We are indebted to Lisbeth (Lee) Schorr, senior fellow at the Center for the Study of Social Policy, who inspired us to write this article, and whose work is reflected throughout.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Although the term intimate partner violence has supplanted the term domestic violence (DV) in much of the academic literature, the term domestic violence is still far more commonly used in the advocacy and service provider arenas. We use the terms interchangeably to refer to the same phenomenon: physical violence, sexual violence, stalking, economic abuse, and psychological aggression (including coercive tactics) by a current or former intimate partner (i.e., spouse, boyfriend/girlfriend, dating partner, or ongoing sexual partner).
2. Of course, there are some instances where randomized control trials (RCTs) or rigorous quasi-experimental designs can produce meaningful knowledge in the DV program context. This is particularly true when the research focuses on discrete, severable program services (Bybee & Sullivan, 2002; Chronister & McWhirter, 2006; Constantino, Kim, & Crane, 2005; Johnson, Zlotnick, & Perez, 2011; Sullivan & Bybee, 1999).
3. There are additional practical aspects of the RCT approach that make it a poor fit for the DV field, including ethical issues related to random assignment of treatment conditions. These considerations have been fully addressed elsewhere and are beyond the scope of this paper.

References

- Blamey, A. M., & MacKenzie, M. (2007). Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation, 13*, 439–455.
- Bybee, D. I., & Sullivan, C. M. (2002). The process through which an advocacy intervention resulted in positive change for battered women over time. *American Journal of Community Psychology, 30*, 103–132.
- Beutler, L. E. (2014). Welcome to the party, but...: A response to Laska, Gurman, and Wampold. *Psychotherapy, 51*, 496–499.
- California Evidence-Based Clearinghouse for Child Welfare. (2016). Retrieved from www.cebc4cw.org
- Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.
- Cattaneo, L. B., & Goodman, L. A. (2015). What is empowerment anyway? A model for domestic violence practice, research, and evaluation. *Psychology of Violence, 5*, 84–94.
- Center for Policy Research, in partnership with National Resource Center on Domestic Violence (2017). *Building evidence for domestic violence services and interventions: Challenges, opportunities, and research priorities*. Submitted to the Office of the Assistant Secretary for Planning and Evaluation Retrieved from <https://aspe.hhs.gov/sites/default/files/pdf/255511/BuildingDV.pdf>
- Chatterji, M. (2007). Grades of evidence: Variability in quality of findings in effectiveness studies of complex field interventions. *American Journal of Evaluation, 28*, 239–255.
- Chronister, K. M., & McWhirter, E. H. (2006). An experimental examination of two career interventions for battered women. *Journal of Counseling Psychology, 53*, 151–164.
- Collins, P. H. (1998). It's all in the family: Intersections of gender, race, and nation. *Hypatia, 13*, 62–82.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.
- Community Defined Evidence Project. (2009). *Preliminary quantitative and qualitative findings*. National Latino Behavioral Health Association and National Network to Eliminate Disparities. Retrieved from http://nned.net/docs-general/CDEP_Final_Report_10-7-09.pdf
- Constantino, R., Kim, Y., & Crane, P. A. (2005). Effects of a social support intervention on health outcomes in residents of a domestic violence shelter: A pilot study. *Issues in Mental Health Nursing, 26*, 575–590.

- Davies, J., & Lyon, E. (2014). *Domestic violence advocacy: Complex lives/difficult choices*. Thousand Oaks, CA: Sage.
- Goodman, L. A., Banyard, V., Woulfe, J., Ash, S., & Mattern, G. (2016). Bringing a network-oriented approach to domestic violence services: A focus group exploration of promising practices. *Violence against Women, 22*, 64–89.
- Goodman, L. A., Bennett Cattaneo, L. B., Thomas, K., Woulfe, J., Chong, S. K., & Smyth, K. F. (2015). Advancing domestic violence program evaluation: Development and validation of the Measure of Victim Empowerment Related to Safety (MOVERS). *Psychology of Violence, 5*, 355–366.
- Goodman, L. A., & Epstein, D. (2008). *Listening to battered women: A survivor-centered approach to advocacy, mental health, and justice*. Washington, DC: American Psychological Association.
- Goodman, L., Smyth, K. F., Borges, A. M., & Singer, R. (2009). When crises collide: How intimate partner violence and poverty intersect to shape women's mental health and coping. *Trauma, Violence and Abuse: Special Issue on the Mental Health Implications of Violence Against Women, 10*, 306–329.
- Goodman, L. A., Sullivan, C. M., Serrata, J., Perilla, J., Wilson, J. M., Fauci, J. E., & DiGiovanni, C. D. (2016). Development and validation of the Trauma Informed Practice Scales. *Journal of Community Psychology, 44*, 747–764.
- Goodman, L. A., Thomas, K., Bennett Cattaneo, L. B., Heimel, D., Woulfe, J., & Chong, S. K. (2016). Survivor-defined practice in domestic violence work: Measure development and preliminary evidence of link to empowerment. *Journal of Interpersonal Violence, 31*, 163–185.
- Hatcher, R. L., & Gillaspay, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16*, 12–25.
- Hess, F. M., & Henig, J. R. (2008). 'Scientific research' and policymaking: A tool, not a crutch. *Education Week*, p. 27, 24, 26, 36.
- Horvath, A., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*, 9–16.
- Johnson, D. M., Zlotnick, C., & Perez, S. (2011). Cognitive behavioral treatment of PTSD in residents of battered women's shelters: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology, 79*, 542–551.
- Kidwell, K. M., & Hyde, L. W. (2015). Adaptive interventions and SMART designs: Application to child behavior research in a community setting. *American Journal of Evaluation, 37*, 1–20.
- Kulkarni, S. J., Bell, H., & Rhodes, D. M. (2012). Back to basics: Essential qualities of services for survivors of intimate partner violence. *Violence Against Women, 18*, 85–101.
- Kulkarni, S. J., Herman-Smith, B. H., & Ross, T. C. (2015). Measuring intimate partner violence (IPV) service providers' attitudes: The development of the Survivor-Defined Advocacy Scale (SDAS). *Journal of Family Violence, 30*, 911–921.
- Laska, K. M., Gurman, A. S., & Wampold, B. E. (2014). Expanding the lens of evidence-based practice in psychotherapy: A Common Factors perspective. *Psychotherapy, 51*, 467–481.
- Macy, R. J., Giattina, M., Sangster, T. H., Crosby, C., & Montijo, N. J. (2009). Domestic violence and sexual assault services: Inside the black box. *Aggression and Violent Behavior, 14*, 359–373.
- Martinez, K., Callejas, L., & Hernandez, M. (2010). Community-defined evidence: A bottom-up behavioral health approach to measure what works in communities of color. *Civic Research Institute, 10*, 11–16. Report on Emotional & Behavioral Disorders in Youth. Retrieved from http://www.civicrosearchinstitute.com/online/article_abstract.php?pid=5&iid=110&aid=702
- McGirr, S., & Sullivan, C. M. (2017). Critical consciousness-raising as an element of empowering practice with survivors of domestic violence. *Journal of Social Service Research, 43*, 156–168.
- Melbin, A., Jordan, A., & Smyth, K. F. (2014, October). *How do survivors define success? A new project to address an overlooked question*. Greenfield, MA: Full Frame Initiative. Retrieved from <http://fullframeinitiative.org/how-do-survivors-define-success-report-recommendations/>

- Messing, J. T., Ward-Lasher, A., Thaller, J., & Bagwell-Gray, M. E. (2015). State of intimate partner violence intervention: Progress and continuing challenges. *Social Work, 60*, 305–313.
- Puddy, R. W., & Wilkins, N. (2011). *Understanding evidence Part 1: Best available research evidence. A guide to the continuum of evidence of effectiveness*. Atlanta, GA: Centers for Disease Control and Prevention.
- Rorie, M., Backes, B., & Chahal, J. (2014). *Services for IPV victims: Encouraging stronger research methods to produce more valid results*. National Institute of Justice Journal. Retrieved from <https://nij.gov/journals/274/Pages/ipv-victim-services.aspx>
- Schorr, L., & Farrow, F. (2014, November 13). *Background paper for the Harold Richman Public Policy Symposium: The Future of Evidence*. Washington, DC. Retrieved from <http://www.cssp.org/policy/evidence/AN-EVIDENCE-FRAMEWORK-TO-IMPROVE-RESULTS.pdf>
- Schorr, L., & Yankelovitch, D. (2010, February 16). What works to better society can't be easily measured. *The Los Angeles Times*. Retrieved from <http://articles.latimes.com/2000/feb/16/local/me-64711>
- Scriven, M. (2008). A summative evaluation of RCT methodology: An alternative approach to causal research. *Journal of Multidisciplinary Evaluation, 5*, 11–24.
- Serrata, J. V., Macias, R. L., Rosales, A., Hernandez-Martinez, M., Rodriguez, R., & Perilla, J. L. (2017). Expanding evidence-based practice models for domestic violence initiatives: A community-centered approach. *Psychology of Violence, 7*, 158–165.
- Silverstein, L., & Maher, E. J. (2008). Evaluation blues: How accountability requirements hurt small, innovative programs the most. *Stanford Social Innovation Review, 6*, 23.
- Sites, B. D., Masaracchia, M. M., & Davis, M. (2017). Distinguishing between efficacy and real-world effectiveness: The case for thinking beyond classic randomized controlled trial design. *Regional Anesthesia and Pain Medicine, 42*, 131–132.
- Smyth, K. F., & Schorr, L. B. (2009). *A lot to lose: A call to rethink what constitutes "evidence" in finding social interventions that work* (Harvard Kennedy School Malcolm Wiener Center for Social Policy Working Paper Series). Retrieved from http://www.hks.harvard.edu/socpol/publications_main.html
- Song, L. Y. (2012). Service utilization, perceived changes of self, and life satisfaction among women who experienced intimate partner abuse: The mediation effect of empowerment. *Journal of Interpersonal Violence, 27*, 1112–1136.
- Substance Abuse and Mental Health Services Administration. (2016). *National registry of evidence-based programs*. Retrieved from <https://www.samhsa.gov/nrepp>
- Sullivan, C. M. (2016). *Examining the work of domestic violence programs within a "social and emotional wellbeing promotion" conceptual framework*. Harrisburg, PA: National Resource Center on Domestic Violence. Retrieved from www.dvevidenceproject.org
- Sullivan, C. M., Baptista, I., O'Halloran, S., Okroj, L., Morton, S., & Stewart, C. (2008). Evaluating the effectiveness of women's refuges: A multi-country approach to model development. *International Journal of Comparative and Applied Criminal Justice, 32*, 291–308.
- Sullivan, C. M., & Bybee, D. I. (1999). Reducing violence using community-based advocacy for women with abusive partners. *Journal of Consulting and Clinical Psychology, 67*, 43–53.
- Sullivan, C. M., Goodman, L. A., Virden, T., Strom, J., & Ramirez, R. (in press). Evaluation of the effects of receiving trauma-informed practices on domestic violence shelter residents. *American Journal of Orthopsychiatry*.
- Thomas, K. A., Goodman, L. A., & Putnins, S. (2015). "I have lost everything:" Trade-offs of seeking safety from intimate partner violence. *American Journal of Orthopsychiatry, 85*, 170–180.
- U.S. Department of Health and Human Services Family & Youth Services Bureau. (2016). *Standing announcement for FVPS domestic violence shelter and supportive services/grants to states*. Retrieved February 6, 2017, from <https://ami.grantsolutions.gov/index.cfm?switch=foa&fon=HHS-2016-ACF-ACYF-FVPS-1115>

- U.S. Department of Justice Office on Violence Against Women. (2015). *OVW fiscal year 2016 transitional housing assistance grants for victims of sexual assault, domestic violence, dating violence, and stalking solicitation*. Retrieved on February 6, 2017, from <https://www.justice.gov/ovw/file/800641/download>
- Whitesell, N. R. (2016, September 22). *Building strong evidence in challenging contexts*. Talk given at the Office of Planning, Research, and Evaluation Conference: Building Strong Evidence in Challenging Contexts: Alternatives to Traditional Randomized Controlled Trials. Retrieved from http://opremethodsmeeting.org/docs/2016/Whitesell_BuildingStrongEvidenceInChallengingContexts.pdf
- Wilson, J. M., Fauci, J. E., & Goodman, L. A. (2015). Bringing trauma-informed practice to domestic violence programs: A Qualitative analysis of current approaches. *American Journal of Orthopsychiatry*, *85*, 586–599.
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.
- Wampold, B. E. (2010). The research evidence for the common factors models: A historically situated perspective. In B. L. Duncan, S. D. Miller, B. E. Wampold, & M. A. Hubble (Eds.), *The heart and soul of change: Delivering what works in therapy* (2nd ed., pp. 49–81). Washington, DC: American Psychological Association.
- Wendt, D., Jr. (2006). Evidence-based practice movements in psychology: Empirically supported treatments, common factors, and objective methodological pluralism. *Intuition: BYU Undergraduate Journal of Psychology*, *2*, 49–62.