

## Research Designs in the Real World: Testing the Effectiveness of an IPV Intervention

by Jill Theresa Messing, Jacquelyn Campbell and Janet Sullivan Wilson

*Many factors can influence study design, particularly when evaluating an intervention in the field. Although randomized controlled trials are considered the gold standard of evaluations, there are practical and ethical considerations that may exclude their use. This case study looks at those factors and their impact on an evaluation of an intimate partner violence intervention.*

Approximately one-third of women experience intimate partner violence (IPV) in their lifetimes.<sup>[1]</sup> Many women call the police when their partners become violent or when the violence becomes more frequent or severe.<sup>[2]</sup> The criminal justice response can hold offenders accountable, but it is not designed to attend to the safety needs of victim-survivors in the same way that domestic violence advocacy agencies are equipped to do.

The Lethality Assessment Program (LAP) is an innovative intervention that occurs at the scene of a police-involved IPV incident and provides risk assessment, followed by advocacy services, for victim-survivors who are at high risk of being killed by their intimate partners. At the program's core is a collaborative partnership between law enforcement agencies and local domestic violence service providers. Police departments and advocacy agencies throughout the U.S. are adopting the LAP,<sup>[3]</sup> but before the current study, little was known about how well this intervention works.

Our NIJ-funded study was the first rigorous evaluation of the LAP. Our objective was to assess the effectiveness of this promising intervention while maintaining the integrity of the LAP and adhering to our ethical principles as researchers and helping professionals. Therefore, choosing the most appropriate research design was paramount.

### The LAP

Developed by the Maryland Network Against Domestic Violence, the LAP brings law enforcement and local domestic violence service providers together to empower IPV victim-survivors in self-care decisions. Near the end of the investigation at an IPV incident scene, the police officer administers a brief risk assessment screen ("Lethality Screen") to gauge the victim-survivor's level of risk for being killed by the IPV offender.<sup>[4]</sup> If a victim-survivor screens in as "high risk," which means having an increased risk of being killed by the intimate partner, then the police officer calls the local domestic violence hotline at a collaborating advocacy organization for information on planning for the victim-survivor's safety ("Protocol Referral").

See "A Closer Look at the Lethality Assessment Program"

### Choosing a Research Design

In our evaluation of the LAP, we examined the intervention's two main goals: (1) decrease the frequency and severity of violence and (2) increase rates of emergency safety planning and help-seeking among women who participate in the intervention. To determine whether the LAP was achieving these goals, we used a quasi-experimental research design in which we could compare two similar groups of people: one group that received the LAP intervention and another group that did not.

Randomized controlled trials (RCTs), also called "true experimental designs," are generally considered the gold standard for evaluation studies because RCTs can rule out alternative explanations for the findings. (See the related article, "Services for IPV Victims: Encouraging Stronger Research Methods to Produce More Valid Results," in issue 274 of the *NIJ Journal*.) In RCTs, the researchers can be relatively certain that any changes found are caused only by the intervention, not by outside influences, because RCTs have three basic characteristics:

- The intervention occurs before measuring the outcome of interest.
- The intervention is given to only some of the participants in the study, creating a comparison.
- The people in the study are randomly assigned into either a group that receives the intervention or a group that does not. Random assignment theoretically ensures that the groups' characteristics are the same before the intervention and that any differences in outcomes between the groups are due to the intervention.

Our ethical obligations as researchers are respect for persons (self-determination), beneficence (do not harm, and maximize the benefits of research), and justice (people should be treated equally).<sup>[5]</sup>

Because the women in our study faced a high risk for homicide due to the fact that they were victims of high-risk IPV cases, we did not feel that we could meet our ethical obligations as researchers or professionals by using an RCT. For instance, if we employed an RCT to evaluate the LAP, we would need to:

- Locate women at the scene of a police-involved IPV incident who would screen in as high risk according to the Lethality Screen.
- Randomize these women into either a group that receives the intervention or a group that does not.
- Gather data from all the women.
- Administer the LAP to the intervention group.
- Gather data from all the women again.

See "Working With Institutional Review Boards"

We could have recruited women at the scene of a police-involved IPV incident, administered the Lethality Screen to determine the women's eligibility, randomized high-risk victim-survivors into intervention and control groups, interviewed the women, placed those in the intervention group on the telephone with a hotline counselor and interviewed everyone again at some follow-up point. In this process, all of the intervention steps would remain intact.

But the LAP is more than the sum of its parts. If we used an RCT design, researchers — not police officers — would administer the Lethality Screen and conduct the Protocol Referral. The intervention would not be administered at the scene of an IPV incident because too many intervening steps would need to occur (first we would need to determine eligibility, and then we would randomly assign the women to groups). Furthermore, practical considerations, such as where the intervention would occur and how to conduct such an intervention with women in high-risk situations, would make study administration difficult.

Moving the LAP out of the field and into a controlled setting would have diminished it in such a way that it would not have been the same intervention. Thus, we agreed that for this research to truly evaluate the LAP, police officers must administer both the Lethality Screen and the Protocol Referral at the scene of an IPV incident for women in the intervention group. Therefore, we would interview women as soon as possible after the police intervened and ask them about their victimization and help-seeking behavior both before and after the incident date.

Still, we struggled with randomization to groups, an important component of an RCT. We considered having officers randomize women into intervention and control groups at IPV incident scenes. However, instructing officers to conduct the LAP with a random selection of participants was logistically impractical. Officers might have chosen to provide the intervention to a participant assigned to the control group, or they might have chosen not to provide the intervention to a participant assigned to the intervention group.<sup>[6]</sup> After being trained on the LAP, officers might also use intervention techniques with the non-intervention group, either consciously or subconsciously.

We considered randomly assigning the intervention by police jurisdiction, but this also made little practical sense. First, there were only two large population centers in the state where we conducted the research, and the regional and geographic differences between them were too large to consider them equivalent. As we moved forward, we discovered that participating jurisdictions had very different operating procedures, implementation fidelity and referral rates. Second, our police and advocacy partners were participating, in part, to receive training and technical assistance on the LAP. To provide this to some partners and not to others — or even to stagger it — would have hindered our researcher-practitioner partnership.

The professional imperatives of our research team (made up of doctoral-level social workers and nurses) and of our advocacy partners also made the idea of random assignment ethically untenable. Both social workers and nurses have ethical obligations to enhance the well-being of research participants and uphold their dignity and worth; the primary commitment of both professions is to help others.<sup>[7]</sup> Determining that women were at high risk for domestic homicide and then withholding a potentially helpful intervention from a randomized group would have been unethical because it placed women's lives at risk.<sup>[8]</sup>

Self-determination is also an important ethical consideration for social workers and nurses. For that reason, we strongly believed that the women should be able to decide independently whether to participate in the intervention, the study or both without one decision affecting another. We wanted the women to be able to choose whether to answer the questions on the Lethality Screen. If they screened in as high risk, they could then choose whether to talk on the phone with the hotline advocate. We also believed that women should be given the choice to participate in the research study regardless of whether they engaged in any aspect of the intervention. Thus, women who received the intervention could choose whether to participate in the study, and women who participated in the study could choose whether to receive the intervention.

In an RCT, a person's ability to receive the intervention is generally contingent upon his or her choice to participate in the study. But because of random assignment, the choice to participate does not guarantee receiving the intervention. In other words, the women might choose to participate in the study in hopes of receiving the intervention, but intervention assignment is not guaranteed. Some RCT designs have attempted to ameliorate this by providing the intervention to the control group after the study ends. But given the high level of risk faced by potential participants and the length of our study (at least six months), we felt that it was important not to withhold or delay intervention for women who wanted to receive it.

### Using a Quasi-Experimental Design

Without random assignment to groups, the study became quasi-experimental; specifically, the study was a nonequivalent-groups quasi-experimental field trial. The groups were nonequivalent because there was no random assignment. Instead, we used an historical comparison group across a previous period.

To create a historical comparison group, we asked the police officers, before training them on the intervention, to refer IPV victim-survivors to researchers when the women evidenced a manifestation of danger (as outlined in the sidebar "A Closer Look at the Lethality Assessment Program") and were willing to speak to a researcher over the telephone. During the study interview, we administered the Lethality Screen but did not score it so that, during analysis, we could determine which women were at high risk and would be included in the comparison group (i.e., those not receiving the intervention).

This ensured that high-risk victim-survivors who later received the intervention would be compared with high-risk victim-survivors who did not.

After we trained the police officers and the advocates on the intervention, the officers then completed the LAP at IPV incident scenes and referred women to the study if the women were willing to have researchers contact them — whether or not the women answered the questions on the Lethality Screen, were determined to be high risk, or talked on the phone to an advocate. Thus, officers pre- and post-intervention used the same criteria to refer women to the study to ensure that the two groups were as similar as possible.

Because we used a historical comparison group, we needed to be particularly attentive to any changes that occurred in participating communities between the times of recruitment of the comparison and intervention groups, such as a high-profile domestic homicide or the closing of a local shelter, because these might affect research outcomes. There were no events that led us to believe that the two groups would differ; however, without random assignment, there were no built-in assurances that they would be similar.

Indeed, the comparison and intervention groups differed in several ways. There were statistically significant differences between the comparison and intervention groups in marital status, immigration status and categories on the Danger Assessment (an IPV risk assessment). We controlled for these differences statistically in our data analysis. However, because participants were not randomly assigned to groups, differences may have existed between the groups that we did not measure and thus could not control statistically.

The risk that we faced with the quasi-experimental research design was that some difference between the groups that we did not measure led to more or fewer protective actions, help-seeking, or frequency and severity of violence among the intervention group but not among the comparison group. Were this to occur, we might have attributed these differences to the LAP when they should instead have been attributed to some other factor. For example, we do not know whether any woman in the comparison group would have agreed to speak with the hotline advocate had she received the intervention. Perhaps the intervention group had some unmeasured characteristic (that the comparison group did not) that affected the women's willingness to participate in the LAP, their decision to take protective actions or their experiences of violence. If that were the case, our research findings would be attributed to the LAP when they should be attributed to this characteristic.

Replication — that is, conducting a similar study with different participants in a different location or with different researchers — is one way to determine whether the results of a study are valid, reliable and generalizable.

- **Valid findings are accurate:** If researchers can replicate study results, then it is more likely that the results reflect real differences between groups or real changes due to an intervention.
- **Reliable findings are consistent:** The same or similar results are found again and again.
- **Generalizable results will translate to different locations and populations:** For instance, an intervention is effective in Oklahoma and Maryland, among Native American women and African American women, and so forth.

Currently, NIJ and the Office on Violence Against Women are collaborating to evaluate two lethality and high-risk assessment models, including the LAP. Two sites will implement the LAP and be rigorously evaluated over the next three to five years.

### How Effective Is the LAP?

Our evaluation of the LAP found that women in the intervention group did, indeed, engage in more protective strategies both immediately after the intervention (e.g., seeking domestic violence services, removing or hiding their partners' weapons) and when we interviewed them approximately seven months later (e.g., applying for and receiving protection orders, obtaining something to protect themselves, seeking medical attention due to violence, going someplace where their partners could not find them). In addition, women in the intervention group had experienced significantly less frequency and severity of violence than women in the comparison group at the follow-up interviews.

To design and conduct this research study, we needed to balance the challenges of engaging in quasi-experimental field research against the requirements of a tightly controlled true experimental design. RCTs have the benefit of controlling for extraneous variables within the design itself and are therefore considered the gold standard for knowing whether an intervention is effective. However, as we discussed above, RCTs require a highly controlled research environment that was neither practical nor desirable in this particular case, which highlights that there is not a single approach to effectiveness trials. To maintain the integrity of the LAP and meet the ethical imperatives of the researchers and community partners, a quasi-experimental design was necessary. Although this design opens the door to outside influences that could affect research outcomes, we believe that this pragmatic field trial provided the best possible information about the effectiveness of the LAP.

NIJ Journal No. 275, posted July 2015

NCJ 248781

### About the Authors

**Jill Theresa Messing** is an associate professor in the School of Social Work at Arizona State University. **Jacquelyn Campbell** is a professor and the Anna D. Wolf Chair in the Johns Hopkins University School of Nursing. **Janet Sullivan Wilson** is an associate professor in the College of Nursing at the University of Oklahoma Health Sciences Center.

[Back to the top.](#)

### For More Information

Read the final report, "[Police Departments' Use of the Lethality Assessment Program: A Quasi-Experimental Evaluation.](#)"

### Notes

[1] IPV is defined here as rape, physical violence or stalking by a current or former intimate partner. Black, Michele C., Kathleen C. Basile, Matthew J. Breiding, Sharon G. Smith, Mike L. Walters, Melissa T. Merrick, Jieru Chen, and Mark R. Stevens, *The National Intimate Partner and Sexual Violence Survey: 2010 Summary Report* (pdf, 124 pages), Atlanta, Ga.: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, November 2011.

[2] Felson, Richard, Steven F. Messner, Anthony Hoskin, and Glen Deane, "Reasons for Reporting and Not Reporting Domestic Violence to the Police," *Criminology* 40 (3) (2001): 617-648.

[3] Virginia Department of Criminal Justice Services, *Review of Lethality Assessment Programs (LAP)* (pdf, 2 pages) [Exit Notice](#), October 2013.

[4] Messing, J.T., J. Campbell, J.S. Wilson, S. Brown, and B. Patchell, "The Lethality Screen: The Predictive Validity of an Intimate Partner Violence Risk Assessment for Use by First Responders" [Exit Notice](#), *Journal of Interpersonal Violence* (May 11, 2015) [epub ahead of print].

[5] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1979.

[6] For example, in the Minneapolis Domestic Violence Experiment, a seminal NIJ-funded randomized control trial on the effectiveness of arrest, officers delivered the intended intervention between 72.8 percent and 98.9 percent of the time, depending on the intervention assigned. Sherman, Lawrence W., and Richard A. Berk, "The Minneapolis Domestic Violence Experiment" (pdf, 13 pages) [Exit Notice](#), *Police Foundation Reports*, April 1984.

[7] American Nurses Association, *Code of Ethics for Nurses with Interpretive Statements*, Silver Spring, Md.: American Nurses Association, 2010; and National Association of Social Workers, *NASW Code of Ethics (Guide to the Everyday Professional Conduct of Social Workers)*, Washington, D.C.: National Association of Social Workers, 2014.

[8] Although the LAP had not been vigorously evaluated, the LAP had been implemented in at least 43 jurisdictions during 2007, the year before the study began. Today, hundreds of jurisdictions across 31 states are using the LAP. The Maryland Network Against Domestic Violence compiles information about participating jurisdictions. During 2007, 3,304 Lethality Screens were administered and 58.2 percent (1,923) of victim-survivors screened in as high risk. Of those victim-survivors who screened in as high risk, 53.6 percent (1,030) talked to the hotline advocate. Of those victim-survivors who talked to the hotline advocate, 25.5 percent (263) went into the collaborating domestic violence agency seeking services (Maryland Network Against Domestic Violence, Lethality Assessment Statistical Information, LAP Report, October 2008). Although experimental research looking at participant outcomes was needed, the research team believed that the available information indicated that the LAP was connecting women with needed resources.