U.S. Department of Justice
Office of Justice Programs
National Institute of Justice

National Institute of Justice

Batterer Intervention: Where Do We Go From Here? Workshop Notes

January 17, 2002

The opinions and conclusions expressed in this document are solely those of the authors and do not necessarily reflect the views of the U.S. Department of Justice.

NCJ 242217

Batterer Intervention: Where Do We Go From Here? Workshop Notes

January 17, 2002

Note: In December 2009, NIJ and the Family Violence Prevention fund co-sponsored a meeting of batterer intervention experts. Read the report from that meeting: Batterer Intervention: Doing the Work and Measuring the Progress.

- Welcoming Remarks
- Batterer Intervention Programs and Strategies for Responding to Batterers: Opening Remarks by Julia Babcock
  - Comments were made by the following participants in response to Dr. Babcock's remarks
- Evaluation Outcomes: Opening Remarks by Dan O'Leary
  - Comments were made by the following participants in response to Dr. O'Leary's remarks
- Evaluation of Recruitment and Retention: Opening Remarks by Chris Eckhardt and Andy Klein
  - Comments were made by the following participants in response to Dr. Eckhardt's and Dr. Klein's remarks
- Evaluation Designs: Opening Remarks by Ed Gondolf and Kevin Hamberger
  - Comments were made by the following participants in response to Dr. Hamberger's and Dr. Gondolf's remarks
- Evaluation Implementation: Opening remarks provided by Rob Davis and Larry Hauser
  - Comments were made by the following participants in response to Rob Davis' and Larry Hauser's remarks
- Recommendations
- Closing Remarks

Welcoming Remarks

Sally T. Hillsman

Dr. Hillsman opened the meeting by thanking the participants for attending. She noted that NIJ acknowledges their important work regarding violence against women and batterer intervention, and that the Institute, along with our colleagues--the Violence Against Women Office (VAWO) and the Centers for Disease Control and Prevention (CDC)--are grateful for the investment of attendees' time and wisdom.

Before Dr. Hillsman discussed the workshop objectives for the day, she brought warm greetings from NIJ's Director, Sarah Hart. She conveyed Director Hart's regrets she could not be at the workshop, as issues surrounding violence against women and batterer intervention are dear to her heart. Dr. Hillsman also briefly discussed Director Hart's past experience in the field, from her position as Chief Counsel for the Department of Corrections to her position as a litigator for the prosecutor's office.

Dr. Hillsman also acknowledged VAWO Director Diane Stuart's presence at the workshop. She noted that VAWO and NIJ have a strong partnership that goes back a number of years and that since many already know Director Stuart, she needs no introduction.

As Dr. Hillsman spoke about the goals and objectives of the workshop, she noted first that NIJ had been thinking about holding this workshop for a while. She talked about how current research findings regarding evaluations of batterer intervention programs (BIPs) have stimulated NIJ's interest in this area. She further noted that part of what has stimulated NIJ's concern is the fact that BIPs are proliferating across the country and made mandatory by a number of statutes, yet those of us in the research field have become uneasy because we don't know how effective these programs are or if there are any negative effects.

Dr. Hillsman also noted that NIJ is concerned about the null effects of these programs and the possibility that there might be no differences between the control and experimental groups. The questions are, she adds, "How do we make things work better?" and "Are there any possibilities that our good intentions are backfiring?"

After stating why NIJ decided to pull this group of individuals together to discuss BIP/evaluation issues, Dr. Hillsman discussed some of the NIJ-funded research in this area. She noted that NIJ has funded a number of evaluations of BIPs: 1) Michele Sviridoff and Rob Davis, who are looking at the impact of BIPs and court monitoring on defendant behavior; 2) Ed Gondolf and Oliver Williams, who are looking at the effectiveness of culturally focused batter counseling for African-American men as compared to conventional batterer counseling; and 3) Chris Eckhardt, who is looking at the stages of change model.

Dr. Hillsman then said that of the completed work to date, the findings are very mixed and that it appears the more rigorous the design, the more likely we are to get a null effect. She also notes that a significant part of the problem is that we do not know whether there are flaws in the methodology or flaws in the treatment program, or both.

Dr. Hillsman said that despite all of these unanswered questions, NIJ remains committed to trying to change the behavior of batterers and to ensuring women's safety. She added that despite some of the limitations mentioned, we must deal with the research finding that these programs may not be working--it is a reality. She said that we must consider, "Are there ways that we can re-conceptualize these programs to make them work better?" In answering this question she notes, it is vital that researchers and practitioners work closely together. She added that the engagement of practitioners is critical to the success of making sure our work is intervening in people's lives and making a difference.

Back to Top

Batterer Intervention Programs and Strategies for Responding to Batterers

Opening Remarks by Julia Babcock

Dr. Babcock began her presentation by listing the five basic types of batterer interventions. The primary intervention, as she noted, is the Duluth model. The Duluth model focuses on patriarchal attitudes as the cause of the violence in the relationship and views changing them as the key to success. The second type of intervention is the cognitive-behavioral model, which looks at things like anger management in a group setting to address batterering. The other types of interventions mentioned were a combination of the Duluth and cognitive-behavioral programs: couples therapy, which involves bringing both partners into counseling together; and individual therapy, which involves the counselor meeting with the batterer alone.

Dr. Babcock then went on to discuss the meta-analysis she conducted, looking at which interventions work the best. Dr. Babcock, along with Dr. Charles Green, examined the findings of studies that evaluate treatment efficacy for batterers. The criteria for inclusion in the meta-analysis were: 1) involvement of a comparison group of batterers and 2) reliance on victim report or police records as a index of recidivism. The research method employed examining 78 empirical studies of efficacy of batterer treatment programs. These studies were then classified according to design: experimental (5) and quasi-experimental (17). The pre-post (48) studies were excluded.

The outcome literature of controlled quasi-experimental and experimental studies was reviewed to test the relative impact of the Duluth model, cognitive-behavioral therapy, and other types of treatment on subsequent recidivism to violence. The findings suggest that the treatment design tended to have only a small influence on effect size. There were no differences in effect size in comparing the Duluth Model versus cognitive-behavioral type interventions using either police records or victim reports. Quasi-experimental designs yielded significantly higher effect sizes than true experiments. Overall, effects due to treatment were in the small range (.10), meaning that current interventions have a minimal impact on reducing recidivism beyond the effect of being arrested. In practical terms, Dr. Babcock noted that the effect size of .10 translates to a 5 percent improvement rate in cessation of violence due to treatment. Dr. Babcock added that while a 5 percent decrease in violence may appear insignificant, this does represent perhaps 42,000 women per year in the United States who are no longer being batterered as a result of treatment.

Dr. Babcock concluded her remarks by noting several caveats. She stated that the effect sizes may be small as a result of measurement error and methodological difficulties common to research in applied settings. BIPs are limited by: variability in the quality of the research studies; high attrition rates; inconsistencies in reporting recidivism for dropouts; low reporting rates at followup; confounds with treatment quality and quality of community response; conservative coding of recidivism as a dichotomous variable; and potential measurement error in both of the recidivism indices.

In closing, Dr. Babcock asked, "Is the low effect size due to measurement problems or the BIPs themselves?" She added that she does not think it is solely attributable to the methods, but rather that batterer treatment programs are not terribly effective. This view, she notes, is backed by Frank Dunford's study, which had drastically different findings.

Back to Top

Comments were made by the following participants in response to Dr. Babcock's remarks

Amy Holtzworth-Munroe commented that there were additional types of treatment, not mentioned by Dr. Babcock, that we should keep track of. She noted Stosny's Attachment Program and psychodynamic and process-oriented treatments.

Rob Davis commented that there are other reasons why we perhaps don't see a bigger effect size. In the first place, there aren't a lot of programs to evaluate--so the problem could be with the programs.

Ed Gondolf commented that the evaluations we have so far are not really rigorous, so the caveats Dr. Babcock offered can go both ways. He noted that the Dunford study is not a

good comparison because the study was conducted in a very different context. It was sponsored by the Navy and done on Navy men, so one could argue that the Navy men are more family oriented. He also added that the conception of batterer programs is that they are very amorphous and thus have lots of issues, noting as examples that no two programs are alike. The operations of the programs themselves vary tremendously administratively and in terms of implementation. Dr. Gondolf concluded his comments by noting that we need to define what we mean by batterer intervention and determine whether we are testing the counseling or the implementation.

Rob Davis commented that the grounds of the discussion seem to have shifted. Do we want to know what kinds of programs work for certain people or do we want to know if the program works?

Dan O'Leary commented that we really need to be careful about how we interpret Dr. Babcock's work. One might conclude--given the small effect size--that we're not doing that much; however, a meta-analysis is a comparison of the different kinds of treatment and this comparative approach suggests that there aren't any differences. The Institutes (NIJ and NIH) need to take seriously the issue of control groups: We will never meet any standards (American Psychological Association) if we lack experimental controls. Dr. O'Leary also questioned the comparison of the Dunford study with the current studies. He notes that the reductions in recidivisms in the Dunford study were so strong, that one might question why. He concluded by noting that we need to look at the experimental designs done with depression studies to get some ideas about how this could best be done.

Amy Holtzworth-Munroe commented that generally, BIPs yield a success rate of two out of three across the board. She said we need to start asking why. What programs are the most beneficial while being most cost effective? What is happening that we aren't reaching the other one-third?

Barbara Hart commented that she was troubled that recidivism was the measure of success. She said that the goal of the criminal justice system, in response to batterer intervention, is not just reducing recidivism, but also safety and restoration. We want to know if these batterers are better at paying their child support. She suggested that for her, the goals should be expanded to include economic reparations, continued support, and safety of the women. We should try to find out if women are still living in fear. Can they make decisions on their own and do the women feel the criminal justice system has helped them? Ms. Hart concluded her comments by noting that we should be asking more nuanced questions and not just asking about recidivism.

David Adams said, "Speaking on behalf of batterer intervention programs, we never thought of ourselves as being about just how well we change batterers. Beyond trying to change batterers, we also try to hold perpetrators accountable." For instance, batterer intervention programs often require their clients to pay child support. Victims are warned about dangerousness, and are notified when batterers are terminated from the program. This information helps victims make more informed choices. One problem is that judges sometimes fail to see accountability as an aspect of change. Dr. Adams also noted that simply saying BIPs don't work implies a too narrow definition of success. He also added that there haven't been many studies that say BIPs don't have beneficial effects in the larger sense. He notes that there is still a lot of trial and error in the field, particularly in determining what motivates change. He adds that BIPs may have a delayed impact on batterers. The notion of delayed positive outcomes is accepted in clinical work with victims, so why not with perpetrators, he asks? Dr. Adams said that he has been approached by

many batterers years after they attended the program, who are only now able to understand and admit their problems. Dr. Adams concluded by saying that long-term effects are very important to consider, even when there are no apparent short-term positive outcomes.

Andy Klein commented that we also have to be aware of the message being sent to the victim when we do place a batterer in a treatment program. One of the effects of sending batterers to treatment is that it makes it harder for victims to leave because they believes the batterer is getting help.

Chris Eckhardt said, "Going back to outcomes, we do not or at least have not fully defined 'treatment,' so that is why we have very ambiguous outcomes."

Rob Davis commented that it seems like we have taken a program type that was originally developed for people who were ready for change and tried to adopt it to very different folks who are not ready for change and then examine the only outcome the system is interested in: recidivism. This may be the problem.

Kevin Hamberger asked, "Do we need to think in terms of a change in long-term recovery as an outcome?" He noted that we need to look at the connections between batterer interventions and resources for women. Batterer intervention does not work in a vacuum and we need to see how we can bring it all together.

Amy Holtzworth-Munroe commented that Andy Klein is exactly right. One of the problems with treatment is the message it sends to the women. The desire to believe a batterer will change entices women back, which is a problem because we do not know if he will change.

Dr. Holtzworth-Munroe also talked about the "quick cure v. delayed impact." She noted that with depression, we get a quick cure but the external problems are much more difficult to deal with--there is no sleeper effect, where the person thinks about it and gets better 6 months later.

Ed Gondolf then noted that there are delayed impacts. All of these measures have a cumulative effect. For example, we may ask a batterer, "How long have you been sober?" He added that when we redefine outcome, we will get a much different picture depending on how outcome is charted--retrospectively.

Radhia Jaaber commented on the fact that there really is not a pure model. Practitioners cross-breed and incorporate different things. She also noted that conducting an experimental design when evaluating BIPs is really artificial. It cuts off the natural ways in which community happens and battering occurs. Measuring recidivism only is very flawed.

Illeana Arias noted that the outcome appears to be a moving target. Is it the point of the programs to treat battering or the batterers? We've started off by saying battering, but it is really the batterers that we want to treat. So if it is batterers, then we must define 'batterer'? We should keep the two concepts separate to pursue our goals.

Barbara Hart said that she would like to look at the process in which batterers engage. She noted that looking at the post-intervention context may make a huge difference and

even considering the community context to which a batterer returns may also have an impact on change.

Sally Hillsman commented that she did not think there were many longitudinal studies conducted on the context of batterering. She noted this is partly because of factors discussed already and whether people who batter over time change. Are these patterns of dissonance and escalation, and where do interventions work?

Amy Holtzworth-Munroe answered that there are very few newlywed studies and most of them only follow couples for 2 to 3 years. She added that what you see is that the guys who were less violent to start, eventually stopped, and the guys who were severely violent at time 1, continued to be severe batterers.

Margaret Zahn commented that in other forms of violence, there is a high correlation between alcohol and other forms of drug abuse. She also noted that the effectiveness of alcohol treatment seems to come in the form of peer support instead of a facilitator's influence and wondered if this was the case for batterer treatment and, if so, how they deal with these additional problems in battering programs.

Ed Gondolf commented that there are a range of ways treatment providers deal with substance abuse. He noted that the batterers may be cycled through a drug and alcohol treatment program as part of the batterers treatment or there may be a separate program that deals with substance abuse treatment.

David Adams commented that the goal of the EMERGE program is to help the batterers create a personal responsibility to help other batterers and to care for self. He added that some batterer programs set out to please the courts by trying to graduate as many people as possible, thereby weakening their standards and possibly not holding the batterers accountable for all they do wrong.

Andy Klein noted that the alcohol question is a key question. He said that judges typically like to refer one person to one program and they have to decide if they will send the person to batterer treatment or substance abuse treatment: The trouble is that the two programs are not compatible.

Larry Hauser responded, "As a Judge, I don't measure success in terms of graduates or recidivism." He noted that he looks at success generally.

Ed Gondolf commented that Julia's [Dr. Babcock's] meta-analysis was very well done, but the caveats that she laid out do not get translated to the field so all the field hears is that BIPs do not work. Also, things we call "small effect sizes," such as .18, might not really be bad. I've seen studies of cognitive-behavioral therapy in a prison population where a .10 treatment effect was considered substantial.

Julia Babcock said there is no way to interpret the effect size, in terms of meaningful impact on the lives of the victims, or in terms of practice or policy decisions about our investment in BIPs.

Sally Hillsman responded by saying we always have that problem when we have statistical significance--we do not know if that is significant to policy or not.

Back to Top

Evaluation Outcomes

Opening Remarks by Dan O'Leary

Dan O'Leary described four of his current projects: 1) He is currently working on a study of kids' aggression towards their parents. This NIMH-funded project studies 450 kids. 2) He is also looking at the strengths and weaknesses of the Conflict Tactics Scale (CTS). 3) He is working with the Air Force on predicting minor and severe aggression. 4) He also is looking at psychological aggression in young married couples and how to prevent it from escalating into physical aggression.

Dr. O'Leary had five points related to evolution and measurement:

1. Measurement of outcomes should be continuous, not dichotomous, and measures should include both physical and psychological aggression. Psychological aggression is a better predictor of a partner's desire or intent to stay in or leave a relationship than physical violence. Outcomes should be measured for both the aggressor and his partner. We should collect information about the context of the violence and have men and their partners write in sentence form exactly what happened and how.
2. Dr. O'Leary also noted that we need to look at issues of control groups as a primary issue. Dan had planned a study of BIP/partner counseling that included a "monitoring, but no-treatment" control group. For political reasons, he abandoned this design. He suggested although no-treatment options may not be feasible for the most violent batterers, where there is less severe risk--as with mild-moderate batterers--a no-treatment group could be feasible. Dr. O'Leary mentioned that when you look at studies of violence over the life course, you see that violence begins around age 12, at the onset of dating, and peaks at around age 25. In the general population, violence declines after young adulthood, although violent behavior may spike again later in life due to Alzheimer's, etc. Court monitoring would be a good control condition for BIP evaluations. Or, second best, a very minimal intervention or an intervention very different from the program being evaluated.
3. Dr. O'Leary then said that we need to look at why some interventions do not work with some people. We need to address the possibility that some severe aggression may not be amenable to psychological treatment. Knowing something about an individual's frequency and severity of violence might be a good predictor of those "untreatable" cases. This fits in with current typology research on batterers. It would be useful to be able to say to a victim, "If you are living in a relationship with [blank] level of violence, it is unlikely that this treatment will change your partner's behavior," and then let the victim decide what to do.
4. Dr. O'Leary asserted that severity of violence should be a primary consideration when assigning men to BIPs.
5. Dr. O'Leary also suggested that the relative risk/predictors for further violence should be considered, so that risk level can be tied to intervention. For example, a self-identified problem drinker is more likely to continue to be violent, and should be treated as such. To say there is no relationship between alcohol and violence is myopic.

Dr. O'Leary added that it would be helpful to assess different systems' impact on aggression: Ed Gondolf's BIPs, courts settings, a batterer's desire to change himself, the family situation--all affect change. I think the same things that make some kids aggressive

may also be general predictors for partner violence: jealousy, inappropriate expectations, and general anger/volatility levels.

Dr. O'Leary added that the value of Julia's meta-analysis is that it shows us there is no one road to change that looks a lot better than any other road. We can see there is no one with a clear answer to the problem, so we should be open to alternatives--this is valuable. Let's look at different things and measurement techniques, and be preemptive.

Back to Top

Comments were made by the following participants in response to Dr. O'Leary's remarks

Radhia Jaaber responded by saying we have to ask women, "How does battering affect how they live, and how does it affect their families and their communities?" We have to ask women questions that are relevant to them.

Andy Klein noted that in interviews with sympathetic researchers, women say things they never say in court, for example, disclosing sexual assault. That whole dimension is being ignored.

Amy Holtzworth-Munroe said we need to look at risk factors, and whether or not men are going back to communities (after attending a BIP) with those same risk factors: e.g., alcohol, violent or aggressive peers, general community violence.

Radhia Jaaber responded by saying even if a batterer is in a program, he still never left that community so we need to know how BIPs relate to men within their communities.

David Adams mentioned that he thought BIPs have become overly dependent on the courts as a source of referrals. Without sufficient community support, BIPs must often bend to the wishes of the courts, which are sometimes more interested in facilitating case flow than batterer accountability.

Oliver Williams reported that in our African-American, culturally focused BIP group they have the men thinking about their communities and what promotes violence. We can look at batterers who have changed, and find out what did it for them.

Barbara Hart saw some research that showed that when women's risk was assessed, only 4 percent of women who were thought to be at high risk were actually beaten again within 4 to 6 months. But 19 percent of those high-risk women were sexually assaulted. We need to ask women to what extent they are under scrutiny or surveillance by their batterer? How much disruptive behavior do they experience by him in the family? Does he keep his promises? Does he undermine her authority with the kids? Does he discredit her? Does he use the children to facilitate his disruptive behavior/surveillance/coercion?

Richard Titus asked whether there are other sources beyond the women themselves from whom we could get this information, or confirm what she says?

Oliver Williams mentioned that different cultures have different manifestations of sexism. Women and men can tell you about these cultural differences. You need to get a sense of what the culture/community "rules" about sexism are so you can understand the broader

outcomes we are talking about. Also, get the men to describe their definitions of womanhood and manhood and use this information to identify outcome measures.

Chris Eckhardt noted that we need qualitative and quantitative measures of all this. But ultimately we are trying to find risk factors to prevent harm to another person. We need to understand individual or additive risk factors, specifically, to prevent violence. We cannot be too general.

Dan O'Leary noted that in *Violence & Victims* there is a review of seven measures for assessing psychological aggression--things like taking her check book, removing spark plugs from the car, restricting her phone use. We do not have measures of sexual aggression used in any systematic way. The new CTS has too few items to be internally consistent. We need to borrow more extensively from the sex abuse community. I do not think we have a good measure yet for assessing fear of partner. We need a more extensive measure of fear, safety and sexual abuse, not just one or two measures.

Andy Klein noted that there are two kinds of deterrence: specific and general. Do BIPs provide general deterrence? If they are like drunk driving programs, they may be better at general deterrence than specific deterrence. My guess is no, but it's an important issue.

Amy Holtzworth-Munroe reported that she used Koss's 13 items on sexual abuse and compared them to the 6 items on the CTS II. "I liked the Koss items better. What is it that makes sex coercive? Is it fear? Previous violence? We have to be careful about what we call coercive/controlling sex. 'I had sex because he wanted it and I'm his wife' may seem controlling to us, but not to the woman. We also need to know if guys who complete BIPs are handling day-to-day situations better."

Chris Eckhardt said we do not know if we should measure coercion by what was said or done or by how the woman felt about or interpreted it.

Ed Gondolf cautioned against instruments getting too long as more and more items are included, and worried that some items may not relate to women's real lives. In this case, a woman may minimize her disclosure or we may not capture the complexity of her experience. We need a qualitative interviewing approach. We have been short on that so far. We need to tap some domains we have not been able to get at with our instrumentation.

Dan O'Leary added that we know enough about psychological aggression and coercion. We have internally reliable scales that are good predictors, e.g., Tolman's 14-item scale. But the CTS II is not enough to gauge sexual coercion, although we can begin to collect data on psychological aggression.

Amy Holtzworth-Munroe responded by saying that those measures are good for continuous measures, but we do not have cutoffs to tell us who is risky.

Diane Rosenfeld said that we need better scales for measuring sexual coercion. For instance, a colleague of mine asked her subjects if their husbands used pornography and got a 75 percent-positive rate.

Andy Klein mentioned that most studies do not get victim report data, but instead rely on official data. But official reports might not include some offenses, for example, civil or

municipal offenses, which do not go on criminal records. You have to know the codes and the arrest and charge rates to understand what the official recidivism rates mean. Low recidivism may speak to low arrest rates more than program success.

Ed Gondolf added that arrest records are problematic from State to State and even within jurisdictions and over time because police behaviors change or crimes are re-classified, etc. Arrest records may be very unreliable. We need to look at a situational model that includes ongoing risk, community, and contextual factors and additional services and interventions the guy may be exposed to. Not just a single-point behavioral measure. And you need all this to be longitudinal. This is difficult and expensive. NIJ has to assess whether it can fund these big longitudinal studies or whether more, smaller studies could be combined in meta-analysis.

Kevin Hamberger said he has never seen a BIP outcome study reporting what men have learned, how they have changed, what skills they acquired. We need to back up and demonstrate they are learning the things we think we are teaching them.

Ed Gondolf retorted that they did that. Ed and his colleagues asked men what they thought they had learned about asking for things, managing conflict, etc., and then asked their partners if the men were using the techniques they reported they had learned. The men did name specific skills taught to them in the BIP.

Amy Holtzworth-Munroe asked whether you could set up role plays, etc., to test how a man would deal with situations post-treatment?

Margaret Zahn noted that Terrie Moffit, Terry Thornberry, and Rolf Loeber each have done longitudinal studies of violent people. Their subjects are now old enough to be married, so the timing is good to do some add-ons on domestic violence. Add-ons would be less expensive and faster than starting again with a new study.

Sally Hillsman mentioned that results are better in a high-risk population than general age cohorts.

Margaret Zahn responded by asking whether we could isolate subjects in these studies who have been through BIPs and study them? These subjects are very cooperative because they have established relationships with the researchers over time.

Amy Holtzworth-Munroe noted three types of studies: 1) BIP studies that have problems with attrition; 2) BIP studies that have no problems with attrition; and 3) longitudinal studies like Terrie Moffit's and mine.

David Adams mentioned a problem with measuring skills. He conceives of battering as a skill set, not a skill deficit. And batterers may co-opt skills they learn in BIPs as part of their battering repertoire. They may incorporate them into their controlling/coercion techniques. He cited research that said *yielding* behaviors on the part of the husband were a better predictor of long-term successful relationships than *communication* skills. This says that maybe we should be teaching humility, not communication/negotiation.

Kevin Hamberger said we should be able to measure humility in behavior. If men are co-opting conflict-solving skills into controlling behavior we should measure that, so that we can change what we are doing in the BIPs.

Andy Klein noted that short-term studies will not tell you whether or not men simply become less violent because they have moved on to new, more malleable partners. While some studies track down victims, few contact new partners, and those that do rely on the batterers to inform them who the new partners are!

Working Lunch—Enhancing Batterer Intervention Programs: The Culturally Competent Batterer Intervention Program

Oliver Williams

Dr. Williams focused on programmatic aspects of batterer intervention. He encouraged BIP counselors to think about influences that shape batterer's lives so that they truly know and understand the people they are working with.

Back to Top

Evaluation of Recruitment and Retention

Opening Remarks by Chris Eckhardt and Andy Klein

Dr. Eckhardt began his presentation by noting that the study outcomes discussed so far have all been tied to recruitment strategies. If outcomes are tied to the percent of batterers who actually complete the program, then we must consider what are the proper elements of a program to examine. Dr. Eckhardt suggested that we need to be looking at who is recruited. Where are they recruited? What is the total number of batterers compared with the sample size? Sixty-five percent of adjudicated cases actually go to a BIP. What is the external validity here? External validity is a real issue. He also noted that the issue of "whom we recruit" keeps expanding. He discussed several options for recruiting: 1) from the BIPs itself, which offers the advantage of gaining access to partners; and 2) from the courts or probationary offices, however the disadvantage of the latter is that it makes it very difficult to gain partner contact. We also need to be looking at the issue of intention to treat vs. completers. It is also hard to contact partners; cold calling is not very effective. Also important to consider is who provides the contact information--there is a real ethical issue here. And we should be looking at who does the screening? He added that we should also consider the extent to which the batterers in the programs are the same as the batterers who are adjudicated--and what is the drop-out?

In response to the "whom we recruit" issue, Dr. Eckhardt offered several different strategies. First, we should consider the idea of broadening recruitment because of the large gap between the numbers in the court system and the numbers in BIPs. Let's work with the courts and probation under the self-change hypothesis. We should figure out a way to work with the courts and probation to include a broader group in the programs. The second strategy he discussed was recruitment narrowing. With this approach, he said that we should engage in a screening process and target the treatment toward the diagnosis: Those with substance abuse problems should have batterer programs especially geared for them and those with emotional problems should have programs especially geared for them. We also need to be looking at what works, for whom, and under what circumstances. Police screen even before these guys get to court.

Andy Klein commented that there are two big issues in the area of retention and recruitment. First, we really need to broaden our definition of victim from the victim of the original abuse to new partners that may also experience abuse by the batterer. Studies

suggest that many batterers are serial abusers. Second, if our goal is to protect victims, we need to broaden our definition of program success. He suggested it is backwards to look at program effectiveness by looking at completers compared with noncompleters. Our goal should not be graduating as many abusers as possible through programs. The research, even the research that shows no treatment effect, agrees that programs identify high-risk abusers--those most likely to continue to abuse. These include those who are referred to treatment but do not attend or complete programs. They are more likely to re-abuse. Therefore, the identification of noncompleters is a crucial program function. Of course, once high-risk batterers are identified by the programs, the criminal justice system must use this information appropriately. In studying batter programs, therefore, we must ask, "What happens when the guy flunks out?" Dr. Klein added that batterer programs are not magic black boxes. They are part of a "forced treatment" system that depends on the "force" as much as the "treatment." Therefore, any BIP study must look at both aspects.

Back to Top

Comments were made by the following participants in response to Dr. Eckhardt's and Dr. Klein's remarks

David Adams noted that researchers sometimes seem exploitative, rather than collaborative with BIPs. For example, he has sometimes been asked by researchers to write letters of collaboration, only to find that the researchers have no wish to collaborate once the grant is awarded.

Rob Davis mentioned that getting a hold of a victim 6 months after an incident happened you find that 15 to 20 percent of the police/court records you are using for her are incorrect; and another 15 to 20 percent have moved, disconnected their phone, etc.

Amy Holtzworth-Munroe retorted that Chris Sullivan, Ernie Jouriles, and others have done well following women.

Dan O'Leary noted that Ernie [Jouriles] is working with moms who have a potentially abused child. This is a different problem than trying to find the partner--or worse yet, new partner--of an abuser.

Ileana Arias asked about Linda Marshall? Linda Marshall has been fairly successful in following a large cohort of women over a period of several years.

Oliver Williams added, Rachel Rodriguez, Julia Perilla? Dr. Williams asked whether there is a way to provide a benefit to the community you are studying so we do not have to take without giving back. Why should the community trust us? Trust has to be established. I try to develop relationships with community-based organizations. They sometimes think my evaluations of their program are for me, not for them.

Margaret Zahn asked whether you have to share race, age, etc., with your subjects to gain their confidence?

Oliver Williams responded by saying he does not know that you have to be like them, but you must demonstrate a competency in their culture and engage them in the research. You do need research partners, interviewers, etc., who are "like" them, for example, using female interviewers with battered women.

Ed Gondolf asked, "At what point do you ask for informed consent from a batterer?" Informed consent is a big issue here that does not get discussed very much. But Institutional Review Boards (IRBs) vary tremendously in what they approve, which has serious implications for this kind of research. This discussion needs to be integrated into the BIP discussion and the field would like some guidance from the feds on IRB issues. If we integrate the informed consent process into the program orientation, we will get much higher recruitment rates. Dr. Gondolf also noted that the question of how much can we compensate research participants is also an issue because, in the medical field, paying more than $30 is considered coercion.

Amy Holtzworth-Munroe stated that she doesn't believe you can ever pay subjects too much for participating because many of the populations we deal with are extremely poor. I know not all IRBs agree with that perspective. Sometimes we hound the participants. We go out to their world. Some may consider this coercive; we consider it something we have to do. There is also the safety issue, the costs involved, and the safety of the interviewers. Ernie Jouriles actually goes to women's homes and he has a nice retention rate.

Etiony Aldarondo said that batterers do not trust researchers. Research feels coercive to them. They are convinced the interviewers are sent by the courts, rather than being independent--even if you tell them otherwise.

Judge Larry Hauser agreed that we need to build up trust among the people working with batterers. Judge Hauser noted that his group meets regularly: the victim advocate, cop, parole officer, etc. The victim advocate should be independent from the criminal justice system. Advocates should be, for example, from the crisis center, not the prosecutor's office. Advocates could help with the recruitment of victims also. The special domestic violence dockets are not defendant friendly. Courts have to show accountability, for example, for due process, and then we can bring in the defense bar. The court does essentially wraparound services. The court should be a resource, not an intrusion.

Oliver Williams agreed that we need to be developing relationships with communities. The University of Minnesota has the BIP academy graduates go into the communities and teach batterers. We must give back to the community. Trust is established by building relationships with agencies, by humanizing each other. We have coffee and Pringles together; that is, we are sociable with each other.

Amy Holtzworth-Munroe stated that there are pros and cons associated with collaboration. The more involved you are, the more *you* change the study and your outcomes. We could use an advisory panel.

Sally Hillsman mentioned an issue that does not get discussed much. That is, universities do not train people in this kind of work.

Back to Top

Evaluation Designs

Opening Remarks by Ed Gondolf and Kevin Hamberger

Kevin Hamberger began this session by saying that there are several issues regarding experimental designs that he wanted to put on the table: threats to design, attrition,

validity of variables, the need for process evaluation, and control/comparison groups. Regarding quasi-experimental designs, Dr. Hamberger noted that there is the issue of dropouts and completers--which is not good. You can get close to what you want with quasi-experimental designs. The intention-to-treat is a real problem. Dropouts neutralize the effect, for example, if you only have a 30 percent completion rate. There is no dose vs. treatment.

Ed Gondolf suggested several alternatives to experimental designs. 1) You can use a dose-response model. People use experiments to control for treatment vs. no treatment, which is essentially a quasi-experimental design. 2) We have to consider logistics. Dropouts confound the results and are also related to other variables, which is another problem. 3) We have to consider context. Experimental designs presume there is no context, which is absurd. 4) We have to consider the systems in which these programs take place. What is the comparison group outside of treatment--it doesn't make sense. 5) Finally, there are analytical tools that enable modeling such as instrumental variable analysis and/or propensity score analysis.

Back to Top

Comments were made by the following participants in response to Dr. Hamberger's and Dr. Gondolf's remarks

Barbara Hart mentioned that giving women information on community resources changes women, so researchers do not give women this information, which is awful. This should be discussed. Batterered women's services and BIP connections are lacking. Recruitment should be easier if these systems were somehow connected. BIP evaluations cannot be at the expense of women's safety.

Dr. Adams notes that program completion rates are influenced by the strength of the court response, not just by the effectiveness of the BIP. EMERGE, for example, deals with 20 different courts in Greater Boston, and program completion rates appear to vary according to the degree of court monitoring, as well as court reinforcement for program goals. He adds that sometimes our decision to keep someone in the program or terminate him depends on whether the referring court is likely to let them off with no penalty. Dr. Adams adds that recidivism is also related to other factors. For instance, many BIPs have extensive contact with victims, and this contact appears to make victims more more sensitized to abuse, as well as more educated about their rights. These victims who are more sensitized might be more likely to report new acts of abuse. Even though he thinks of this as a positive outcome of sorts, this could be misconstrued by some researchers as a program failure.

Larry Hauser added that pure experimental designs are very difficult if not impossible in a court setting. Defense lawyers will not go along with it.

Sally Hillsman further added that we have an imperfect "delivery system" compared to a drug trial, for instance. We need to know does the pill work and can the real world deliver the pill. Absent from our research is a thoughtful theoretical base for why we think an intervention will work. We need to link logical theories to why we think interventions will work, and then we would better understand what we are learning.

Etiony Aldarondo agreed that we do not have empirically/theoretically driven benchmarks we can work with. We can only work with one program at a time to make that program

better. But we are not at a stage where we can decontextualize that program and say anything about "what works" globally.

Rob Davis suggested that the experimental designs we have done so far have not been perfect; and there is a role for quasi-experimental designs when experimental designs are impossible to implement. But under some conditions you can do experimental designs and they are the gold standard.

Sally Hillsman added that if you are going to do experiments, you want to be sure what you are studying is also the programmatic gold standard. Does theory say the program will work?

Dan O'Leary noted that Jacobsen's, Leonard's, and his research, which follows the progress of, but does not treat, violent men finds that 60 to 80 percent of the men continue to be violent. Julia says two-thirds of people in the Navy study's no-treatment control group improved. "Steven Beech and I got approved for a wait-list control group, with emergency phone contact if necessary, for a study of suicidal people. If the subjects used the emergency phone contact more than X number of hours, we did not count them as no-treatment controls. This might be possible for women assigned to lesser or alternative treatments. You could work with people who volunteer for services. Admittedly, their situation is less severe, but they are easier to track, and you might be able to document some effects."

David Adams asked what is the program vs. the Gestalt, that is, the broader system in which the program resides? Hopefully, victims become more sensitized to violence through participation in the evaluation.

There are two different things we are doing: We are measuring a social process and we are asking whether a particular kind of treatment works. There is likely imperfect delivery of the program, recruitment, etc.

Dan O'Leary reiterated that what we have here is an absence of a theoretical base for uncovering why treatment is going to work. We desperately need research on the theoretical underpinnings of BIPs. There is a large literature on aggression that researchers could draw on.

Etiony Aldarondo reported what works in his program, but not BIPs generally, using experimental designs. But to do experimental designs, we need a program that is based on theory and research.

Barbara Hart suggested that communities should get to help in designing research. Of course communities/courts resist no-treatment control groups because of concern that women will be further endangered. Controlled experiments seem to be indifferent to the safety of women. Science should not impose system changes that practitioners think might be dangerous. It might help by finding communities where BIPs and battered women's programs have good connections with each other.

Dan O'Leary mentioned again that they successfully used a wait-list model with emergency treatment. If the participant used the emergency phone consultation, then he was no longer in the control group. Eighty-five percent of the participants remained on the wait list. We

could develop a design in which we would assign batterers to a BIP or a lesser treatment. We could also use volunteers for services.

Ed Gondolf noted that although there are no treatment differences between types of programs, many men do exhibit less violence in treatment. Two-thirds of those who do not get treatment reabuse, and two-thirds who are in treatment do stop abusing their partner.

Back to Top

Evaluation Implementation

Opening remarks provided by Rob Davis and Larry Hauser

Judge Larry Hauser provided an analogy: Doctors did not know aspirin helped reduce the risk of heart attacks. We have an ethical obligation to see if BIPs work.

Judge Hauser referred to the study that Rob Davis is undertaking and commented that it is important to compare the effectiveness of court monitoring with court control. He went on to ask whether we should be applying the maximum amount of pressure? There is more we could do between post-plea and pre-sentencing. What we do not know is what accounts for the change in batterers: Violence suppression or behavior change?

Judge Hauser commented on ways to build trust. He noted that they hold weekly meetings that everyone but he attends: as the Judge, he has to remain impartial. Meeting attendees include police, prosecutors, probation, service advocates. They review the cases for that week. They established trust with the defense bar by sharing information--sharing information with the right person to the right program--not indiscriminate sharing of information. He added that the defense bar really likes this. It wasn't easy and took about 2 ½ years.

Back to Top

Comments were made by the following participants in response to Rob Davis' and Larry Hauser's remarks

After being asked to build that trust, **Judge Hauser** replied that there should be weekly meetings of all the players, including cops, advocates, parole/probation supervisors, treatment providers, etc. These regular meetings should be used to monitor men's progress. This will keep women safer. Also, having advocates active in the process can "grease the skids" for the researchers, by making victims more accessible. It should also be noted that the defense bar needs to see that there is due process and accountability if they are to be more likely to accept research.

Oliver Williams asked a question of Judge Hauser: "In Connecticut you've invested in providing wraparound services. I wonder what the community thinks of this?" Dr. Williams went on to say that he thinks if people think you're a resource and not an intrusion, they will be more likely to participate in research. People are concerned about the consequences of research. We have to make communities feel like partners in research, where academics go into communities and are taught by communities. Communities have to see what the benefits are for them--their agenda may not mesh with our academic agenda. We need to be invested, long-term, in the community.

Kevin Hamberger added that researchers need to develop trusting relationships with agencies--shelter, probation, etc.--and make the research useful to them with minimal risks. This will enhance recruitment and retention.

Amy Holtzworth-Munroe cautioned that there is an ethical issue here: You cannot become too involved with your subjects or you may change the condition you are studying. You are trying to study the process without intervening. It is hard to track how much we are changing the process by keeping tabs on people and building relationships with them.

Ed Gondolf added that our own human biases complicate the research and affect our results. We have to weigh the benefits of being part of a collaboration and be able to maintain the objectivity of the work. We do this by using a scientific panel that is not involved with the research to review the work to make sure our biases are not coming out. Dr. Gondolf went on to say that quasi-experiments have earned a bad reputation for falling short of the gold standard, when they are trying to achieve the same goals as the experimental designs. Quasi-experimental designs are easier, and more realistic, and you get closer to what you are trying to do in real life. Experimental designs test treatment compared with no-treatment, but what we really want to know is how much do the BIPs add? Dropouts do not receive the full treatment, and therefore cloud the results. This requires a dose-response analysis. Logistic regression that accounts for dose response is a naive type of analysis and is confounded because dropping out is linked to other variables you are controlling for. There is also a problem with experimental designs and context. Experimental designs assume the BIP is independent of its context. Quasi-experimental designs allow you to account for context. Instrumental variable analysis allows you to account for complexity, also propensity score analysis allows you to create in the data comparison and control groups. It is becoming harder and harder to get a pure control group: systems resent the intrusion, and the no-treatment folks seek out other help, such as extra counseling for the victim, alcohol treatment, etc.

Ed Gondolf suggested that we need to be looking at all the components of the system, not just the program alone. This really is a systems issue. We know that 20 percent of the men reoffend and continue to reoffend. What we do not know is how to identify and contain these men. But most guys do better after going through a program.

Margaret Zahn noted the issue of desistance: Older individuals commit less crime--perhaps because of maturation, either biological or psychological--and being employed predicts use of less violence.

David Adams asserted his preference for longer programs. In short programs, the batterer sometimes is merely "holding his breath" until he completes the program. Longer programs create more opportunities for batterers to distinguish the program for the court. One technique used by EMERGE is having group members do relationship histories in which they give information about each intimate relationship they've had. This exercise helps batterers to recognize long-term patterns of abuse and, therefore, to develop more of an internalized motivation to change. Longer programs also help batterers to appreciate how self-defeating their behavior is. A longer program also has the added benefit of providing monitoring over a longer period of time. By seeing clients every week, BIPs provide more intensive monitoring than is possible by the probation officer, who typically sees the probationer only once per month.

Rob Davis interjected that comparisons allow you to say something concrete, in contrast to trend lines, which only show you generalities.

Dan O'Leary wanted to know what is an alternative to treatment.

Andy Klein responded that measuring physical violence alone may be misleading. Although breaking someone's jaw is different than threatening to break her jaw, once you have broken your partner's jaw, you don't have to do it again, just the threat may intimidate, control, and terrorize your partner. Reductions in violence alone are not valid measures of behavior change.

Barbara Hart agreed completely. Maturation, physical violence vs. coercive control. We should look at outcomes other than desistance. How about including in treatment programs men's relationships with their children? Can we change batterers by getting them to focus on fatherhood?

David Adams responded by saying that they have the fathers in their program look at the effects of abuse on their children. The guys are more empathetic when we discuss the impact of violence on their children.

Barbara Hart asserted that men should want their female partners to walk in power so their kids do better. Kids who have strong mothers have kids who are better adjusted.

Ed Gondolf asked a public policy question that he felt the feds should be considering. We are always quoted using little sound bites. When the media call us and ask us whether BIPs are working, what should we say? Is there a consensus statement? [The audience was silent--no response and the issue was tabled.]

Back to Top

Recommendations

Each participant gave a one or two sentence summary of their recommendations for which direction the field should be headed.

Barbara Hart said that performance evaluation for program improvement is essential.

Chris Eckardt noted that some people are too confident that we have all the answers. We have to love uncertainty or we will not get anywhere. The other important area is matching a client with a specific treatment: a treatment algorithm.

David Adams said that we still do not know what we are learning and suggested an approach of interviewing batterers who are known to have changed, to see what they've learned. Collaboration is important: Practitioners have a lot of ideas and they need to know which aspects of the program are working. We need evaluation. We need to know what the batterers are really learning? We need to look at more than just outcomes. How are the victims affected? What happens with the system, both judicial and prosecutorial? We need to look at the actual practices of judges and prosecutors, not at just what they say they are doing.

Diane Rosenfeld noted that there is a question of political will here. Do we want to stop battering? England has stopped funding BIPs because it does not think they are effective. Diane proposed a Detention Facility that would include: active judicial involvement and oversight, programs, and community victim services. We need to adjust the institutional

response to battering. [Two of Diane's papers were distributed to the participants at the meeting.]

Andy Klein asserted that a 70 percent success rate is not good. He suggested that we need to look at BIPs compared with other alternatives. What do we do with the 30 percent of batterers who do not successfully make it?

Dan O'Leary suggested that we need to be looking at the persistent nature of violence, from childhood and adolescent through adulthood. We should also be focusing on alternatives to BIPs. We also need to know what is accounting for change: psychological change vs. change because of monitoring. [There was a comment made about class-action suits regarding BIPs.]

Larry Hauser mentioned his quandary in that recidivism is high among this population and judges have to defend BIPs, but this is also a resource issue. We need to build trusting relationships between the criminal justice system and researchers. And we need to pay attention to recruitment and retention issues for rigorous evaluations.

Shannon Morrison indicated that we should be looking at other programs to see what has worked for them. We should also be assessing a program's ability to be evaluated and include process evaluation and theory in the model.

Radhia Jaaber suggested the need to look at the culture of research and have an exchange of information and ideas. We need to respect who they are. Also, this is a young field and we cannot forget that. Let's look at what groups BIPs work for: the criminal justice system, advocates, and communities. What is our social consciousness here?

Kevin Hamburger asserted that process research is really needed. We need to know what our clients are learning. We only look at cessation and desistance, but there are other things to look at. We also need to talk to successful batterers who have changed and learn how they did it.

Bill Riley reminded us that there are some programs that are not called BIPs, but the community works with these programs. We need to pull all these nuggets together and call them something else.

Jerry Silverman invited participants to think about what kinds of programs make sense? There is money from the courts to go through prescribed programs. We also need to look at how people change. We need to embed men in systems that are functional. Men change when they are a part of something. But which components of the program lead to change? Learning compassion, for example, may help men and it may spill over into community involvement and parenting. What elements of the program are resulting in change?

Linda Melgren informed us that she works at HHS in fatherhood programs and these programs are not hooked up with BIPs. How can we address violence in these other programs? Should we develop programs between BIPs and fatherhood programs, for example, offering peer support. We do not measure domestic violence in the fatherhood programs. We need more conversations between these programs.

Ed Gondolf concluded that these conversations reinforce how difficult it is to do evaluation research. He believes in action research. Dr. Gondolf finds real treatment effects in his

research, and he is confident that change is due to treatment and not some other variable. But what we need to know is how to help the less effective programs work better. Our research suggests that treatment matching is necessary for unresponsive batterers. We have to predict in research, but we have no answers--no impact. To stop some men from reoffending, we may need containment measures rather than treatment.

Julia Babcock reiterated that current program treatments have small effects. BIPs are having an effect, but it is small. She reminds us that all hope is not lost, however. Chris Murphy's study found large effects and Waldo found relationship enhancement. These studies need replication, but it is a start. We need to challenge clinicians to look at a number of outcomes, including violence, psychological adjustment, accountability, and women's feelings of safety. We need more basic research and must develop treatment programs based on the basic research.

Etiony Aldarando noted that peace is not the opposite of violence. Justice and equality are harder to attain than peace. BIPs work because they promote equality and justice. He further asserted that we do not really know why they work. We do need to improve programs. Perhaps we could identify programs that exceed a 60 percent-success rate and study how they work. We should also devote time to evaluating programs for African-American men and programs with integrated programs. Dr. Aldarondo asserted that BIPs do work better than aspirin.

Back to Top

Closing Remarks

Margaret Zahn

Dr. Zahn suggested that maybe we need a small booklet on best practices. She further reiterated the need to develop theoretical models. There are three theories that this field could build upon: learning theory (we know fear is a good motivator), violence theory (general violence), and power and control, which theorizes about such topics as victim selection.

Back to Top

Date Modified: July 6, 2011