MARTI
Mid-Atlantic Addiction Research and Training Institute
Edward Gondolf, EdD, MPH; Research Director
1098 Oakland Avenue
Indiana University of Pennsylvania (IUP), Indiana, PA 15705

OFF: 724-357-4749
FAX: 724-357-3944
E-Mail: EGONDOLF
@IUP.EDU

Nov. 11, 2005

Michael Rempel, Research Director
Center for Court Innovation
520 8th Avenue
18th floor
New York, NY 10018

RE: Rempel, M., Labriola, M., & Davis, R. (2008). Does judicial monitoring deter domestic violence recidivism? Results of a quasi-experimental comparison in the Bronx. *Violence Against Women, 14*,185-2007.

Dear Michael,

I have been meaning to write or call about some questions that have been posed to me about the Bronx BIP study. A couple of people recently emailed me asking whether your study sufficiently verifies that batterer programs "don't work". These questions come in response to the letter from Hon. Judy Kluger accompanying a summary of your study. The letter cautions judges that "batterer intervention programs should not be relied upon to change offenders' behavior." The inference seems to be that judges should not refer domestic violence offenders to batterer programs because there is no additive effect over open probation. Similarly, you conclude in your ASC presentation and the summary report on the study: "Programs using popular educational and/or cognitive-behavioral models do not reduce future domestic violence." And you recommend instead community service as a suitable alternative. It seems important to recheck some issues with you given the study's apparent impact on policy and practice.

Implementation Issues

Your Court study, as I see it, makes an important contribution, as we've discussed. But a few implementation problems compromise the results and warrant qualifications. I gather your position is that the implementation problems are countered or not substantial, and that the results are substantiated by the other experimental studies (which I would claim have serious limitations of their own). I know we talked about several of these issues at the DC meeting, and perhaps you have sufficiently addressed them.

1) The outcomes of “no effect” hinge primarily on the police arrest records. The catch, as I see it, is that the base rate of arrest records for domestic violence tend to be relatively low (they were approximately 13% in our 4-year multi-site follow-up), and they are complicated by a variety of circumstances. For instance, the women with program dropouts, in our study, were less likely to continue with their partners and have less access to them; some professed discouragement that the CJS doesn’t do anything. Thus, fewer calls to the police in the follow-up. Women with men in the program, on the other hand, had higher expectations and were more likely to call the police in response to the men’s violation of the program requirements or continued abuse. Therefore, more police calls were made by the women with program participants despite an actual decrease in abuse incidents. We found also, like your study, that the arrest rate for other crimes remains high and does not appear to be affected by the domestic violence intervention. This finding seems to reflect the criminality among a substantial portion of the men that warrants more comprehensive and extensive intervention.

2) As you point out the actual incidence report by victims is substantially higher and more reliable than arrest records. An important issue in your study is, therefore, the small follow-up response rate of approximately 25 men per option and the short duration of the follow-up period. There is of course valuable information in that follow-up, but it seems hard in my mind to draw a conclusive recommendation about the program effectiveness based on such a low response rate. The small sample size makes the outcome vulnerable to confounding influences. The responding women across the samples are not necessarily equivalent (maybe you have checked on that) in terms of their characteristics and circumstances. Again, the women respondents with partners in batterer programs may be the women with a gripe because the man dropped out, or more likely to be living with their partners. The women with no treatment may be more likely to have sought and received compensating intervention and services, like a protection order or victim services. Overall, there may be related factors in the dynamics of the follow-up that are likely to vary in the different groups as a result of differential system reactions that influence the outcomes. A small response group would be, it seems to me, very sensitive to any of these differences. All you need is one or two cases in another direction to get a substantial percentage difference, and you would need a lot of cases to get the power to detect significant differences. Interestingly, the current percentage differences, although small, are in the direction favoring the batterer programs.

3) The main issue in my mind is still the “intention-to-treat” assumption in the experimental designs comparing batterer programs to “no treatment.” The research on the drug courts and domestic violence courts is pretty clear that program attendance is sensitive to compliance oversight and sanctions. The question then becomes what effect does <u>attending</u> the program have and how much does attendance matter—this becomes, a question of dose-response. The fairly substantial dropout rate of 45% in your study helps to offset the program effect being tested with the treatment group (including completers and dropouts) and the non-treatment group. I continue to be impressed by the moderate program effect detected in the instrumental variable analysis, and confirming propensity score analysis in our study (along with a deterrence analysis and an “attribution of change” study with our data). (An article in the current Journal of Experimental Criminology

endorses Instrumental Variable analysis as an alternative approach to experimental approaches that face implementation problems.) Other indicators of success like the women's perceptions of change and safety seem highly important as well, especially since these correlate so high with future abuse and well-being. The women's higher satisfaction with the batterer programs, that we also found, may say something in this direction, as well as decreases in other kinds of abuse and increases in positive behaviors of the men over time.

Again, my question is not with the study and its importance, but with its qualifications and interpretation. Should it be used to direct court policy? I'd say we certainly should reinforce longstanding cautions about simplistic notions of intervention (e.g., "just send all the men to any batterer counseling"), but we are not to the point where we have an "evidence-based" rejection of batterer counseling programs. Yet it seems like that is where we are headed based on the judge's letter and your preliminary report.

My sense is that the judge's statement is literally true but perhaps misleading. That is, batterer programs have never in and of themselves been sufficient to change the behavior of all the men sent to them. There is, to my knowledge, no intervention in the CJS system that does. Do batterer programs reinforce, extend, or initiate change in a substantial portion of men, compared to doing nothing? I say "yes" for a substantial portion of men (in some programs), given the dose-response analyses.

The issue that our research points to is the need to identify and contain the repetitive or chronic abusers that are unresponsive to intervention in general. As your arrest rates, and ours arrest rates, remind us, there is a core of severely violent men with criminal histories that have failed in previous interventions and are not good candidates in this one---batterer counseling. This of course begs the question about risk assessment and ultimately the need for on-going risk management. Moreover, we've found that further re-assaults tend to be overlooked or not reported to the CJS. When they are, the containment, supervision, or sanctions are generally not increased for the re-offense. As a result, batterer programs are not sufficient to change SOME men's behavior. But I'd say that the intervention process as a whole (police, courts, batterer program, and follow-up services) are associated with change in the vast majority of men over time, as our 4-year trend analysis suggests.

Conclusion drawing on previous research

Your conclusion seems to be that your study adds confirmation to the previous experimental studies that show little or no effect for batterer programs: "Batterer programs have not demonstrated success at rehabilitating batterers. Given what we know about batterer programs, it makes sense to think about experimenting with new options…" I would say that we have limited, ambiguous and contradictory findings, and that it is just as easy to argue that the fundamentals of batterer programs need to be better implemented and built on. (What follows is the counter research that I've compiled for another presentation.) It is worth noting that cognitive behavioral approaches (CBT), which I would argue comprise the mainstream of batterer programs, are recommended and deemed relatively effective with other offenders. Treatment manuals have recommended them for individuals with antisocial and/or narcissistic tendencies (e.g., Choca and Van Denburg, 1997), and a current review of treatment approaches for violent offenders (Elliott, 2005) and a review of sex offender treatment (Marshall & Serran, 2000) conclude that CBT is more effective than other approaches with these populations. Also, the reviews of drug courts research generally conclude that coerced treatment is more effective than voluntary treatment or no treatment (Wilson et al., 2005).

You mentioned in your previous email: “since our results on the batterer programs echo the other experiments, we'll still probably write a strong conclusion of no effect.” I’ve see this line of thought frequently at conferences and in papers. My impression is that this is building on sand rather than solid ground, and also ignoring alternative research. The NIJ special report on the batterer program experiments (Jackson et al., 2003), two of our published critiques (Gondolf, 2001, 2003), the recent DOD review (Saunders, 2005), and a CDC review (Morrison et al., 2003) all point out the limitations of the experimental studies and raise cautions about their interpretation.

Using an evidence-based research guide (Briss et al., 2000, p.4), the CDC review concluded: “The diversity of data, coupled with the relatively small number of studies that met the inclusion criteria for the evidence-based review, precludes a rigorous, quantitative synthesis of the findings. However, the rudimentary analytical strategy used suggests that the majority of BIP studies reported positive intervention effects for behavioral (i.e., re-assault) and psychosocial outcomes for at least on follow-up period.”

Eckhardt et al. (2005) gives some more specifics in their review of the experiments: “Careful review of these experimental studies further indicates that they fall short in important ways from the state-of-the-art RCT (random clinical trial) methodology outlined above, most notably in not demonstrating treatment adherence and therapist competence, providing inadequate specification of interventions, having low partner contact rates, and/or having high levels of attrition from the treatment and research protocols.”

We add the fundamental issue of the “intention-to-treat” assumptions (compounded by high dropout rates), what our statisticians call “naïve” statistical analyses, and the external validity questions raised by the context of the intervention system (Gondolf, 2001). (In acknowledging these issues, a recent article by Berk [2005] redefines experimental designs in criminal justice research as the “bronze standard” instead of the “gold standard.”)

These sorts of conclusions also raise questions about the two published meta-analyses. The Babcock meta-analysis itself cites five major caveats that qualify its results, and meta-analytic proponents like Lipsey and Cohen warn about over interpreting or misapplying meta-analyses (and particularly the effect sizes) under these conditions. A recent article on research and policy for child services concluded its critique of the misuses of meta-analysis: “In a brilliant essay, Jacob Cohen (1990) reflected on the statistical lessons he has learned and offered the following advice: ‘Finally, I have learned that there is no royal road to statistical indication, that the informed judgment of the investigator is the crucial element in the interpretation of the data, and that things take time’ (page 1304). Let us use our best judgment when we bring research to bear on policy questions—and, when we do, let us take the time to evaluate effect sizes in context." (McCartney & Rosenthal, 2000, p. 179).

There are moreover several lists of criteria for achieving evidence-based treatment or dismissing it, and the current batterer evaluations fall far short of these (e.g., Briss et al., 2000; Chambless & Hollon, 1998; Eliott, 2005; Lipsey, 2005).

Please excuse this long-winded wondering. It is an extension of thoughts and issues that have surfaced in wrestling with the "do batterer programs work?" question in general. They come out of my sense that we need to keep the door open on batterer programs at this point (although rightfully raise some cautions), and interpret their contribution in a broader context.

Sincerely,

Ed

Dr. Edward W. Gondolf, Research Director
Mid-Atlantic Research and Training Institute (MARTI)
Indiana University of Pennsylvania, Indiana, PA 15705