Emotion-focused therapy for incarcerated offenders of intimate partner violence: A 3-year outcome using a new whole-sample matching method

Article *in* Psychotherapy Research · May 2011

DOI: 10.1080/10503307.2011.572092 · Source: PubMed

CITATIONS
22

READS
412

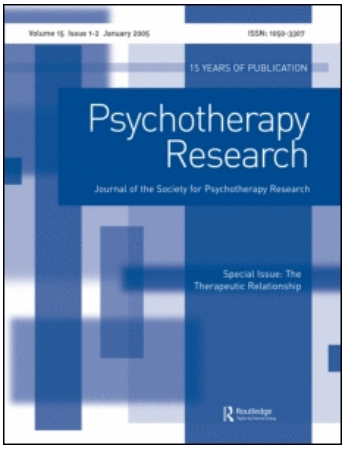
4 authors, including:

Antonio Pascual-Leone
University of Windsor
61 PUBLICATIONS **940** CITATIONS

SEE PROFILE

Robert Arnold
University of Windsor
12 PUBLICATIONS **220** CITATIONS

SEE PROFILE



Psychotherapy Research
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713663589

Emotion-focused therapy for incarcerated offenders of intimate partner violence: A 3-year outcome using a new whole-sample matching method

Antonio Pascual-Leone[a]; Ralph Bierman[b]; Robert Arnold[c]; Eugene Stasiak[d]
[a] University of Windsor, Psychology, Windsor, Ontario, Canada [b] Independent practice, Toronto, Ontario, Canada [c] University of Windsor, Sociology, Anthropology, and Criminology, Windsor, Ontario, Canada [d] Ontario Correctional Institute, Brampton, Ontario, Canada

First published on: 18 April 2011

To cite this Article Pascual-Leone, Antonio , Bierman, Ralph , Arnold, Robert and Stasiak, Eugene(2011) 'Emotion-focused therapy for incarcerated offenders of intimate partner violence: A 3-year outcome using a new whole-sample matching method', Psychotherapy Research, 21: 3, 331 — 347, First published on: 18 April 2011 (iFirst)
To link to this Article: DOI: 10.1080/10503307.2011.572092
URL: http://dx.doi.org/10.1080/10503307.2011.572092

Psychotherapy Research, May 2011; 21(3): 331–347

Routledge
Taylor & Francis Group

Emotion-focused therapy for incarcerated offenders of intimate partner violence: A 3-year outcome using a new whole-sample matching method

ANTONIO PASCUAL-LEONE,[1] RALPH BIERMAN,[2] ROBERT ARNOLD,[3] & EUGENE STASIAK[4]

[1]*University of Windsor, Psychology, Windsor, Ontario, Canada;* [2]*Independent practice, Toronto, Ontario, Canada;* [3]*University of Windsor, Sociology, Anthropology, and Criminology, Windsor, Ontario, Canada &* [4]*Ontario Correctional Institute, Brampton, Ontario, Canada*

(Received 12 April 2010; revised 3 March 2011; accepted 8 March 2011)

Abstract
A 3-year follow-up was conducted for Relating Without Violence, an emotion-focused group psychotherapy program for incarcerated men who have a history of intimate partner violence. This is the only known manualized experiential treatment for incarcerated batterers. A sample of 66 men who completed the treatment was compared to 184 men from the same prison. Although the initial research design was quasi-experimental, new procedures were used to match the groups' averages on all known pre-treatment variables. At 7 and 8 months post-release, the treatment group recidivated by assault and/or sexual assault significantly less than controls. Treatment effects are comparable to those of best practices. The study also demonstrates methodological developments for statistically creating matched groups not previously used in psychotherapy research.

Keywords: domestic violence; recidivism; emotion-focused therapy; follow-up; matched groups; propensity scores; generalized boosted regression

Research suggests that most violence against women falls from the hand of an intimate partner. Furthermore, the number of victims is staggeringly high (Sartin, Hansen, & Huss, 2006). According to the National Victimization Against Women Survey (NVAWS), almost one-quarter of women in the United States have been physically assaulted and/or raped by an intimate partner in their lifetime. Moreover, although other methods have produced even higher estimates the NVAWS indicates that approximately 1.3 million women are physically assaulted by their current or former spouses and boyfriends annually (Tjaden & Thoennes, 2000). To better contextualize these devastating figures, it is important to acknowledge that intimate partner violence is a complex problem, involving both male and female perpetrators. Indeed, it has long been recognized that, in heterosexual relationships, women are slightly more likely than men to initiate violence (see Archer, 2000, for a review of this topic). Men, however, are more likely to inflict injury and to cause serious harm. On that note, this paper focuses on male violence against women in the context of intimate relationships. In addressing this issue, shelter-related services for abused women and children will remain of high importance, although they do not prevent domestically violent men from continuing to be abusive in subsequent relationships. Thus, there continues to be an urgent need to address domestic violence at its source by preventing repeat offences by men who have been violent at home.

We begin this paper by outlining related treatment research and providing a rationale for examining a new, emotion-focused, treatment approach to working with intimate partner violence. Next, we highlight some of the issues related to treating intimate partner violence in a prison setting, a treatment milieu which is very rarely explored in the literature. Both a focus on emotion and an in-prison setting are characteristics of Relating Without Violence, the treatment program used in this study. Then, to

Correspondence concerning this article should be addressed to Antonio Pascual-Leone, University of Windsor, Psychology, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4 Canada. Email: apl@uwindsor.ca

ISSN 1050-3307 print/ISSN 1468-4381 online
DOI: 10.1080/10503307.2011.572092

better frame this paper, we consider a few methodological issues related to this area of research, before finally describing the current study.

Overall, this paper offers three main contributions. (1) This is the first long term outcome study of a new, *emotion-focused and experiential* treatment for intimate partner violence (IPV). Moreover, (2) this is the only published outcome study to date conducted on a treatment of IPV that is provided *during incarceration*. Finally, (3) the paper demonstrates a new methodological solution that has virtually never been used in psychotherapy research; it involves statistically creating matched groups such that a quasi-experimental design can be "upgraded" to control for a large, albeit discrete, number of extraneous variables.

Treating Intimate Partner Violence (IPV): A New Perspective

A current meta-analysis of 22 treatment outcome studies concluded that treatment for the perpetrators of Intimate Partner Violence (IPV) has a significant albeit disappointingly small impact on the rates of recidivism. Still, even a small impact on the prevalence of IPV could be clinically meaningful for some victims of family violence (Babcock, Green, & Robie, 2004). There are currently two major treatment approaches for perpetrators of IPV that are by far the most common and established (Sartin et al., 2006). On one hand, feminist psychoeducational approaches (i.e., the Duluth model of power and control) regard IPV principally as a social problem related to power and inequality among partners. On the other hand, cognitive and behavioral therapies (CBT; i.e., anger management and skills-based programs, among others), understand IPV as mainly a problem of behavioral self-control related to a lack of skills or faulty assumptive frameworks. Although both these formulations clearly have merit and garner comparable amounts of empirical support (see Babcock et al., 2004), recent literature suggests a new view that is becoming increasingly coherent. This third perspective conceptualizes IPV as also being related to problems in emotional functioning—the abuser has poor emotional awareness coupled with inadequate and maladaptive attempts at regulating emotional distress. A number of research directions support this perspective.

It is known that traumatic early life experiences in which parents have been abusive, cruel, and coercive in attempts to control a child are a frequent part of the personal histories of many abusive adults (Dutton, 1995; Lewis, 1992). These issues remain largely unaddressed by CBT or psychoeducation programs, which already suggests an *emotion-focused* approach to treatment could make a new contribution. Moreover, a corpus of research from several empirical lines of inquiry has shown that a history of this kind, of complex relational trauma, is related to deficits in facets of emotional intelligence such as the awareness, regulation, and processing of feelings (Paivio & Pascual-Leone, 2010). Consistent with this, research has shown a link between the personality psychopathologies that describe batterers and an array of dynamic emotional characteristics, such as avoidance, jealousy, attachment-related anxiety, and compulsive care seeking (Waltz, Babcock, Jacobson, & Gottman, 2000). Furthermore, in a review of his life's work, Dutton (2008) highlighted the finding that men convicted of wife assault scored significantly lower than men in the general population on all facets of a measure of emotional intelligence; namely on indices of emotional self-awareness, empathy, stress tolerance, and impulse control, among others. These observations reflect *criminogenic needs* that are in-line with the intermediate treatment targets recommended by Andrews et al. (1990).

In presenting a very recent development in treatment, Tollefson, Webb, Shumway, Block and Nakamura (2009) observed that IPV often erupts in a violent outburst that to some extent is precipitated by a lack of awareness and an inability to regulate emotional arousal. These authors then go on to describe the promotion of mind-body awareness as a promising treatment direction when working with IPV perpetrators. In another example, efforts have begun to develop an attachment-oriented psychodynamic treatment to help abusive men develop an understanding of a link between their internal experiences and their use of violence (Sprout, 2007). Drawing on a body of empirical work, Paivio and Pascual-Leone (2010) have described the role an emotion-focused and attachment-oriented treatment can play in promoting emotional processing, meaning exploration, and in overcoming emotional deficits that may have been the unfortunate legacy of complex trauma. Furthermore, research has demonstrated *emotion-focused therapy* (EFT) as an empirically supported treatment and recent work has been expanding its application.

McMurran and Gilchrist (2008) have identified defensive anger as a dynamic factor that is important as a target for the treatment of IPV. From the perspective of EFT, hostile anger in the context of an intimate relationship is understood as *secondary defensive anger*; that is, a reaction which clouds (or even prevents) some underlying and more vulnerable primary emotion. In cases such as these, underlying primary emotions are often experiences of maladaptive shame or fear (Pascual-Leone, Gillis, Singh, & Andreescu, 2011). Research on experiential and

emotion-focused intervention principles has shown them to be effective in addressing this underlying and emotionally dysfunctional process, although they have not yet been fully applied to the problem of reducing violence (Paivio & Pascual-Leone, 2010). For these reasons, there appears to be a strong rationale for developing experiential and emotion-focused treatments for perpetrators of IPV in order to address emotional damage (i.e., personal shame, insecure attachment fears, and poor emotional awareness) that underlies abusiveness as well as the behaviors and cognitions that maintain it (Wolfus & Bierman, 1996). Indeed, in the conclusions of their meta-analysis, Babcock and colleagues (2004) propose that "more emotion-focused, rather than cognitively focused, interventions may increase the effect size of batterers' treatment" (p. 1064), adding that more research is still needed.

Treatments in Prison

Although IPV is believed to be substantially underreported, those individuals who have come to the attention of the legal system are an easily identified group for treatment (Sartin et al., 2006). In addition to being easily identified, men with histories of IPV who are also incarcerated represent a population that has often committed relatively more severe forms of IPV. At other times they may be incarcerated because of a failure to comply with less restrictive sanctions in the community (Babcock & Steiner, 1999). With regard to treatment, however, Babcock and colleagues (2004) observe that since the introduction of mandatory arrests and court-mandated counseling, intervention with perpetrators of IPV has became a fusion of punishment and rehabilitation. Thus, whether in-prison treatments are voluntary or not, it comes as no surprise that incarceration itself may exacerbate abusive attitudes. *Toxic masculinity* (i.e., misogyny, homophobia, greed, and violent domination) plays a significant role in prison culture and "this training in toughness does not help prisoners prepare for postrelease adjustment as caring, loving men" (Kupers, 2005, p. 717). Unfortunately, despite a significant need for unique kinds of in-prison mental health care, treatment programs, including those for domestic violence, are relatively scarce (Kupers, 2005). Finally, the *risk principle*, supported by meta-analyses by Andrews et al. (1990) among others (i.e., Lipsey & Cullen, 2007), posits that treating incarcerated offenders of IPV should have a greater impact than on lower risk offenders in the community.

For these reasons, a prison setting actually provides an important treatment opportunity at a time when some men may need it most. Perhaps the first evaluation of a treatment program for incarcerated offenders of IPV was presented by Wolfus and Bierman (1996). Their pre-post study examined treatment related changes while participants were still incarcerated (as described below). More recently, Ley (2005) completed a post-release outcome study using a 26-week CBT group treatment for incarcerated domestic violence offenders. She found a significant reduction in the rate of recidivism at 1 year post-release of the CBT group as compared to a control group (18.6% vs. 30.0% reoffence, respectively). Similarly, the current study is a longer term post-release outcome study that follows the initial work of Wolfus and Bierman (1996) using an experiential and emotion-focused treatment called Relating Without Violence.

Relating Without Violence: An Experiential and Emotion-Focused Treatment

Relating Without Violence (RWV) is a group treatment program for incarcerated men who admit to having a history of IPV and was developed in response to an initiative in 1986 by the government of Ontario (Canada) to reduce family violence. In the targets for this treatment, violence is broadly defined to include physical, sexual, or psychological assault. The treatment is an experientially based psychotherapy that integrates several approaches. First, focusing (Gendlin, 1996) is used with empty-chair techniques from emotion-focused therapy (EFT; Greenberg, Rice, & Elliott, 1993) to promote experiential awareness, self-reflection, role reversal, and victim empathy. Second, eye movement desensitization and reprocessing (EMDR; Shapiro, 1995) is used to help clients overcome psychological effects of traumatic life experiences (many men in the target population meet formal DSM-IV criteria for Post Traumatic Stress Disorder). Third, as treatment progresses these approaches are integrated with other techniques modified from imago couples therapy to help promote relational skills (Hendrix, 1990). The treatment goals are (1) to improve psychological and emotional factors that seem to contribute to domestic violence and (2) to develop conflict resolution skills in perpetrators of domestic abuse (Wolfus & Bierman, 1996).

To date, RWV is the only known manualized experiential treatment for incarcerated batterers (see Bierman & Cheston, 1996; Wolfus & Bierman, 1996). It is structured using a time-limited, closed-group format with two modules. Thus, the focus of the first 6 weeks is *therapy for emotional change*, which then serves as the foundation for the focus of the last 6 weeks, *therapy for relationship change*. The specific tasks used in this treatment can be summarized as

aiming to (a) process emotion related to traumatic life experiences (i.e., of parental shaming, physical/sexual abuse, or abandonment); (b) develop focusing and empathy as experiential skills; (c) develop behavioral strategies for emotion regulation (i.e., time outs, mindfulness); (d) enhance awareness of other's emotions; (e) use cognitive restructuring to reduce self-shame and improve self-esteem; (f) articulate feelings and personal needs through imaginary dialogues with significant others; (g) develop an experience-near understanding of the relationship between primary shame and secondary anger/rage; (h) develop dialoguing skills for conflict resolution; (i) develop empathy for the abused partner. Finally, this is done in the context of a group, which allows the treatment to make use of mutual support, modeling, and a buddy system, in which participants practice listening and empathy skills with each other as homework between the treatment sessions.

Wolfus and Bierman (1996) conducted an initial pre-post study of the RWV treatment using a treatment sample of 57 domestically abusive men comparing them to both 30 abusive and then 30 non-abusive no-treatment controls, all from the same institution. Findings showed that over 13 weeks, the RWV program changed the way men who had been domestically violent dealt with conflict within the confines of the prison institution and produced favorable changes on measures of aggressive behavior and defensiveness. Indeed, correctional staff at the facility also reportedly perceived program participants as less aggressive in their interactions (Bierman & Stasiak, 2006). Moreover, the rate of absenteeism and attrition in this voluntary program was low, around 10%, and usually associated with either dropouts or the transfer of a prisoner to a different location, or both. This compares favorably with the commonly reported attrition rates of 40 to 60% (Babcock et al., 2004). Of course, although participation in RWV was voluntary, the fact that treatment was run in prison promised few competing interests; which makes it difficult to truly compare attrition rates with voluntary treatments in the community.

Comparing Outcomes: Difficulties in Research Design

Before discussing the current study in detail, we highlight a few methodological issues related to this area of research. Difficulties in finding suitable comparison groups, as well as the unique nature of small treatment effects on recidivism are each critical background issues.

Finding a Suitable Comparison Group in IPV Research

A number of published reviews have observed that random assignment in a treatment study for IPV offenders is extremely difficult to achieve, especially given the involvement of the legal system (Babcock et al., 2004; Rosenbaum & Leisring, 2001; Sartin et al., 2006). Court-ordered treatment and very high attrition rates are both common obstacles to sound clinical trials. In addition, although clients in most clinical trials voluntarily accept the potential consequences of being in a no treatment or waitlist control group, the non-treatment of IPV offenders is likely to have a much more significant impact on victims than on the actual participants. Thus, although it is sometimes possible, researchers must be mindful that the design of non-treatment control groups in this field can pose certain ethical problems (Sartin et al., 2006).

To address these widespread problems Chen and colleagues tried using statistical corrections of the inherent selection bias that results from a court-mandated treatment (Chen, Bersani, Myers, & Denton, 1989). Their method accounted for the likelihood of a judge to assign a given offender to treatment but was still unable to control directly for the host of predictors associated with domestic violence. More typically, studies on the treatment of domestic violence involving incarcerated men have either not used control groups or used comparison groups without measuring extraneous variables that are known to increase the risk of domestic violence (Donnelly et al., 2002, cited in Ley, 2005; Jenkins & Menton, 2003; Wolfus & Bierman, 1996). These include variables such as age, education level, marital status, prior felonies, a history of alcohol or substance abuse, and mental health difficulties (particularly personality disorders).

One exception to this is the treatment study by Ley (2005), mentioned above. In her study, a treatment group of 70 men were matched one-to-one by selecting another 70 men from a database of comparable men. By using the traditional case-matching method *some* of the key aforementioned variables could be controlled. Unfortunately, there are limits to the number of variables that can be controlled using this method. Consequentially, the study was unable to control for other key variables that have also served as risk predictors in prior research, including: employment, association with deviant/criminal peers, social skills, and attitudes regarding violence against women, among others (cf., Holtzworth-Munroe & Stuart, 1994; Jenkins & Menton, 2003; Sartin et al., 2006). The

method used in the current study addresses this continual concern in a new way.

High-Grade Quasi-Experimental Designs: A Primer on Whole-Sample Matching

When conducting a randomized controlled trail (RCT) is not practical or feasible, the researcher must depend on quasi-experimental designs. Indeed, the need for high-grade secondary alternatives to the RCT is an emerging area of methodological interest in health care research at large (see West et al., 2008, for a review). One viable alternative is the use of quasi-experimental designs that are augmented with statistical corrections to create matched *samples* (McCaffrey, Ridgeway, & Morral, 2004; West et al., 2008). This method has begun to receive widespread attention and statistically creates a suitable comparison group using *propensity scores* (i.e., the probability of a participant's group membership given a set of observed characteristics; McCaffrey et al., 2004). West and colleagues (2008) explain the basic idea of this approach in their paper, "Alternatives to the Randomized Controlled Trial":

> If the researcher can accurately construct propensity scores that balance the treatment and control participants on all potentially relevant baseline variables, the difference between the response in the treatment condition and the control condition (conditioned on the propensity scores) will represent the causal effect. In essence, conditioning on the basis of the propensity scores attempts to create homogeneous units whose responses in the treatment and control groups can be directly compared. (p. 1362)

Concretely, a simple approach for doing this is to include propensity scores directly as a variable in regression analyses, although unpaired cases can result in the loss of a substantial amount of information due to reduced sample sizes (see Bartak et al., 2009, 2010). Another approach, however, is to weight cases (rather than including or dropping) thereby retaining the full sample; we refer to this as "*whole-sample matching*." Since this method was first introduced by Friedman (2001), only a handful of psychological studies (mostly published in *Psychological Methods*) have used it to negotiate the challenges of causal effect modeling with naturalistic data (for review see Stuart, 2010). For example, researchers in mental health (Harder, Stuart, & Anthony, 2010) recently used the method to conduct a longitudinal study on the causal relationship between cannabis problems during early adolescence and negative mental health outcomes later in life. To that end, they statistically corrected a comparison group to control for the covariates of early cannabis use disorders, thereby addressing a question for which an RCT was patently not possible.

In another study of naturalistic data, McCaffrey et al. (2004) presented the only other paper to apply this method for treatment evaluation. In a design similar to that of our current study, the authors evaluated a residential substance-abuse treatment program for young probationers, comparing it to other residential treatments. By using data from an extensive intake interview on substance abuse and related risk factors, the authors successfully used weights (i.e., propensity scores) to match the target and comparison samples on all available data and then estimated a corrected treatment effect. Although the study by McCaffrey and colleagues was pioneering, this method of retaining all cases while "controlling" naturalistic data for an accurate outcome evaluation remains virtually unheard of in treatment research. Even so, we believe that whole-sample matching could be a useful tool given the emerging directions of practice-based research. In the current study, we use this method to provide suitable controls for naturalistic data collected through the legal system.

Interpreting Outcomes: How Large an Effect Should One Reasonably Expect?

Finally, a note about treatment effects for IPV will also be useful here. An intuitive way of evaluating treatment effect sizes is to compare them to those of other treatments for other populations. When examining psychotherapeutic treatments for violence, however, it is critical to consider the unique nature of clinical significance, particularly when effects are small. On one hand, a large effect size (i.e., $d = .80$) has come to be regarded as the average effect associated with robust psychotherapies when compared to no treatment (Wampold, 2001). On the other hand, in the field of forensics, a considerable body of treatment outcome research suggests the overall effect is small (i.e., around $d = .10$) when treatment (i.e., *correctional rehabilitation*) is compared to no treatment, although the range of variability is very wide (see e.g., Lipsey, 2009; Lipsey & Cullen, 2007). Work on best practice CBT programs in forensics reach higher effects, usually ranging between $d = .20$ and .30 (Landenberger & Lipsey, 2005).

Furthermore, as Babcock and colleagues (2004) have argued, it may be unfair to directly compare the effects of treatments for IPV with those of psychotherapy in general for at least three reasons. First, most psychotherapies address internalizing problems

(e.g., anxiety, depression) rather than externalizing problems such as anger and aggression, which are notoriously difficult to treat and still represent an underdeveloped treatment area. Second, client factors such as motivation contribute considerably to therapeutic change (Wampold, 2001); and while some treatments for IPV are voluntary, the treatment samples inevitably represent a skeptical group with mixed motives and ambivalent feelings about participating in treatment (Gondolf, 2001). Third, and perhaps most importantly, the clinical impact of an effect size in treatments for IPV arguably holds a different meaning than effects in most other psychotherapies. In the case of treatments for depression, for example, a very small albeit reliable effect (i.e., a 5% reduction) would indicate only the slightest change in depressive symptoms as compared to no treatment and would not amount to any clinical significance. In contrast, changes in violent recidivism are arguably more dramatic and polarized than changes in psychological symptom severity. As Babcock et al. (2004) point out, the clinical significance of a reduction in recidivism can be concretely understood as successfully averting the injury or death of a certain number of persons. Thus, although a 5% treatment effect for IPV may appear insignificant (when compared with treatment effects on depression), given all reported cases of domestic violence in the United States it would equate to approximately 65,000 women not being beaten or killed each year. For these reasons, one may expect smaller treatment effects but they must be carefully interpreted.

The Current Study

Now, after 15 years of running Relating Without Violence (RWV) groups we are in a position to examine the longitudinal effects of treatment on recidivism rates following participants' release from prison. This paper addresses the question: Were participants in the treatment able to sustain nonviolence when they returned to their normal social milieu?

Since Wolfus and Bierman's (1996) initial study of RWV as an experiential and emotion-focused treatment for incarcerated men, other cognitive and behavioral approaches have been developed and tested. However, none have provided a long term post-release follow-up. The current study was conducted in the context of an existing, albeit unique, prison-based approach to partner violence. Working in a naturalistic setting imposed certain parameters, one of which is that randomization was not feasible. In short, this research is designed to contribute to the psychotherapy outcome literature on the treatment of perpetrators of IPV and also to explore a new method for comparing outcomes in a non-randomized research design.

First, it is hypothesized that a suitable comparison group can be created by statistically matching the treatment group's *risk profile* (i.e., a wide range of variables know to be predictive of recidivism). Second, it is expected that upon clients' release from prison, the RWV treatment program will decrease the rate of recidivism of violence offences by assault and/or sexual assault as compared to the control group. Third, it is also expected that recidivism rates following the in-prison treatment will be maintained over time in a manner that is at least comparable to those reported in the literature for community-based treatments.

Method

Participants

A sample of 250 men was drawn from the database of the Ontario Correctional Institute (OCI), a medium security prison in southern Ontario, Canada. Inclusion criteria required that all participants had sentences of incarceration of more than 11 months served between 1996 and 2003. All participants had a reported history of, and were at risk for, violence against their current or former partners (i.e., physical or sexual assault) according to explicit self-report and/or through routine assessment by OCI staff. The most severe offence related to incarceration was not always a change related to IPV, although for many it was (see below).

Treatment and comparison sub-samples. Sampling was conducted to include 66 men from the prison who admitted to abusing their partners and voluntarily participated in a treatment program for domestic violence (i.e., Relating Without Violence) during the period of their incarceration. The treatment sample used in this follow-up study adds additional cases to those reported on in Wolfus and Bierman's (1996) pre-post study. As a comparison group, another 184 men meeting the same study inclusion criteria were chosen at random from an already existing database at the same prison. Additional inclusion criteria for the comparison group were that men had been identified by OCI staff as having a history of, and being at risk for, partner abuse and that they did not attend the RWV program. All clients in the treatment sample had given informed consent to participate in both the treatment program and its evaluation, while de-identified data from individuals in the comparison group was made available for research purposes

through the Corrections and Conditional Release Act of Canada.

Sample demographics. The full research sample of 250 men were between the ages of 20 and 64 ($M = 36$, $SD = 8.56$). Upon admission to OCI, 60% of the sample was assessed as having some combination of: marital dissatisfaction, non-rewarding family relationships, and a criminal spouse/family (as rated on the Levels of Service Inventory, see Measures). Similarly, 58% of the sample was assessed as having a history of no regular employment, less than grade 12 education and/or at least one suspension/expulsion, and problem interactions with peers and/or authority. Average prison sentence length of these men was 532 days ($SD = 140$, range = 334 to 809).

As indicated, all participants in this study had a reported history of partner abuse. However, an examination of the most severe offences pre-incarceration showed that for 35.6% of the full sample ($n = 250$) assault and/or sexual assault was the primary reason for incarceration. In further detail: 18.4% ($n = 46$) were charged with sexual assault and 17.2% ($n = 43$) were charged with assault; both charges were examined again later in the study's post-release follow-up. The remaining most frequent charges prior to incarceration included 14% (35) of break and entry, 8% (20) of robbery, 6% (15) of uttering a death threat, and 5.2% (13) of forcible confinement in addition to other less frequent crimes.

Demographics are only presented for the full sample at this point in the paper because any pre-treatment differences between the treatment and comparison groups on key variables (which may have been related to volunteerism or other factors) were later corrected through statistical procedures of whole-sample matching. As such, between-group differences are presented at length and addressed in the procedures and results sections that follow. However, as we shall see, the general treatment needs of the sample (i.e., Andrews et al.'s (1990) principles of *risk* and *criminogenic need*) are similar to those of other groups of violent offenders in many ways.

Setting

In addition to being a medium security prison setting, OCI is a 220-bed residential treatment centre for incarcerated male offenders. The men at this centre are primarily violent offenders, sex offenders, and property offenders. Sex offenders are routinely referred to OCI, while others are referred to the institute on the grounds that they suffer from emotional or other psychological difficulties. As a treatment centre, OCI offers a range of rehabilitation approaches including individual therapy, group therapy, milieu therapy, skills training, and education; while being an inmate at OCI was contingent on participating in a minimum number of these programs. One of those treatments was the Relating Without Violence program (Wolfus & Bierman, 1996).

Treatment

The Relating Without Violence (RWV) program for domestic abusers was followed by the treatment sample during their incarceration. The treatment is designed for men who voluntarily participate and who have admitted to being abusive in their family relationships and with their romantic partners in particular. In accordance with a manual (Bierman & Cheston, 1996) the treatment took place over 12-week periods and involved both group and individual therapy components, with approximately 12 men per cohort. A closed-group format was the major component and consisted of weekly sessions of 4 hours (including breaks). All treatment cohorts were jointly facilitated by the second author (R.B.), who was the principal group leader, and a female co-facilitator. In addition, two 45-minute individual sessions were available to group members each week (Wolfus & Bierman, 1996). Absenteeism and attrition was 10%.

Measures

Using an integrative conceptual framework that bridges mental health and correctional treatment programs, Magaletta and Verdeyen (2005) have pointed out that, unlike assessment practices used in mental health service and research, those in a correctional setting focus on criminogenic risk/need. The predominant assessment tool for that purpose is the Level of Service Inventory-Revised coupled with a risk/needs interview. The authors also argued that recidivism is the most widely used, and most appropriate, measure of treatment outcome in a correctional setting. The current study follows these recommendations.

Level of Service Inventory (control variables). The Levels of Service Inventory, Ontario Revision, (LSI-OR; Andrews, Bonta & Wormith, 1995) is a risk/needs assessment tool designed to predict recidivism among offenders. This instrument is a revision of the original Levels of Supervision Inventory (LSI; Andrews, 1982), one of the most widely used and well-researched risk/needs assessment tools in the field. The tool has

been adopted by many jurisdictions in Canada, the United States, and the United Kingdom (Girald & Wormith, 2004). Like its parent measure, the LSI-OR is based on a semi-structured interview and a review of official records and case files. It includes 65 binary items describing criminogenic risk and needs within 10 different domains grouped on two factors: General Risk and Specific Risk. Domains of special interest for research on domestic violence include: family/marital status, substance abuse, antisocial patterns, personal problems with criminogenic potential, and history of perpetration. Individual items of interest indicate: diagnosis of personality disorders, anger management deficits, intimidating/controlling behavior, poor social skills, sexist behavior, as well as intra- and extra-familial sexual assault.

In this practice-based research study, evaluators using the LSI-OR were trained to professional standards and used the tool following standard procedures in the field. Internal validity, temporal consistency, and convergent validity with other measures of risk are all well established for the LSI-OR. Research on inmates serving less than 2 years in Canadian prisons has repeatedly demonstrated the sound psychometric properties of this tool (see Girald & Wormith, 2004, for a full list of references). At a 31-month follow-up, general risk scores on the LSI-OR have been show to correlate with general recidivism ($r = .39$). Similarly, the specific risk/needs scale, which has been tested using special subgroups including domestic violence offenders, also provides relatively high correlations ($r = .34$) for violent recidivism as compared to other measures in the field (Girald & Wormith, 2004).

Most Severe Offence (control variable). The most severe *pre*-treatment offence of each member of the sample was appraised using the Offence Severity Scale (OSS). This scale consists of 25 categories ordered by criminal seriousness and was developed by the Research Department of the Ministry of Correctional Services of Ontario after examining the category of offence and average sentence length of 60,000 offenders (Ministry of Correctional Services, 1983; Stasiak, Winter, & Nixon, 1996). As a measure, the OSS is an objective classification tool based on the recorded convictions of inmates and it does not rely on interviews or test administration. As such, this measure holds high construct and high face validity.

Recidivism (outcome variable). Although there are different approaches to measuring the recidivism of domestic violence, many studies in the literature have relied primarily on legal records. A number of previous researchers have considered conviction for any violent behavior as evidence of recidivism (Sartin et al., 2006), and the current study also adopted this approach.

Records of convictions and dispositions taken from the Ontario Ministry of Community Safety and Correctional Services' Offender Tracking Information System were reviewed for each of the 250 men for criminal episodes prior to incarceration and up to 3 years post release. Conviction records were available for all participants over the 3 years. Because of their relevance to intimate partner violence and the thrust of the treatment program, all offences of *recidivism by assault and/or sexual assault* were examined irrespective of whether it was family related violence: A total of 20 different possible assault charges (including assault, aggravated or simple, with or without a weapon, murder, hostage taking, etc.) and 11 different possible charges of sexual assault (including with or without harm, threatening harm, etc.).

Procedures

Data collection. Pre-treatment data were collected upon participants' incarceration. The RWV treatment program was run with two consecutive cohorts per year as described above, over several years. Inmates were made aware of the treatment program as part of their involvement in the OCI treatment centre, and word of mouth seemed to have a positive impact on recruitment. Participation was voluntary and steps were taken to ensure that there was no secondary incentive (i.e., participants were informed that therapists would not write letters for judicial hearings etc.).

Whole-sample matching procedures. Although the design was quasi-experimental, statistical methods were used to match the groups' means and distributions on a large range of pre-treatment variables. The treatment and control groups were compared on 19 pre-treatment variables (16 from the LSI-OR, in addition to three others). These included, for example: birth year, most severe offence before incarceration, sentence length at the prison, and a range of indices representing both general and specific risk factors for recidivism (see Table I for a complete list). When the two groups differed, it was necessary to control for the difference.

For this purpose, a procedure developed by Friedman (2001) and incorporated within Ridgeway, McCaffrey and Morral's (2006) R package, *twang*, was used. This procedure, known as *generalized boosted regression*, provided a robust estimate of the probability that a given case pertained to either the treatment or control group, given its scores on

the full set of variables which predict group membership. Logistic regression coefficients from many bootstrapped samples were used to create a final set of regression weights for the predictors. Based on these coefficients and a subject's scores on the predictors, each case in the control group was assigned a weight (i.e., a propensity score). When applied, these weights produced a control sample closely resembling the treatment group. This method of statistically generating a profile-matched control group has almost never been used before in psychotherapy research. Note that a traditional case-by-case matching method, using 19 variables, would certainly have required us to discard a good number of potential control cases as poorly matched. However, using the current approach we were able to retain all relevant cases in the analysis (thereby gaining maximum statistical power), with a control group very similar to its demonstration counterpart.

In summary, weighting the control cases was done such that the two groups did not differ significantly on age, sentence length, general risk level, specific risk level, employment, education, and a range of other relevant variables. To date, this relatively new method has been used in a few other longitudinal studies in psychology (e.g., Harder et al., 2010; McCaffrey et al., 2004). Moreover, the method sometimes creates weighted control groups that are even more comparable to treatment groups on their pre-treatment differences than if random assignment had been used (Friedman, 2001).

Results

On the 25-point Offence Severity Scale, where low scores indicate more severe offences, the *pre*-treatment distribution of the full study sample ($n = 250$) had a mode of 3 (Category: violent sexual crimes), and a median of 4 (Category: break and enter). However, the sample ranged in the seriousness of their offences from homicide (ranked 1) to traffic crimes involving alcohol (ranked 18) on the severity scale. The ranking of participants' most severe offence on the Offence Severity Scale was used with other variables to match the treatment and control groups.

Establishing a Pre-Treatment Match on Key Variables

Given the nature of what initially was a traditional quasi-experimental design with a post hoc control, it became important to take into account any pre-treatment differences. As one might expect, given the nature of a self-selecting voluntary treatment group and of a non-randomized comparison group, the two groups differed significantly on a number of important pre-treatment variables. The first three columns of Table I show pre-treatment differences between the treatment group ($n = 66$) and the comparison group ($n = 184$) before unique weights were applied to each control case. Individual statistics for these preliminary analyses are not shown in the table but a series of *t*-tests showed the two groups to be significantly different at the .05 level ($|t| > 2.00$, $df = 248$) on 11 of the 19 variables. These differences generally tended to be small but significant and they indicated that the initial unmatched (i.e., uncontrolled) comparison group had slightly higher scores than the treatment group on variables which have been related to criminal risk for domestic violence (cf. Girald & Wormith, 2004; Rosenfeld, 1992).

Thus, after Friedman's (2001) procedure was implemented, different weights were applied to each of the 184 individual cases in the comparison group. For example, while some cases were weighted as full comparison cases (weight = 1), others were weighted only as partial cases (e.g., = .4), and still others counted as more than a single case (e.g., = 2.1), etc. It follows that applying weights to each case provided a new set of means for each of the variables in Table I (see column 4). In this way the 184 comparison cases were weighted in a manner that matched their collective group means to those of the treatment group (whole-sample matching), thereby controlling for pre-treatment variables.

To ensure that this group-based matching was effective, the weighted control group was then compared as a whole to the treatment group on the basis of both central tendency and distribution. The right-hand of Table I (see columns 5 to 8) shows these findings. The reported *t*-tests for differences in mean were conducted by *twang* after weighting the cases, so weights were taken into account when defining the resulting *p*-values. Differences between treatment and the weighted control groups were all non-significant, with all *p*–values > .13 and a median $p = .55$. *Twang* also provides a Kolmogorov-Smirnov (*ks*) test for differences in distribution. This test is sensitive to any observable difference in distributions, but when differences in means are trivial it is primarily sensitive to differences in shape (for dichotomies the standard deviation is a function of the mean so the *ks* test provides no additional information and is reported for the sake of completeness). Again, *ks* differences between treatment and weighted control groups were all non-significant, with all *p*–values > .44. It is useful to note that given the large number of variables being controlled for, a randomizing process could not have been expected to produce more equal groups. In fact, a randomization design would likely have yielded two to three variables for which *t*-tests

Table I. Establishing a match between control and treatment groups on pre-treatment variables

Predictor	Treatment $M^{\star}$	Unweighted control M	Weighted control $M^{\star}$	t	p-value	ks	p-value
Birth year	1966.05	1967.47	1966.22	−.14	.89	.08	.95
Most severe offence[a, b]	6.14	6.56	6.47	−.53	.60	.08	.96
Sentence length (days)[a]	590.55	511.05	575.93	−.66	.51	.09	.91
General risk[c]	21.44	24.91	22.00	−.42	.68	.08	.96
Criminal history	5.49	5.96	5.52	−.13	.90	.09	.88
Education/employment	4.23	4.96	4.49	−.58	.56	.07	.99
Family/marital	1.91	2.25	1.97	−.32	.75	.04	1.00
Leisure/recreation	1.32	1.48	1.31	.03	.97	.04	1.00
Companions	1.49	1.81	1.57	−.54	.59	.07	.98
Procriminal attitude/ orientation	1.30	1.85	1.36	−.26	.80	.08	.94
Substance abuse	4.21	4.84	4.39	−.48	.63	.11	.74
Antisocial pattern of behavior	1.50	1.77	1.38	−.69	.49	.06	.99
Specific risk[c]	5.88	6.77	5.87	.02	.98	.08	.94
Personal problems with crime risk	2.06	3.72	2.60	−1.46	.14	.12	.64
History of perpetration	1.76	2.58	2.08	−1.14	.26	.14	.45
Anger management deficits[d]	.52	.77	.54	−.34	.74	.04	1.00
Intimidating/controlling[d]	.38	.43	.33	.54	.59	.03	1.00
Physical assault (extrafamilial)[d]	.36	.60	.44	−.93	.35	.06	.99
Physical assault (intrafamilial)[d]	.38	.64	.48	−1.27	.21	.10	.79

Note. Treatment group means (column 2) are contrasted here with first unweighted (column 3) and then weighted (column 4) versions of the control group. Treatment and unweighted groups were substantially different (tests not shown). $^{\star}t$- and *ks*-tests shown here examine differences between *treatment* group vs. *weighted* control group means; [a]variables not part of the Levels of Service Inventory LSI-OR; [b]according to the Offence Severity Scale; [c]broad subscales on the LSI-OR that subsume domain areas listed below; [d] items of special interest subsumed under the domain area *History of perpetration*. All other variables are domain areas on the LSI-OR; $df = 248$; all p-values in column 6 (corresponding to t-tests) are $>.13$; all p-values in column 8 (corresponding to *ks*-tests) are $>.44$.

showed p–values $<.13$ and eight to nine variables for which *ks* tests showed p–values $<.44$.

These findings, which support the first hypothesis, indicate that by weighting cases in the control group, we successfully matched the two groups on all relevant variables for which data were available. This effectively controls for all linear effects in the measured pre-treatment differences through a statistical procedure that matches groups on their profiles, making them amenable to a comparison of outcomes that controls for a range of key variables.

Recidivism

Preliminary examination of the data (i.e., prior to the application of weights through the above matching procedure) revealed a 15.6% ($n = 39$) recidivism rate of the full sample ($n = 250$) by assault and/or sexual assault over a 3-year period. Furthermore, of those who re-offended by the end of that follow-up period, 35 were charged with assault (14.0%; including six counts which involved bodily harm or a weapon) and four were charged with sexual assault (1.6%). These were considered poor outcomes for the treatment of intimate partner violence.

In using the weighted-control procedure during analyses, even though recidivism per case remained binary, weights were applied to cases and then carried on to affect treatment outcomes in that some cases are given more weight than others, thereby impacting the measure of outcome for the group as a whole. The remainder of this paper will compare recidivism rates between the two groups after matching weights have been applied to the control group (i.e., $n = 66$ treatment cases vs. $n = 184$ weighted control cases). Figure 1 presents the assault and/or sexual assault recidivism rates of each of these groups over three years post-incarceration.

Differences in Recidivism at 7, 8, 9 Months Post-Release

Preliminary analyses of all predictor variables ensured that variables met assumptions for all analyses. Only the subscale "General risk" needed to be addressed by truncating the variable's range to

preserve linearity. The first violent offences occurred in the comparison group at 6 months. Although logistic regression would have been the preferred method of analysis at all follow-up times, statistical packages do not allow one to run analyses with no cases in the treatment group and at early measurement times there were an insufficient number of recidivists to support predictors other than the treatment condition in a logistic regression. For this reason chi-squares were used at measurement times prior to the first year post-release, while logistic regression was used at subsequent measurement times.

Conventionally, the observed values in a chi-square test are compared to naive, "expected" values based on pure chance. However, a more controlled test of our treatment's observed values was to set the expected values of chi-squares (2×2) to those recidivism rates found in the control group. In this way, patterns of observed recidivism in the treatment group could be compared to the expected control group's values at 7, 8, and 9 months. Findings showed that when weights are used to control all known pre-treatment variables, the treatment group recidivated by assault and/or sexual assault less than the control group at 7 and 8 months post-release (chi-squares $=4.86$, and $=4.28$, respectively, both $p < .05$). This initial difference indicated a significant benefit for those in the treatment group. However, by the time participants in the RWV treatment program reached 9 months post-release their rate of recidivism could no longer be distinguished at the .05 level from that of the control group (chi-square $=2.77$, $.10 > p > .05$). The differences in rates of recidivism during the first year are shown on the left-hand side of Figure 1.

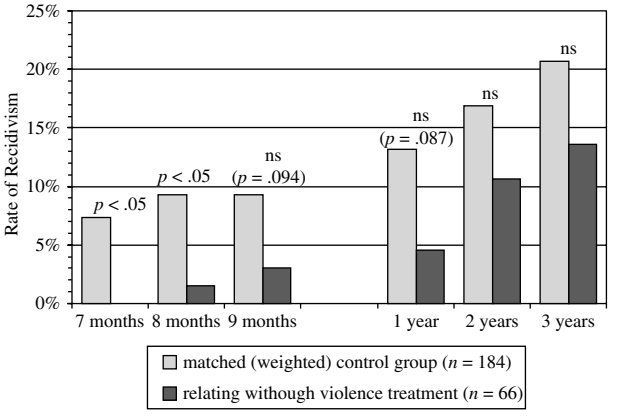


Figure 1. Assault and sexual assault post-incarceration: treatment vs. matched control group.

Recidivism at 1, 2, 3 Years Post-Release

Having applied weights for matching the treatment and control groups, analyses were used to partial out co-variates for the 1, 2, and 3 year measurement times—given that the data at these times could support more comprehensive analyses than data at earlier points, as discussed. Even though the two groups were matched on key demographic and risk variables, as one would expect, those variables still had predictive power over recidivism. In other words, although the two matched groups could not be differentiated on any recorded variables, it was possible that the pre-treatment variables could account for enough variation in recidivism to appreciably increase the power of a test for group differences. Partialing out covariates was also necessary because the study aimed to examine specifically the treatment's predictive power over and above the usual predictors of recidivism (or non-recidivism, as the case may be). This was done through a backward selection procedure using logistic regression. The backward selection procedure began by including all the variables presented in Table I in addition to the presence (or absence) of assault/sexual assault as the most severe offence pre-incarceration, and the group membership variable. At the end of each analysis, backward selection resulted in no significant predictors in addition to group membership except at 3 years where a participant's pre-incarceration record of assault/sexual assault was significant.

The long term rates of recidivism for each group are shown on the right-hand side of Figure 1. The logistic regression findings showed that the RWV treatment had no significant predictive power in distinguishing the two groups on the outcome variable of recidivism at 1, 2, or 3 years post-incarceration; $F_{2,249} > 2.53$, all p–values for RWV treatment $> .05$ (see Table II). The variables that predicted violent offences pre-incarceration remained more or less the same variables that predicted recidivism post-incarceration, and having been in treatment did not seem to have a significant effect at 1, 2, or 3 year evaluations post-release. This means that beyond a certain point, the rates of decay by re-offence in Figure 1 are comparable for both groups.

Even so, although the RWV treatment was no longer significant at the .05 level as a predictor at years 1 through 3, its coefficient continued to have a negative value across time. Given that this variable shows a trend of diminishing significance over all six testing periods, it may be fair to assume that the direction of the treatment effect is not arbitrary even at the 3 year mark (see Figure 1).

Table II. Logistic regression predicting assault or sexual assault at 1 and 3 years post release

	1 year		3 years	
Predictor	b	*p*-value	b	*p*-value
Assault/sexual assault pre-incarceration			1.024	.031
RWV treatment	−1.162	.087	−.578	.226
Intercept	−1.882	.000	−1.874	.000

Note. Significance test at 1 year, $F_{1,249} = 2.96$, *NS*. Significance test at 3 years, $F_{2,249} = 2.53$, NS. Analyses done at 2 years post release were consistent with the 1 to 3 year pattern, and as such revealed no significant effect for the treatment group variable at the .05 level.

Finally, it is informative to note that the initial benefit in favor of treatment at 8 months post-release—which was an absolute reduction in recidivism of 7 to 8% (control minus treatment group rates of recidivism)—is a between-group difference that was actually sustained for the following 3 years. Of course, in the context of an increasing number of re-offenders in each group over time, the sustained difference becomes less significant in terms of raw numbers. In keeping with this, if we dropped from the analyses all cases who recidivated in the first 8 months of their release and only examined those who were still crime-free (i.e., 65 from the treatment group, 172 from the control group), we find that *beyond* 8 months the rates of recidivism differ very little between the two groups. Indeed, after accounting for the initial differences at 8 months, rates of the two groups at 9, 12, 24, and 36 months only differ by an average of 1%. This means that the full impact of treatment seems to occur in the first 8 months and, although the difference between groups favoring the treated seems to persist, it was due to initial rather than ongoing positive effects. In summary, these findings provide some support for the second hypothesis, that the RWV treatment group has decreased recidivism rates as compared to the control—although the differences are most noteworthy in the first 8 months.

Comparison to Other Treatments

The third hypothesis stated that recidivism rates following the in-prison treatment would be maintained over time in a manner that is comparable to those in community-based treatments. In a recent meta-analysis on treatments for domestically violent men, Babcock and colleagues (2004) calculated the effect sizes at both short term (less than 12 months) and long term follow-up periods (12 or more months); they did this in terms of *d* using Hedges and Olkin's (1985) meta-analytic method. This provides a method and basis of comparison for our findings.

In keeping with Babcock et al.'s (2004) measurement approach for short vs. long term effect sizes, we calculated recidivism rates from the current study by averaging all three measurement times that were less than 12 months (i.e., 7, 8, 9 month follow-ups) for treatment and control groups ($M = 1.52\%$ and 8.63%, respectively). Recidivism rates were also averaged across all times that were at 12 months or greater (i.e., 1, 2, 3 year follow-ups) for treatment and control groups ($M = 9.60\%$ and 16.94%, respectively). Then, using the same method as Babcock et al. (2004), effect sizes for the current treatment were calculated as $d = .28$ for the short term follow-up (i.e., <12 months) and $d = .25$ for the long term (i.e., >12 months).

According to Babcock and colleagues (2004), when 20 outcome studies that used police reports of recidivism were considered, the effect size of treatment for domestic violence was estimated to be $d = .18$ (95% CI = .11–.25). This overall effect combines short and long term follow-ups, and no significant difference was found in the effects of experimental vs. quasi-experimental designs. However the case may be, effects of RWV at either follow-up seem comparable to those reported elsewhere. In short, these findings support the third hypothesis.

Discussion

Summary of Findings

This study examined the long term effect of a new emotion-focused and experiential prison-based treatment for offenders of intimate partner violence (IPV) and compared their recidivism patterns post-release to those of a matched comparison group. By about 6 months, participants in the comparison group had already committed their first violent offences in the study (i.e., assault or sexual assault). In contrast, participants who had attended the RWV treatment while in-prison seem to have abstained from violent offences until their eighth month post-release and still maintained a significantly lower rate of violent offending up until their ninth month.

Furthermore, at six points of measurement across the 3-year study, recidivism rates for the treatment group were consistently between one- and two-thirds less than those of the weight-matched controls even after the effects faded out of statistical significance. In short, the findings suggest there is a small yet enduring initial advantage from being in the treatment.

In the introduction of this paper, we highlighted key issues for interpreting the clinical significance of effects in psychological treatments for violence as compared to other psychotherapies for mental and emotional difficulties. In essence, while small effects in symptom reduction may not be very noteworthy, the same sized effects in measures of recidivism may have a significant human impact. This study showed that at 12 months post-release the difference in offence rates between the treatment and control groups was a decrease of 7 to 8%. To a clinician this means that a man who was arrested, imprisoned, *and treated* was a bit less likely to commit physical or sexual assault during that year than a man who was only arrested and imprisoned. However, given the high prevalence, recurrence, and severe impact of violence at home, a small effect on the number of individuals being traumatized, hurt, or killed as a result of IPV is worth noting. Babcock et al. (2004), for example, have argued that even a 5% reduction may be clinically or socially significant.

Comparisons Based on the IPV Treatment Literature

Offenders treated with RWV showed a one-quarter to one-third standard deviation in improvement in recidivism as compared to non-treated batterers. One of the unique characteristics of RWV is that it is the only known emotion-focused and experiential therapy for intimate partner violence, which is important to consider in the context of its comparable effects vis-à-vis other treatment approaches. In their meta-analysis, Babcock and colleagues (2004) examined the overall effect size due to treatment types within quasi-experimental designs that used police reports of recidivism: They reported that Duluth (feminist-psychoeducational) treatments had an effect size of $d = .32$, CBT had an effect size of $d = .12$, and "other" (i.e., supportive, eclectic) treatments had an effect size of $d = .27$. However, given the variability of scores that went into these average effect sizes and their wide confidence intervals, Babcock et al. concluded that these effects did not differ significantly. Thus, our study's findings regarding an emotion-focused treatment (i.e., $d = .25$–$.28$) seem to fit very well among the average effects of other approaches.

At a long term follow-up of more than a year post-release the treatment effect of RWV, when provided to an in-prison population ($d = .25$), was identical to the long-term treatment effect found by Babcock et al. (2004) for community-based treatments (95% CI = .14–.35). However, at a short term follow-up, of a year or less, the effect of RWV ($d = .28$) was somewhat larger than the effects found in community-based treatments overall ($d = .13$; 95% CI = .04–.22; Babcock et al., 2004). Thus, the current prison-based treatment appears to have yielded a slightly larger effect than community-based treatments in the short term. Although this might be accounted for by the new treatment, it may simply reflect the population and circumstances post-release that are associated with a prison-based treatment.

Incarcerated Men: Treatment and Recidivism

Studies included in the meta-analyses by Babcock and colleagues (2004) reflect treatments that are voluntary or court-ordered but they are primarily provided in a community-based setting; conducting treatments in prison, as done in the current study, is very uncommon. Thus, in considering this treatment's effectiveness, it is paramount to be mindful of the fact that there may be fundamental differences across the populations being served. Domestically violent men living in the community may represent a different subgroup than offenders who are incarcerated. Authors who have described subtypes of domestically violent men suggest that members of the more *generally violent subgroup* are likely to be over-represented in a prison sample (Holtzworth-Munroe & Stuart, 1994). As such, men in the current study should be considered higher risk than those represented by the findings of Babcock et al. (2004). Even so, the relative effect of treatment on the current study sample does not seem to support the Andrews et al. (1990) *risk principle*—the observation from other research that higher risk clients have more to gain and therefore are more impacted by treatment. Even so, the current study sample may also have represented a subgroup that was more *difficult* to treat. In addition, prisoners have also been subjected to a punitive sanction for their offence, which presumably interacts with treatment. Despite the fact that effect sizes in the current study are on par with those reported in the literature, these observations actually suggest that comparing the effects of this prison-based treatment to those of community-based treatments may be more stringent than is called for.

Perhaps the only other study on a treatment for IPV which was provided during incarceration is the one conducted by Ley (2005), who managed to control for

a limited number of extraneous variables using one-to-one case-matching design. Her study of CBT reported a 38% decrease in reoffence rates (i.e., percent of non-overlapping data) at 12 months post-release, whereas the current treatment showed a 66% decrease in recidivism at 12 months (relative to controls). This difference favors the experiential treatment (i.e., RWV) although without replications any such conclusion remains tentative. However the case, this study presents one of the few treatments designed for incarcerated individuals. Furthermore, it appears that community-based treatments do not help individuals transfer knowledge from the treatment context to *real life* any more effectively than this treatment administered in a prison setting.

Clinical Implications

Positive findings in this field always need to be regarded with some degree of caution; particularly given that, upon final analysis, the success of IPV treatment programs remains inadequate. All too often, victims of violence express high levels of confidence in treatment providers and programs, some making decisions to continue or re-establish relationships with violent partners based on the belief that treatment will have "worked." This confidence can sometimes be misguided, placing partners at risk of further victimization. Even so, with the current study as one of only a few exceptions, there is no evidence base from which to guide clinicians or policy makers who are interested in the treatment of IPV in a prison setting. Given that prisons are likely to house some of the highest risk and most serious offenders of domestic abuse, it is particularly important to continue work in this area.

One obvious possible development is to extend the length of the initial treatment. Furthermore, in order to enhance the duration of reduced rates of violent and sexual offence recidivism, we propose that *booster* sessions may be useful in helping treated men maintain their gains. Thus far, the literature on treatment for IPV has offered little guidance on whether or when follow-up treatment (i.e., booster sessions) might be helpful in a field where the rapid decay of treatment gains is so discouraging (cf., Babcock et al., 2004; Sartin et al., 2006). Moreover, given the fact that resources are so limited for convicts once they are released from prison, any ongoing support must be delivered sparingly and strategically (Kupers, 2005). We have observed that the RWV treatment has its full impact within the first 8 months post-release and the extent of that modest positive impact may be sustained over several years. As such, this study suggests 7 to 8 months post-release, prior to any re-offence by the treatment group, might be a critical time to provide additional treatment. Follow-up sessions strategically offered during that period (e.g., offering 3 or 4 group sessions at weekly intervals) could serve as both a "refresher" and as a way of generalizing treatment effects from prison-life to the outside world.

Methodological Contributions

The method used in this study of matching whole groups based on a *profile of variables* in order to control quasi-experimental designs post hoc has important implications for psychotherapy research and could help address two perennial problems in our field. First, while psychotherapy researchers are increasingly looking toward practice-based evidence, treatment data collected in a naturalistic setting often means that an RCT may not be a feasible or practical option (Bartak et al., 2009; McCaffrey et al., 2004; West et al., 2008). Some of the more common examples from practice-based research where true randomization becomes very difficult to ensure are when treatment assignment is confounded by self-selection, legal constraints, or attrition. This is especially true when treating hard–to-reach populations such as domestic abusers but also those suffering from addiction, concurrent disorders, or personality disorders. As in the current study, weighted controls may be used in studies of those populations to create more suitable, sample-matched, comparison groups (for more demonstrations and detailed discussions of this method see Harder et al., 2010; Haviland et al., 2007; McCaffrey et al., 2004; for a discussion of a related method using propensity scores see Bartak, 2009).

Second, even when randomized clinical trials of psychotherapy are successful, it is often difficult to confidently compare the effect sizes from one treatment study to another if their respective study samples differ in complex ways. So, between-study comparisons are usually done in the context of a narrative or meta-analytic review. That being the case, unless a randomized study is run to directly compare two or more treatments, any comparison of different treatment approaches is indirect and limited by the degree to which the studies have each controlled for a few select variables. However, using the *whole-sample matching method* presented here, two (or more) independently drawn treatment samples (e.g., from separately conducted outcome studies) could be directly compared by weighting cases across groups so as to effectively match the two treatment samples on a very wide range of variables. For example, while Bartak et al. (2010) do not use whole-sample matching as such, they use a related

method to compare multiple treatment modalities, pooling data from five different sites.

Limitations and Future Directions

While the program is effective in a manner that is comparable to existing treatments, a dismantling study is now needed to determine which components of the current treatment program and its setting are producing the apparent advantage over no treatment. Future studies should also use measures of anger reduction as intermediate or process-related outcomes. Our findings suggest that even beyond the point where treatment no longer produced a statistically significant difference there appeared to be a pattern of positive treatment effects. A larger sample would increase statistical power to detect these differences by conventional criteria. By the same token, it is well established that incidents of IPV go largely under-reported and that criminal records, as were used in this study, routinely yield lower reported rates than measures that use victim reports (Babcock et al., 2004; Sartin et al., 2006). However, victim/partner reports were unavailable for this study and so it is reasonable to assume that the recidivism rates reported here are underestimates of the true rates for reoffence by assault or sexual assault.

Arguably, one limitation of this study is that is has not directly controlled for volunteerism through randomization. While this is true, it is important to allow that the issue of volunteerism in treatment research gains much of its gravity from the host of variables that systematically correlate with it, suggesting that volunteers are somehow different from the target population. In this case, however, after matching groups on a substantial profile of 19 variables, the method has statistically controlled for many of the correlates which might otherwise have been related to volunteerism. This makes the volunteerism critique quite vague in the face of hard empirical data, although skeptics may speculate about "unknown" correlates.

This having been stated, however, measurements themselves are still subject to reporting bias. Furthermore, the adequacy of a weighted comparison group will always depend on the quality of the initial baseline assessment and the inclusion of variables believed to be related to volunteerism and outcome (West et al., 2008). Thus, upon final analysis, one must acknowledge that even with powerful methods of adjusting for many covariates, we cannot unequivocally guarantee that selection bias or other confounds do not exist (also see McCaffrey et al., 2004).

Despite the close matching between treatment and control groups, and the fact that participants were drawn from the same prison setting, samples in this study are likely to have represented a heterogeneous group of offenders. However, there is an increasing recognition that different approaches to treatment may better suit different kinds of offenders and future research should examine treatment responders by subtype. For example, the formulation that IPV follows from reactive anger which stems from traumatic shame or fear and therefore could be addressed using an emotion-focused approach may hold true for a number of individuals, but there are certainly other pathways to IPV. In their recent study on a public health approach to partner violence, Coulter and VandeWeerd (2009) have provided a practical example for delivering a multilevel treatment. Finally, as indicated in the introduction, female perpetration of domestic violence is also a topic of importance and it needs research attention. This is particularly the case because although women are more likely to sustain greater injury in this context, men are more likely to be killed (as a result of gunshot wounds, knife wounds, etc.; Archer, 2000). Thus, future directions should also investigate the treatment of female perpetrators of partner violence.

Concluding Remarks

An emotion-focused and experiential therapy for IPV, such as Relating Without Violence, is a unique approach in the field and follows on the heels of other emotion-focused approaches that have demonstrated efficacy for a number of disorders (see Greenberg & Pascual-Leone, 2006). IPV is disturbingly widespread and very few individuals proportionally volunteer for treatment. However, incarceration provides a unique, contained, environment to do this treatment and ensures high attendance as a result of few competing interests. Still, it will be important to investigate the impact of an emotion-focused and experiential treatment on individuals who have been domestically violent but continue to reside in the community.

Acknowledgements

Ralph Bierman passed away on December 25, 2008. The RWV therapy he developed reflects his conviction that treatment should be, "hard on problems, soft on people." This paper is dedicated to him.

The views expressed in this report are those of the authors and are not necessarily those of the Ontario Correctional Institute or the Ministry of Community Safety and Correctional Services.

References

Andrews, D.A. (1982). *Levels of Supervision Inventory (LSI): The first follow-up.* (Report). Toronto, Canada: Ontario Ministry of Correctional Services.

Andrews, D.A., Bonta, J., & Wormith, J.S. (1995). *Levels of Service Inventory – Ontario Revision (LSI-OR): Interview and scoring guide*. Toronto, Canada: Ontario Ministry of the Solicitor General Correctional Services.

Andrews, D.A., Zinger, I., Hoge, R.D., Bonta, J., Gendreau, P., & Cullen, F.T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, *28*, 369–404.

Archer, J. (2000). Sex differences in aggression between heterosexual partners: A meta-analytic review. *Psychological Bulletin*, *126*, 651–680.

Babcock, J.C., Green, C.E., & Robie, C. (2004). Does batterers' treatment work? A meta-analytic review of domestic violent treatment. *Clinical Psychology Review*, *23*, 1023–1053.

Babcock, J.C., & Steiner, R. (1999). The relationship between treatment, incarceration, and recidivism of battering: A program evaluation of Seattle's coordinated community response to domestic violence. *Journal of Family Psychology*, *13*, 45–59.

Bartak, A., Spreeuwenberg, M.D., Andrea, H., Busschbach, J.J.V., Croon, M.A., Verheul, R., Emmelkamp, P.M.G., & Stijnen, T. (2009). The use of propensity score methods in psychotherapy research: A practical application. *Psychotherapy and Psychosomatics*, *78*, 26–34.

Bartak, A., Spreeuwenberg, M.D., Andrea, H., Holleman, L., Rijnierse, P., Rossum, B.V., Hamers, E.F.M., Meerman, A.M., Aerts, J., Busschbach, J.J.V., Verheul, R., Stijnen, T., & Emmelkamp, P.M.G. (2010). Effectiveness of different modalities of psychotherapeutic treatment for patients with cluster C personality disorders: Results of a large prospective multicentre study. *Psychotherapy and Psychosomatics*, *79*, 20–30.

Bierman, R., & Cheston, J. (1996). *Relating without violence: A manual for a treatment program for domestically abusive men*. Ontario Correctional Institute, Brampton.

Bierman, R. & Stasiak, E. (2006). *Reduced violence after Relating Without Violence: A prison-based domestic violence integrative therapy program*. Workshop at the Conference of the Society for Exploration of Psychotherapy Integration, Los Angeles, USA.

Chen, H., Bersani, C., Myers, S.C., & Denton, R. (1989). Evaluating the effectiveness of a court sponsored abuser treatment program. *Journal of Family Violence*, *4*, 309–322.

Coulter, M., & VandeWeerd, C. (2009). Reducing domestic violence and other criminal recidivism: Effectiveness of a multilevel batterers intervention program. *Violence and Victims*, *24*, 139–152.

Donnelly et al., (2002). A batterer education program for incarcerated African-American men. In E. Aldarondo and F. Mederos (Eds.), *Programs for men who batter: Intervention and prevention strategies in a diverse society.* Kingston, NJ: Civic Research Institute.

Dutton, D.G. (1995). *The batterer: A psychological profile*. New York, NY: Basic Books.

Dutton, D.G. (2008). My back pages: Reflections on thirty years of domestic violence research. *Trauma, Violence Abuse*, *9*, 131–143.

Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*, 1189–1232.

Gendlin, E.T. (1996). *Focusing-oriented psychotherapy: A manual of experiential method.* New York, NY: Guilford Press.

Girald, L., & Wormith, J.S. (2004). The predictive validity of the Level of Service Inventory – Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice and Behaviour*, *31*, 150–181.

Gondolf, E. (2001). Limitation of experimental evaluations of batterer programs. *Trauma, Violence and Abuse*, *2*, 79–88.

Greenberg, L.S., & Pascual-Leone, A. (2006). Emotion in psychotherapy: A practice friendly research review. *Journal of Clinical Psychology: In-session*, *62*, 611–630.

Greenberg, L.S., Rice, L.N., & Elliott, R. (1993). *Facilitating emotional change: The moment-by-moment process*. New York, NY: The Guilford Press.

Harder, V.S., Stuart, E.A., & Anthony, J.C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, *15*, 234–249.

Haviland, A., Nagin, D.S., & Rosenbaum, P.R. (2007). Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychological Methods*, *12*, 247–267.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hendrix, H. (1990). *Getting the love you want: A guide for couples*. New York, NY: Harper Perennial.

Holtzworth-Munroe, A., & Stuart, G.L. (1994). Typologies of male batterers: Three subtypes and the differences among them. *Psychological Bulletin*, *116*, 476–497.

Jenkins, J.A., & Menton, C. (2003). The relationship between incarcerated batterers' cognitive characteristics and the effectiveness of behavioral treatment. *Corrections Compendium*, *28*, 1–9.

Kupers, T.A. (2005). Toxic masculinity as a barrier to mental health treatment in prison. *Journal of Clinical Psychology*, *61*, 713–724.

Landenberger, N.A., & Lipsey, M.W. (2005). The positive effects of cognitive–behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, *1*, 451–476.

Lewis, M. (1992). *Shame: The exposed self*. New York, NY: Free Press.

Ley, L.F. (2005). *A study of domestic violence recidivism following treatment among incarcerated men who batter.* Unpublished dissertation. University of Miami, Coral Gables, FL.

Lipsey, M.W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders*, *4*(2), 124–147.

Lipsey, M.W., & Cullen, F.T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science*, *3*, 297–320.

Magaletta, P.R., & Verdeyen, V. (2005). Clinical practice in corrections: A conceptual framework. *Professional Psychology: Research and Practice*, *36*, 37–43.

McCaffrey, D.F., Ridgeway, G., & Morral, A.R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, *9*, 403–425.

McMurran, M., & Gilchrist, E. (2008). Anger control and alcohol use: Appropriate interventions for perpetrators of domestic violence? *Psychology Crime & Law*, *14*, 107–116.

Ministry of Correctional Services. (1983). *Offence Severity Scale*. Toronto, Canada.

Paivio, S.C., & Pascual-Leone, A. (2010). *Emotion-focused therapy for complex trauma: An integrative approach*. Washington DC: American Psychological Association.

Pascual-Leone, A., Gillis, P., Singh, T., & Andreescu, C.A. (2011). *Hate, rage, and destructive anger in psychotherapy: An emotion-focused perspective*. Unpublished manuscript. University of Windsor, Canada.

Ridgeway, G., McCaffrey, D., & Morral, A. (2006). *Toolkit for weighting and analysis of nonequivalent groups: a Tutorial for the twang package*. (R Package Version 1.0 –1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://finzi.psych.upenn.edu/R/library/twang/doc/twang.pdf.

Rosenbaum, A., & Leisring, P.A. (2001). Group intervention programs for batterers. *Journal of Aggression Maltreatment & Trauma*, *5*, 57–71.

Rosenfeld, B.D. (1992). Court-ordered treatment of spouse abuse. *Clinical Psychology Review*, *12*, 205–226.

Sartin, R.M., Hansen, D.J., & Huss, M.T. (2006). Domestic violence treatment response and recidivism: A review and implications for the study of family violence. *Aggression and Violent Behaviour*, *11*, 425–440.

Shapiro, F. (1995). *Eye movement desensitization and reprocessing*. New York, NY: Guilford Press.

Sprout, B.K. (2007). *Healing the origins of intimate violence: A model using attachment and psychoanalytic theory.* Unpublished dissertation. Alliant International University, San Francisco, CA.

Stasiak, E., Winter, K., & Nixon, B. (1996). *Awaiting trial: Accused persons remanded to custody.* Toronto, Canada: Ontario Ministry of Correctional Services.

Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*, 1–21.

Tjaden, P., & Thoennes, N. (2000). *Full report of the prevalence, incidence, and consequences of intimate partner violence against women: Findings from the National Violence Against Women Survey (NCJ 183781)*. Rockville, MD: U.S. Department of Justice, Office of Justice Programs.

Tollefson, D.R., Webb, K., Shumway, D., Block, S.H., & Nakamura, Y. (2009). A mind-body approach to domestic violence perpetrator treatment: Program overview and preliminary outcomes. Journal of Aggression. *Maltreatment & Trauma*, *18*, 17–45.

Waltz, J., Babcock, J.C., Jacobson, N.S., & Gottman, J.M. (2000). Testing a typology of batterers. *Journal of Consulting and Clinical Psychology*, *68*, 658–669.

Wampold, B. E. (2001). *The great psychotherapy debate: Model, methods, and findings*. Mahwah, NJ: Erlbaum.

West, S.G., Duan, N., Pequegnat, W., Gaist, P., Jarlais, D.C.D., Holtgrave, D., Szapocznik, J., Fishbein, M., Rapkin, B., Clatts, M., & Mullen, P.D. (2008). Alternatives to the randomized controlled trial. *American Journal of Public Health*, *98*, 1359–1366.

Wolfus, B., & Bierman, R. (1996). An evaluation of a group treatment program for incarcerated male batterers. *International Journal of Offender Therapy and Comparative Criminology*, *40*, 318–333.