

Effectiveness of Risk-Needs-Responsivity-Based Family Violence Programs with Male Offenders

Lynn A. Stewart · Natalie Gabora · P. Randall Kropp · Zina Lee

Published online: 4 February 2014
© Springer Science+Business Media New York 2014

Abstract The following presents the outcome of an evaluation of family violence prevention programs for male offenders. The moderate and high intensity programs were designed and implemented to conform to the Risk-Needs-Responsivity (RNR) principles. Results showed that program participation significantly reduced attitudes that supported violence against women and improved pro-social skills related to non-abusive relationships; treatment effects were moderate to high. Parole officer feedback generally reported positive changes in behavior and attitude associated with treatment. A post-release follow-up indicated that program completion significantly reduced spousal violence and general violent recidivism. Program participation did not, however, have a significant impact on non-violent crime. Results indicate that domestic violence programs respecting the RNR principles may be effective in reducing partner violence.

This report is in partial fulfillment of contract 21120-02-7743-INH-00 between the BC Institute Against Family Violence (BCIFV) and the Correctional Service of Canada. Funding for the project was provided by the Department of Justice, Canada and Health Canada. The following individuals made valuable contributions to this project: Jessica Ius and Melanie Bania assisted with the parole officer interviews; Katherine Rossiter and Dayna Medveduke were responsible for coding recidivism data; Stephen Hart, Ph.D., provided consultation on statistical analyses. The views and opinions expressed in this report are those of the authors and should not be construed to represent the views of the Correctional Service of Canada or the Canadian government.

L. A. Stewart (✉) · N. Gabora
Research Branch, Correctional Service Canada, 340 Laurier Ave.
West, Ottawa, Ontario K0A 0P9, Canada
e-mail: lynn_stewart@sympatico.ca

P. R. Kropp
Forensic Psychiatric Services Commission of British Columbia and
Simon Fraser University, Burnaby, BC, Canada

Z. Lee
School of Criminology and Criminal Justice, University of Fraser
Valley, Abbotsford, BC, Canada

Keywords Intimate partner violence · Correctional treatment outcome · Domestic violence treatment outcome · Male offenders · RNR · Effective corrections

The General Social Survey (GSS) on victimization has shown the extent to which domestic violence is a serious social issue in Canada. While the latest survey found that rates have declined over the last 10 years among Canadians over 15 years of age, 7 % of women and 6 % of men report having been assaulted by their partners within the last 5 years (Statistics Canada 2011). The same survey found that Aboriginal-Canadians were disproportionately affected by domestic violence, with those who self-identified as an Aboriginal person being more than twice as likely as non-Aboriginals to report being the victim of spousal violence (Brennan 2009). In light of this concern, many provinces and territories have enacted legislation to address domestic violence. As a result, perpetrators of domestic violence are being arrested, charged, and convicted, and institutions are then tasked with managing and treating these individuals.

Research on the effectiveness of intervention programs that address domestic violence is still in the early stages, with few well-controlled studies completed to date. The more rigorously designed evaluations found weak or no treatment effects, depending on whether the outcome was assessed through official records or victim reports (Carney and Buttell 2006; Davis and Taylor 1999; Davis et al. 1998; Dunford 2000; Feder and Dugan 2002; Palmer et al. 1992; Stover et al. 2009; Taylor et al. 2001). In a recent update of their research, the team of Maxwell et al. (2010) found that even the weak effect they identified in their original analysis did not persist beyond the participants' period of court control. In contrast, several quasi-experimental designs comparing dropouts to completers have shown treatment effects, even when differences between completers and dropouts were controlled

(Gondolf 1999). A meta-analysis of controlled batterer intervention program (BIP) studies found that the overall effect size of BIPs is small, but significant (Babcock et al. 2004) and a review by the Centres for Disease Control concluded that the intervention was effective (Morrison and Nesius 2003). In contrast, a meta-analysis of court mandated programs (Feder and Wilson 2005) and a systematic review (Feder et al. 2008) concluded that there was no consistent evidence of the effectiveness of these programs. Overall, the most optimistic conclusion is that these programs have only a modest impact on reducing repeat domestic violence. Without doubt, more needs to be done to identify practices that produce stronger treatment effects.

One explanation for poor treatment results is that in the outcome studies data were aggregated across all participants in a treatment condition. Specifically, none of the studies evaluated the effect of different types of intervention on participants of differing backgrounds or typologies (Holtzworth-Munroe, and Meehan 2004) who may respond differently to the standardized intervention model being evaluated (Buttelt and Carney 2006; Cavanaugh and Gelles 2005; Stuart 2005). Other suggested reasons for the disappointing outcomes are: failure to assess or account for untreated substance abuse and mental disorders, the impact of poverty and stake in conformity, failure to sanction non-compliance, and inclusion of generally violent men in programs not designed to address general antisocial behavior (Bennett et al. 2007). Problems with inconsistent program implementation, including variable attention to the training and monitoring of program delivery staff, lack of integrity of program materials, and very high dropout rates have also been noted (e.g., Day et al. 2009). One recent analysis of dropouts in correctional programs determined that an average of 37.8 % of participants never complete domestic violence programs (Olver et al. 2011) and for many programs these rates are much higher. Recidivism rates for men who drop out of domestic violence programs are generally higher than for those who complete the programs (Gordon and Moriarty 2003; Jewell and Wormith 2010; Olver et al. 2011). These are precisely the men we need to retain in treatment. Few studies examine components within programs that may contribute to successful outcome (Bowen et al. 2008) making it impossible to know what aspects of these programs drive reductions in future violence. In general, design and implementation of programs to address domestic violence are not as rigorously implemented as general offending behavior programs, and could incorporate the principles and approaches of the effective corrections literature as one way to improve on their results (Day et al. 2009).

Recently, there are calls for domestic violence programs to “grow up”, adapt a paradigm shift, shed ideology, and determine how the maximum benefits can be obtained from programs to address partner violence (e.g., Babcock et al. 2007; Bowen 2011; Dutton and Corco 2006; Stuart 2005). This paper will present the outcome evaluation of two family

violence programs delivered nationally to male federal offenders in the Correctional Service Canada (CSC). The program curricula, assessment procedures, and program implementation strategy were designed to meet standards derived from the effective corrections literature, adhering to the Risk-Needs-Responsivity (RNR) principles (Andrews and Bonta 2010; Gendreau 1996).

Programs to Address Family Violence within the Correctional Service of Canada (CSC)

The risk factors associated with domestic violence are similar to those associated with general criminal behavior (Dutton and Hart 1992; Hanson and Wallace-Capretta 2000; Harris et al. 2011). It is not surprising, therefore, that an estimated 40 % of male federal offenders have a history of violence against their intimate female partners (Stewart et al. 2011). CSC has mandated treatment for all male offenders who have been identified as being at continued risk to be abusive in intimate relationships. CSC has endorsed a process of program accreditation that sets out criteria, derived from the RNR principles in the effective corrections literature, that programs must meet (Andrews and Bonta 2010). For example, to be accredited, programs must provide sufficient intensity of service based on risk level (Risk Principle); target criminogenic needs (i.e., dynamic factors empirically linked to criminal behavior—the Need Principle); and use effective methods and address responsivity issues (Responsivity Principle). The moderate- and high-intensity family violence prevention programs were accredited by an international panel of correctional experts convened in 2001 and then implemented nationally within CSC.

In CSC, all offenders undergo an extensive assessment at intake that determines their risk level and their criminogenic needs, that is, the dynamic risk factors associated with their criminal histories. Based on the results of these assessments, offenders’ correctional plans are completed. These plans include the programs and interventions the offenders will be referred to that will address their risk and needs prior to their appearance before the national parole board. All the national correctional programs have clear referral criteria set out in policy as part of the Commissioner’s Directive on Correctional Programs. This policy includes the Nationally Recognized Correctional Program guidelines. Only offenders meeting these criteria are offered a place in the programs.

The moderate-intensity program is a 25-session program for moderate risk offenders as assessed on the Spousal Assault Risk Assessment guide (SARA; Kropp et al. 1999), and the high-intensity program is a 78-session program for high-risk offenders (as assessed on the SARA). Both programs adopt the nested ecological model (Dutton 1995) that explains domestic violence as determined by multiple factors. The CSC

programs attempt to address factors at all levels defined by the model. The programs' initial modules establish offender motivation for change and educate offenders on the range of abusive behaviors and factors that contribute to their abusive patterns towards women. Other modules train offenders on cognitive restructuring of attitudes and beliefs that condone the abuse of women, and in skills that address the management of emotions of jealousy, anger, and fear of relationship loss that are associated with abuse. Participants are trained on key social and communication skills that underlie healthy relationships. Later modules involve the offenders in the development of relapse prevention plans that include the planning of coping strategies to avoid or manage high-risk situations, the identification of people to avoid who contribute to the risk of further abuse, and conversely, the identification of a support network who will assist in maintaining a commitment to healthy relationships. The high-intensity program includes a more detailed examination of the influences on the development of abusive patterns through autobiographies and a longer module on parenting. Graduates of both programs follow a maintenance program for a prescribed period.

Both programs incorporate a stage-of-change approach that recognizes that not all offenders will be equally ready to address their history of abuse against their partners. A stage-appropriate treatment primer is available for offenders who meet the referral criteria, but are not ready to attend the program. Additionally, the introductory section of the program itself gives the offenders time to evaluate their goals for change and assess how they want to change.

The curricula for both programs are delivered by trained facilitators whose adherence to the manual and to the principles of effective correctional program delivery is monitored through videotape review of key sessions. Regional trainers conduct detailed quality review reports of the facilitators' work and provide feedback on their program delivery to assist in professional development and also ensure that they meet a specified standard in their assessments and final program reports. Programs are delivered to a group of up to 12 offenders by teams made up of a male and a female co-facilitator. Sessions are 2–3 h, delivered 3 to 5 times per week. At least 3 (moderate-intensity program) to 10 (high-intensity program) individual sessions are conducted with one of the facilitators assigned as a primary counselor. Offenders participate in a detailed assessment of their pre-program attitudes and skill levels and are reassessed upon completion of the program.

Method

Participants

Participants were 572 male offenders incarcerated at various federal institutions across Canada who participated in either

the high- (41 %, $n=237$) or moderate- (59 %, $n=335$) intensity family violence prevention programs between 1999 and 2003. Offenders' risk on the SARA determined allocation to the moderate or high intensity program. Offenders assessed as moderate risk attended the moderate program and those assessed as high risk and who had more than one domestic violence assault were referred to the high intensity program. The mean sentence length of the total treatment group was 4.24 years ($SD=2.99$) and ranged from 2 to 22 years. These sentences were not necessarily for a spousal assault, but to meet the referral criteria, the participants had to have at least one incident of assault of an intimate partner in their personal history. In other words, the assault of an intimate partner did not have to result in a conviction. There was no significant difference between the high- and moderate-intensity participants with respect to sentence length or on any of the demographic variables. Overall, the men ranged in age from 20 to 69 years, with an average age of 37 years ($SD=9.50$). Most of the offenders were Caucasian (62 %), with the remainder Aboriginal (25 %) or of "other" ethnicity (13 %), which included a variety of ethnic groups. Only 25 % of the offenders had a high school education. The mean IQ of the sample estimated from the Shipley Institute of Living Scale (Shipley 1967) was 96.91 ($SD=20.59$). Of note, 79 % of the high-intensity and 60 % of the moderate program participants reported being under the influence of drugs or alcohol when they assaulted their partners.

The high-intensity program participants were assessed for personality disorders in order to determine whether there was a differential response to treatment for the Borderline or Antisocial Personality offenders. This assessment was not completed on the moderate-intensity program participants. The mean score on the measure of Borderline Personality Disorder (BPD) was 62.04 ($SD=18.36$). Using mean scores reported by Oldham et al. (1985), an estimated 26 % of the CSC sample met the diagnosis of BPD. Missing data made it more difficult to determine the percentage of offenders who met the criteria for Antisocial Personality Disorder (APD). However, previous research has shown that up to 75 % of male offenders in CSC meet the broad criteria for APD (Motiuk and Porporino 1991), compared to lifetime rates in the general US population of 3.6 % among American men (Compton et al. 2005).

The offenders had substantial histories of assaultive behavior towards both their partners and others. Most (57 %) reported having been in 10 or more fights; the majority of these were with other males (73 %). Although participants in both programs admitted to substantial physical and emotional abuse of their partners, the high-intensity program participants acknowledged greater, and more serious, abuse than did the moderate intensity participants, confirming that the offenders were allocated to the appropriate treatment option. For example, the high-intensity program participants were more likely

to admit to hitting (72 % vs. 53 %), choking (31 % vs. 19 %) and use of a weapon (40 % vs. 19 %) than the moderate-intensity participants. Thirty percent of the high and 17 % of the moderate-intensity participants admitted having assaulted their partners more than 5 times. Both treatment groups also admitted to equal levels of emotional abuse ranging from yelling (89–90 %), swearing (86–88 %) and threatening (58–59 %), to isolating her (17–25 %) and property damage (47–49 %). Participants in both programs were reluctant to admit to sexual abuse, with 16 % of both the high- and moderate-participants admitting to forcing sex or pressuring sex when their partner did not want it. Their partners had called the police because of their abuse at least once in 67 % of the cases for the high-intensity and at least once in 61 % of the moderate-intensity participants. A majority of the men in the high-intensity program (58 %) and 41 % of the moderate-intensity participants had been assaultive towards more than one partner. As expected, the offenders' social histories included experience of personal victimization and family disruption. Most of the high-intensity participants (61 %) and half of the moderate-intensity participants had witnessed the abuse of their mothers and two-thirds of the men from both programs had themselves been abused as a child. More than half of the participants in both programs reported that family members had criminal records.

On a positive note, almost all of the offenders in both the moderate and high intensity programs at the initial interview (96–99 %) reported believing that the program could be "somewhat" or "very" useful for them. This is reflected in the completion rates for the program. Relative to other batterer intervention programs reported in the literature (Daly et al. 2001; Olver et al. 2011), the dropout rates were respectable. Only 18 % of offenders dropped out of the high-intensity program and 14 % of offenders dropped out of the moderate-intensity program.

Measures

The background information, including the offense history, was extracted from components of the Offender Management System (OMS), the official electronic record on all federally sentenced offenders. All measures that contributed to the profiles were completed at intake by specially trained parole officers on all consenting offenders. Profile information on the risk and need variables were drawn from the Offender Intake Assessment (OIA), which is a comprehensive evaluation conducted on all incoming offenders in CSC by qualified parole officers. The Dynamic Factors Identification and Analysis (DFIA) component of the OIA assesses a variety of dynamic risk factors grouped into seven criminogenic need domains (i.e., employment, marital/family, associates, substance abuse, community functioning, personal/emotional, and attitude). Each domain is divided into principal components that are

further broken down by indicators. A series of yes/no self-reference indicators and, in some cases, help messages accompany the indicators to enhance rating clarity (Brown and Motiuk 2005). The need level for each domain is assessed on a four point rating scale: asset, no need, some need, and considerable need. This measure provides an overall need level (dynamic risk level) categorized into low, medium, or high. Meta-analytic reviews have confirmed the content validity of the DFIA, indicating that the DFIA contains factors that have been identified in the research literature as strong predictors of criminal reoffending (Brown and Motiuk 2005). Statistical reliability (as measured via internal consistency) of each domain ranged from acceptable to superior within three release cohorts: men, women, and Aboriginal offenders. Brown and Motiuk (2005) report that the predictive validity for all DFIA domains range from moderate to strong across all three release cohorts, meaning that individuals who were rated as having more serious problems across the seven domains were significantly more likely to be readmitted to federal custody during the 3-year, fixed follow-up period.

The second component of the OIA is the Static Factors Assessment (SFA), which provides in-depth information regarding an offender's criminal history and the seriousness of the offenses. The overall rating of static risk is made by the intake parole officer based on the following: consideration of the results of the Statistical Information on Recidivism (SIR-R; Nuffield 1982) measure; the criminal history record that documents the nature and extent of the criminal history including adult and juvenile record, type, and number of offenses and crime-free periods; the offense severity record; the sex offense history checklist; reoffense if released; and the criminal profile report, which provides details of the crime(s) for which the offender is currently sentenced. The offense severity assessment consists of an index of offense severity, the type of conviction(s), sentence length, the number and types of victim(s), the degree of force used on victim(s), and the degree of physical and psychological harm to victim(s). The sex offense history checklist assesses type of sex offenses, victims, serious harm, and assessment and treatment history. The final assessment yields an overall rating of low, medium, or high static risk (Motiuk 1997). Information from the DFIA and SFA is used to inform an offender's correctional plan, with offenders assessed as high risk and high need being prioritized for correctional programs.

The SIR-R scale is an actuarial tool that anchors the static risk assessment described above. It assesses risk for general reoffending within 3 years of community release. The scale contains 15 static risk items that are related to recidivism (e.g., volume of previous offending, types of offending, age at first offense, and current age). The Burgess method is applied to weight each item according to the strength with which it is associated with recidivism. The total score results in an overall numerical risk rating that places each offender in a risk category from very poor to very good. The latest validation study

confirmed the tool has good internal consistency with a standard Cronbach's alpha reliability coefficient of 0.75 (Nafekh and Motiuk 2002). The measure is significantly correlated with general recidivism (Pearson r coefficient with SIR-R1 groupings = 0.36, $p < .0001$). Area Under the Curve (AUC) results for federally sentenced male offenders were good at 0.745 (Nafekh and Motiuk 2002).

Since the validity of the SIR-R tool has not been established for Aboriginal offenders, CSC policy does not permit the use of the SIR-R for these offenders. The estimate of risk for Aboriginal offenders is therefore provided through the overall static risk rating, omitting consideration of the SIR-R. Research has shown that the combined ratings of risk and need using the measures in the OIA are highly predictive of outcomes on release (Motiuk 1997).

Following the general risk and need assessment completed on all offenders, specialized assessments are also conducted at intake, depending on the offender's criminal history. For example, a screening of offenders, and, if necessary, a further assessment is conducted to determine risk for sexual reoffending and also for domestic violence.

The Family Violence Risk Assessment Screening (FVRAS) is conducted at intake on all incoming male federal offenders. It consists of reviewing the offender's history for three indicators suggestive of a history of domestic violence: documented partner abuse, documented victim of abuse, and suspected partner abuse perpetrator (based on parole officer's observations of attitudes and fear expressed by partner or ex-partner). A positive response on any one of these indicators prompts a more in-depth assessment, which is conducted using the SARA. The SARA is a structured risk assessment tool that assesses the risk for domestic violence for the purpose of guiding monitoring and intervention. The measure consists of 20 risk factors for spousal assault that are scored from zero to two depending on the strength of the evidence for the presence of the item. The final evaluation provided by the assessor is an overall risk rating (low, moderate, high) that determines risk for future abuse against an intimate partner or former partner. The SARA was developed based on retrospective data on offenders with previous histories of partner abuse. The tool has been used internationally to assess the dynamic risk of offenders with histories of partner abuse and case managers are meant to use the tool to determine the focus of intervention and supervision. Research has confirmed that the measure has good interrater reliability and internal consistency (Kropp and Hart 2000). Currently, the SARA is only validated for men. In CSC, the tool has been incorporated systematically into the risk assessment of all incoming offenders since 2002 and results are recorded in the OMS. An updated SARA is required if there is a change of risk status related to additional incidents of abuse or concerns related to a deterioration in a relationship, or alternatively, if the risk is lowered because of stabilization of risk factors. Prior to assessing offenders on the SARA, parole

officers are required to take an online training module on the administration of the tool and pass a test for certification that involves accurate assessment and analysis of a case.

All offenders referred to the family violence prevention programs complete a battery of tests designed to assess the impact of the program on the program targets: attitudes towards abuse against women, and lack of skill in managing emotions and social situations and coping with high-risk situations. To provide divergent sources of evaluation, a variety of measures were chosen, including self-report, scenario-based skills assessment (vignettes), and facilitator ratings. Correlational analyses confirmed that one of the strengths of including the vignettes and rating scales in the battery was that, unlike the self report measures, the scores were unrelated to offenders' deliberate attempts to "fake good" as measured by the Paulhus Deception Scale (Paulhus 1990). The pre- and post-measures are described briefly below.

Interpersonal Relationship Scale The Interpersonal Relationship Scale (Hupka and Rusch 2001) is a 27-item self-report measure that assesses six aspects of jealousy: threat to exclusive companionship, self-deprecation/envy, dependency, sexual possessiveness, competition and vindictiveness, and distrust. Each item is rated on a 6-point scale ranging from 1 (*strongly agree*) to 6 (*strongly disagree*), with higher scores indicating lower levels of jealousy. Split half reliability is reported by the authors as 0.79. Concurrent validity was established with a correlation of 0.68 with self reported levels of jealousy (Hupka and Rusch 2001).

Abusive Relationships Inventory The Abusive Relationships Inventory (Boer et al. 1993 unpublished) is a 33-item self-report measure that assesses four scales related to abusive relationships: rationale for hitting, need for control, legal entitlement, and batterers' myths. Each item is rated on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), with higher scores indicating more negative attitudes toward relationships. On a sample of 101 federal offenders referred to the moderate intensity family violence program at pre-treatment, Ferguson (2004) found that reliability alphas were high (0.95 for Rationales for Hitting, 0.92 for Need for Control, 0.80 for Legal Entitlement), with the exception of the Batterers' Myths scale with a reliability alpha of only 0.42. It should be noted that Boer et al. had earlier reported an alpha coefficient of 0.75 for the Batterers' Myths scale.

Personal Reaction Questionnaire The Personal Reaction Questionnaire (Blackburn and Fawcett 1999) is a 125-item, multi-trait, self-report inventory designed to measure cognitive, affective, and behavioral dispositions relevant to offender populations. It provides a brief assessment of deviant personality traits and is intended to facilitate the identification of an offender's problems and the planning and monitoring of

clinical intervention. Each item has a dichotomous response choice (*yes/no*), and the measure assesses eight personality traits: resentment, aggression, self-esteem, avoidance, paranoid suspicion, extraversion, deviance, and self-control. The authors report that scales have satisfactory reliability (α) ranging from 0.79 to 0.86 on the scales, and significant correlations with measures of personality disorder, observer ratings of interpersonal style, and criminal career data supporting their construct validity (Blackburn and Fawcett 1999).

Paulhus Deception Scale The Paulhus Deception Scale (PDS; Paulhus 1990) is a 40-item self-report measure designed to assess two subscales of socially desirable responding: self-deceptive enhancement and impression management. Each item is rated on a 5-point scale ranging from 1 (*not true*) to 5 (*very true*). Self-deceptive enhancement refers to a person's tendency to give honest but exaggerated positive self-reports, whereas impression management refers to a person's tendency toward purposeful manipulation of answers to appear more socially acceptable. The PDS has been used with men and women offenders and has been found to have sufficient reliability (Cronbach's alpha ranged from 0.58 to 0.84) and validity, although these results are based on the previous version of the PDS (Kroner and Weekes 1996; Lanyon and Carle 2007).

Borderline Personality Organization (BPO; Oldham et al. 1985). This instrument is a 30-item self-report measure that assesses components of a borderline personality. Each item is rated on a 5-point scale ranging from 1 (*never true*) to 5 (*always true*), and items are summed to yield three subscales: Loss of Reality, Primitive Defences, and Identity Confusion. The three scales can be summed to yield a total score. Oldham et al. (1985) report Cronbach's alpha for the BPO subscales as: Loss of Reality, 0.84; Primitive Defences, 0.87; and Identity Diffusion, 0.92.

Antisocial Personality Symptoms This measure is a checklist of 11 symptoms described in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association 1994) under Antisocial Personality Disorder. The items are scored in a *yes/no* format. Endorsed items are summed to yield a total score. Research established that the number of items endorsed is related to poorer outcomes on release among federal family violence program participants (Stewart et al. 2005) and weakly, but significantly, related to spousal reassault upon release. Ferguson (2004) found a Cronbach's alpha 0.85 based on results of program participants.

Relapse Prevention Test, Family Violence Vignettes and Empathy Scales These measures were developed within CSC to assess skills and attitudes targeted by the program (CSC 2001). They are based on structured interviews in which the respondent

is asked how he would respond in a number of situations related to family violence. There are two versions of the tests (Version A and Version B), with one version administered at pre-treatment and the second at post-treatment. Administration of versions A and B is alternated across offenders to counterbalance the learning effect and level of difficulty. More detailed descriptions of these measures follow.

Relapse Prevention Test There are a total of eight scenarios, two for each of four potential risk situations. Four of the scenarios are administered at pre-treatment and four are administered at post-treatment. The test is designed to assess an offender's (a) recognition of situations leading to violence; (b) effective and systematic use of skills such as problem solving, emotions management, and network support; and (c) ability to evaluate the effectiveness of his solutions. In addition, it requires the facilitator to count how many potentially effective responses the offender suggested. Higher scores indicate that the respondent has good recognition and application of skills. Ferguson (2004) reports good reliability estimates at pre- and post-treatment for all scenarios ranging from 0.68 (Scenario 1) to 0.89 (Scenario 4), and the reliability for the total score at pre-treatment was 0.92 and post-treatment was 0.90. Ferguson (2004) and Connors et al. (2012) found positive change on all scenarios pre- and post-treatment.

Family Violence Vignettes The Family Violence Vignettes examine responses to aspects of relationship violence in situations involving jealousy, employment, and sexual issues. Higher scores indicate that the respondent assesses his partner with respect and as an equal, and uses appropriate behaviors. At pre-treatment, the reliability estimates ranged from 0.52 to 0.86. At post-treatment, Cronbach's alpha range from 0.60 to 0.95 (Ferguson 2004).

Empathy Scales The Empathy Scales involve a structured interview in which the participants are given a number of situations and asked how they would respond. The situations represent a partner in distress, a child in distress, and a person outside the family in distress. Responses are scored on a 3-point scale (0 = *no evidence of empathy skills or abusive response*, 1 = *some evidence of empathy*, 2 = *strong evidence of empathy*) to assess perspective taking, sincerity of affect, and coping with distress. Total scores range from 0 to 72, with higher scores indicating better empathy skills. At pre-treatment, Cronbach's alpha ranged from 0.75 to 0.89. At post-treatment, the reliability estimate ranged from 0.76 to 0.91 (Ferguson 2004).

Treatment Readiness, Treatment Responsivity, and Treatment Participation and Gain Scales (Serin and Kennedy 1997, unpublished). These scales are structured interviews designed to assess treatment readiness and responsivity factors that

could affect an offender's response to treatment. These scales were used for the high-intensity participants only. For the Treatment Readiness Scale, at pre-treatment, the reliability coefficient for the 11 domains ranged from 0.56 to 0.95. At post-treatment, the internal consistency estimates ranged from 0.70 to 0.89. At pre-treatment, the Treatment Responsivity Rating Scale reliability estimate ranged from 0.68 to 0.89. At post-treatment, the reliability coefficients for the 11 domains ranged from 0.74 to 0.90 (Ferguson 2004).

Goal Attainment Scale-Family Violence (GAS-FV; Correctional Service Canada 2001). The GAS-FV is a facilitator rating scale assessing participants' behavior and attitudes on 14 treatment goals associated with the CSC family violence prevention programs. Each goal is rated on a 5-point scale ranging from -2 (*"very risky" attitude or very low skill achievement*; 0 *minimal acceptable attitude or skill achievement*; +2 *very prosocial attitude or high skill achievement*). The scale is completed for each participant with input from both facilitators to ensure greater reliability. The tool has been demonstrated to be dynamic (i.e., is sensitive to treatment change; Connors et al. 2012). Results of a similar tool used in CSC to assess all correctional programs (i.e., the Generic Program Performance Measure) demonstrated that offenders who improved their scores with treatment had better outcomes on release (i.e., lower recidivism rates) than those whose scores remained below minimal acceptable attitude or skill achievement level (Usher and Stewart 2011).

University of Rhode Island Change Assessment-Batterers (URICA-B; Levesque et al. 2000). This instrument, based on the more generic URICA, measures offenders' stage of change with respect to their willingness to stop abusive and violent behavior against an intimate partner. The self-report inventory presents statements characteristic of various stages of change (precontemplation, contemplation, etc.), and the offender rates the extent to which the item is like him on a 5-point scale. The measure is administered during the initial interview for both the moderate and high intensity program participants. Levesque et al. (2000) suggested that the measure is valid based on their findings that participants in the most advanced stage clusters were most likely to have used strategies to end the violence in the last year and engaged in less partner blame.

Rating of Motivation Following the initial interviews and assessment, offenders were rated by facilitators on a 5-point scale ranging from 0 (*no motivation to change to avoid partner abuse*) to 5 (*excellent level of motivation*) on three areas: motivation for the program, extent of abusive thinking, and strategies for managing emotion and resolving conflict. The results of this assessment were provided to the offender so that he could work with the facilitator in agreeing to the goals of treatment. Recent research has shown a significant

relationship between scores of motivation level on this measure and outcomes in a moderate intensity family violence program (Connors et al. 2012).

Quizzes During the program, participants completed quizzes at the end of every module to determine the extent to which they could understand and apply the material in the program (CSC 2001). These quizzes measured various aspects of the program: motivational enhancement, awareness and education, cultural issues, thinking skills, social skills, managing emotions, parenting, and relapse prevention-risk management. The high-intensity participants completed motivational enhancement, awareness and education, cultural issues, thinking skills, social skills, managing emotions, parenting, and relapse prevention-risk management. The moderate-intensity participants completed motivational enhancement, awareness and education, social skills, managing emotions, and relapse prevention-risk management.

Parole Officer Assessment Feedback was solicited from parole officers supervising the program participants on their ratings of the impact of the program on offenders in achieving the goals of the programs related to abusive intimate behavior and risk for further abusive behavior. Parole officers were mailed a letter by independent contractors that explained the purpose of the study and requested their participation in a short interview. A random group of parole officers that represented 65 % of offenders who participated in either the high or moderate intensity family violence program were contacted. Of these parole officers, only 19 % ($n=74$) agreed to participate in the confidential interview.

Participant Feedback Offenders were requested to complete detailed feedback on specific aspects of the program and overall feedback on the general utility of the program. The feedback from the offenders was collected using confidential questionnaires completed anonymously at the end of the program.

Post-Release Follow-Up

The final phase of the evaluation was a post-release follow-up of 246 family violence offenders falling into two groups. The first group was the treated offenders who had completed either the high-or moderate- intensity family violence program and had been released. This group included 84 participants from the high-intensity program and 76 participants from the moderate intensity program. The second group consisted of untreated offenders who met the admission criteria for the programs (i.e., they had a history of spousal violence and a SARA rating of moderate or high), but who, for various reasons, did not begin or complete the program. This group was comprised of 35 offenders who were eligible for the high intensity program but did not attend: 13 who received no treatment

for administrative reasons (e.g., institutional transfer) and 22 who had dropped out of treatment. The group also included 51 offenders who were eligible for the moderate intensity program (22-received no treatment, and 29 dropped out).¹

To be eligible for this phase of the evaluation, offenders were required to have spent at least 6 months in the community following their release. The average time at risk was approximately 1 year. The difference between time at-risk for the two groups was not statistically significant. Researchers were unable to contact enough spouses of the released program completers to include their input in the analyses. Data on recidivism therefore came from Canadian Police Information Centre (CPIC) and CSC OMS records. Information from both sources was used to record new criminal offenses and supervision violations following release.

Initially, new offenses were tallied in five categories: (a) confirmed spousal violence, including any actual, attempted or threatened physical or sexual violence against a spouse; (b) other spousal-related crime including harassing behaviors and breaches of no-contact conditions; (c) non-spousal violence; (d) non-violent crime; and (e) non-spousal-related conditional release violations. Due to the low reported base rate of re-offending, however, three nested offense categories were chosen for analyses of recidivism: spousal violence, which includes all spousal-related crime; any violence, which includes spousal violence and other violence; and any infraction, which includes all violent and non-violent crime, regardless of whether or not the crime resulted in arrest.

Results

Intermediate Measures

The pre-treatment and post-treatment performance of the program participants on the test measures were compared by conducting a series of *t*-tests, and applying a Bonferroni correction for the number of analyses. Almost every comparison found that offenders improved significantly after participation in treatment. In general, upon treatment completion, offenders in both the high- and moderate-intensity programs evidenced significantly lower levels of jealousy (Interpersonal Relationship Scale), fewer negative attitudes about relationships (Abusive Relationships Inventory), better recognition

and application of relapse prevention skills (Relapse Prevention Test), more respect for their partners (Family Violence Vignettes), greater treatment readiness and responsivity (Treatment Readiness Scales), and more engagement in positive behaviors and attitudes (Goal Attainment Scale).

Given the number of comparisons, Table 1 presents only the calculated effect sizes (Cohen's *d*) for the comparisons instead of presenting the average pre- and post-treatment scores for all of these measures. This allowed us to standardize the change scores so that they could be compared directly. The effect size is an estimate of the level of change from pre to post treatment. Cohen's *d* may be interpreted as follows: an effect size of 0.3 is considered "small"; an effect size of 0.5 is considered "moderate"; and an effect size of 0.8 is considered "large." Effect sizes exceeding 1.0 are considered exceptionally high. According to these guidelines, most of the measures had medium to large effect sizes, with the exception of the Interpersonal Relationship Scale and the Abusive Relationships Inventory, which had small effect sizes. Offenders were tested only once on the Relationship Style Questionnaire, the Personal Reaction Questionnaire, and the Borderline Personality Organization measures to profile participants and determine their relative response to treatment.

Quizzes

The content quizzes were designed to determine whether offenders understood the key concepts and could apply them to common relationship situations they had described. The scores for both the moderate- and high-intensity programs indicated that most offenders understood most of the content quite well. The mean average scores on the quizzes were between 65 and 77 % for the high-intensity program and between 63 and 77 % for the moderate-intensity program.

Parole Officer Feedback

The average number of months the parole officers supervised the offender was 10.03 (*SD*=6.55). Attempts were made to ask parole officers about offenders' attitudes and behavior in the context of their current relationships; however, very few offenders (30 %) were currently involved in a relationship. Of those who were, the majority of parole officers (71 %) stated that they either had no concerns about the relationship or that the offender's behavior in the relationship had improved.

In addition to assessing offenders' attitudes and behavior in the context of relationships, parole officers were asked to judge offenders' attitudes and behavior post treatment on the items identified in the Goal Attainment Scale (GAS) using the following categories: does not appear to, possibly/partially appears to, or definitely appears to exhibit the attitude or behavior). Many officers were unable to comment on changes

¹ Preliminary analyses indicated that there were no differences in recidivism rates between those who had received no treatment and those who had dropped out of treatment. Therefore, we chose to combine these offenders into the group labelled "Untreated Offenders" in order to increase the sample size for subsequent analyses. The reasons for dropping out of treatment varied: of the 51 who dropped out of either the moderate or high program, about 30 % did so because they were transferred to another institution, had a conflict with scheduling, or had problems understanding the language of instruction, variables that are not theoretically linked to risk to reoffend).

Table 1 Moderate- and high-intensity program effect sizes

Scales	Effect size moderate program (n=282) ^a	Effect size high program (n=156) ^b
Interpersonal relationship scale		
Threat to exclusive companionship	0.43**	0.59**
Self-deprecation/envy	0.33*	0.48**
Dependency	0.52**	0.66**
Sexual possessiveness	0.30*	0.27*
Competition and vindictiveness	0.56**	0.78**
Distrust	0.13	0.25*
Abusive relationships inventory		
Rationales for hitting	0.37**	0.36**
Need for control	0.56**	0.67**
Legal entitlement	0.38**	0.32**
Batterers' myths	0.37**	0.52**
Relapse prevention test		
Total score	1.09**	1.22**
Family violence vignettes		
Total score	0.88**	0.71**
Empathy scales		
Total score	0.60**	0.84**
Goal attainment scale (Pre vs. post)		
Accepts responsibility for abusive behavior	1.01**	1.18**
Acknowledges use of power and control tactics	1.20**	1.32**
Shows empathy for victims	0.96**	0.92**
Extent of skills development	0.99**	1.27**
Minimizes consequences	0.91**	1.16**
Understands life style dynamics	1.07**	1.20**
Understands abusiveness pattern	1.30**	1.38**
Identifies relapse prevention concepts	1.21**	1.54**
Discloses personal information	0.92**	0.94**
Participation in treatment	0.80**	0.73**
Motivation to change behavior	0.79**	0.65**
Overall quality of relapse plan	1.12**	1.10**
Overall program performance	0.83**	0.85**
Treatment readiness scales		
Treatment readiness		0.78**
Treatment responsivity		0.77**

^a The size of the group varies because of missing data^b The size of the group varies because of missing data* $p < .05$. ** $p < .01$

since the commencement of treatment, since they were often unfamiliar with or could not recollect the offenders' pre-treatment attitudes and behaviors. Table 2 illustrates the percentage of offenders post treatment who were rated in each of these categories on the various items. The parole officers'

judgments were primarily favorable, with the vast majority of parole officers observing at least some positive attitudes and behaviors on the GAS dimensions. Parole officers reported that most offenders had not been charged or convicted of an offense since their participation in treatment (78 %). Of the 22 % who were charged or convicted since treatment, the type of charge/conviction varied, but none was spousal-related. Finally, parole officers were asked to make judgments about whether they felt the offender benefited from treatment. The majority of parole officers stated that they believed that the treatment was either somewhat or very effective (73 %), 17 % stated it was ineffective or somewhat ineffective, and 9 % were unable to say.

Participant Feedback

At the end of every program, offenders completed a confidential feedback form on their opinion of the program. Offenders who completed the moderate program were asked: (a) Was the program interesting? (b) Did the program do a good job? (c) Would you be able to use the skills you have learned in the program in the community? The results indicated that most offenders thought the program was very interesting (81 %) and that the program did a very good job (84 %). All (100 %) of the moderate-intensity offenders who responded to the last question indicated that they would be able to apply the skills learned in the program in the community upon release. Offenders who participated in the high-intensity program were asked to comment on the helpfulness of the program. Almost all participants (98 %) stated that the program was helpful. Thus, overall, participants in both the moderate- and high-intensity programs had positive comments about the specific program content.

Table 2 Parole officers' judgments on goal attainment scale items – post treatment (N=74)

	No	Partially	Definitely
Accepts responsibility for his abusive behavior	9.5 %	33.8 %	56.8 %
Acknowledges use of power and control tactics	13.5 %	35.1 %	51.4 %
Shows empathy	13.5 %	36.5 %	50.0 %
Uses appropriate problem solving skills	13.5 %	41.9 %	44.6 %
Uses skills to manage and control his emotions	12.2 %	37.8 %	50.0 %
Uses appropriate social skills	10.8 %	36.5 %	52.7 %
Recognizes cognitive distortions	12.5 %	50.0 %	37.8 %
Does not minimize consequences of his behavior	16.2 %	25.7 %	58.1 %
Understands lifestyle dynamics	9.5 %	23.0 %	67.6 %
Understands his abusiveness pattern	10.8 %	32.4 %	56.8 %

Recidivism

Table 3 presents the recidivism rates for the three categories of recidivism for the treated and untreated groups in the moderate- and high-intensity programs. Within the high-intensity group, there was a statistically significant difference between the treated and untreated offenders both for spousal violence ($p<.05$) and any violence ($p<.05$). Four percent of the treated offenders committed a new spousal violence offense compared to 14 % of the untreated offenders, which is a 71 % decrease in spousal assault recidivism. Further, 11 % of the treated high-intensity group and 26 % of the untreated offenders committed a new violent offense, resulting in a reduction of violent recidivism of nearly 60 %. For the moderate-intensity group, the same trends were observed but the differences were not statistically significant. However, when the moderate- and high-intensity groups were combined, once again there were statistically significant differences between the treated and untreated groups in the spousal violence $\chi^2(1, N=246)=6.74, p\leq.05$; the results indicated a 69 % reduction in spousal violence for the combined results of the two programs and a 47 % reduction in any violence. The third category, which included parole violations of any kind, did not indicate a treatment effect.

Table 4 displays the recidivism data using odds ratios, which are measures of the relative likelihood of recidivism. For example, within the moderate-intensity group, untreated offenders were 3.25 times more likely than treated offenders to commit spousal violence, and 1.57 times more likely to commit any violence. Within the high-intensity group, untreated offenders were 4.5 times more likely than treated offenders to commit spousal violence, and 2.88 times more likely to commit any violence. Overall, when the moderate- and high-intensity groups are combined, untreated offenders were 3.76 times more likely than treated offenders to commit further spousal violence.

Table 5 presents the results of a binary logistic regression that tested how much the intensity level of treatment (moderate or high) or treatment completion in either program (treated

Table 4 Recidivism odds ratios for treated versus untreated groups $N=246$

	Moderate intensity	High intensity	Combined groups
Spousal violence	3.25 ($p=.09$)	4.50 ($p=.03$)	3.76 ($p=.01$)
Any violence	1.57 ($p=.43$)	2.88 ($p=.04$)	2.06 ($p=.06$)
Any infraction	1.10 ($p=.82$)	1.89 ($p=.14$)	1.39 ($p=.27$)

versus untreated) contributed to the overall reductions in violent recidivism. The results suggest that family violence treatment, regardless of intensity level, had a significant effect upon reducing spousal violence ($OR=3.83, p=.01$). The level of significance for the individual effects in this model is associated with their Wald statistics. In other words, this model confirmed the results in Table 4 indicating that offenders who complete either one of the programs are almost 4 times less likely to commit an act of spousal violence in the follow-up. It appears that treatment completion affected spousal violence recidivism in particular, but also generalized to reduce all violence ($OR=2.18, p=.04$). The contribution of intensity level, however, was not significant. This could be interpreted as meaning treatment length and intensity level do not necessarily matter. Alternatively, the high-intensity group contained higher risk offenders. It could be, therefore, that the CSC has been successful at matching the risk and needs levels of offenders with the intensity of treatment.

Overall, the recidivism analyses suggest that the CSC high- and moderate-intensity programs are achieving the stated goal of reducing violence. The results are consistent with other evidence suggesting that treatment had a positive effect on the majority of the participants.

Factors Related to Drop Out and Success on Release

A correlational matrix and a series of t -tests examined the relationship of a number of factors the literature suggested

Table 3 Recidivism rates for treated versus untreated groups

	Moderate intensity		High intensity		Combined groups	
	Treated ($n=76$)	Untreated ($n=51$)	Treated ($n=84$)	Untreated ($n=35$)	Treated ($n=160$)	Untreated ($n=86$)
Spousal violence	4 %	12 %	4 %*	14 %	4 %*	13 %
Any violence	9 %	14 %	11 %*	26 %	10 %*	19 %
Any infraction	24 %	26 %	24 %	37 %	24 %	30 %

“Spousal Violence” is defined here as any actual, attempted or threatened violence towards a past or current intimate partner; “Any Violence” is violence (including threats) toward spouses and others; “Any Infraction” includes all criminal behavior and conditional release violations not necessarily resulting in arrest

* $p<.05$ (two-tailed); $N=246$

Table 5 Overall effect and relative effects of treatment completion and intensity level on recidivism $N=246$

Effects computed using logistic regression

	Overall effect (χ^2)	Completion (<i>OR</i>) (Yes vs. no)	Intensity (<i>OR</i>) (Moderate vs. high)
Spousal violence	6.74 ($p=.03$)	3.82 ($p=.01$)	1.10 ($p=.85$)
Any violence	4.93 ($p=.09$)	2.18 ($p=.04$)	1.58 ($p=.24$)
Any infraction	1.73 ($p=.42$)	1.43 ($p=.24$)	1.24 ($p=.47$)

would be related to drop out. Since the number of drop outs was low, those from both the moderate and high intensity programs were combined. Correlational analyses as well as t -tests showed that ethnicity (Aboriginal or Other Ethnic), IQ, antisocial or borderline personality disorder diagnosis, spousal assault risk rating (SARA), and the extent of participants' substance abuse (either drugs or alcohol) were not related to drop out. The primary demographic difference between dropouts and completers was that dropouts were significantly younger (dropouts: $M_{age}=34.20$ vs. completers: $M_{age}=37.05$; $t=9.50$, $p<.05$).

A correlational matrix and a series of t -tests examined the relationship of a number of factors to spousal violence recidivism. Given reported low base rates of reoffending the results for both the moderate- and high-intensity program participants were combined in the correlational analysis. Criminal risk rating (SIR scale score for non-Aboriginal offenders and general criminal risk rating for Aboriginal offenders), spousal violence risk rating (SARA), age, extent of substance abuse, ethnic group membership and almost all the scores on the assessment battery including Borderline Personality Disorder were not significantly related to spousal violence on release. The only variables even weakly related to spousal violence on release were an interview based rating on the extent of abusive thinking patterns ($\rho=-.25$, $p<.08$), and the measure of antisocial personality disorder symptoms ($\rho=-0.23$, $p<0.05$).

Discussion

Overall, several lines of evidence provide encouraging support for the positive impact of both the high- and moderate-intensity violence prevention programs for this offender sample. Intermediate results suggest that participants improve on almost all the measures tapping dynamic factors related to intimate partner violence, including reductions in attitudes in support of spousal abuse, improvement in skills related to conflict resolution and communication, reductions in jealousy, and a better understanding of the factors related to their offending pattern and the development of plans to address these factors. Results of content quizzes indicated that most offenders understood the content of the program, and the offenders themselves overwhelmingly stated that they found the program useful. Preliminary analyses of recidivism demonstrated that participants in the two programs were 69 % less

likely than untreated offenders to be involved in an incident related to spousal abuse following release from prison. Completion of the program also significantly reduced rates of general violence on release.

However, the program does not appear to have an impact on general criminality and suspensions related to technical violations (i.e., violations of parole conditions such as a positive urinalysis tests that are not reoffenses). Profiles of most federal offenders who have incidents related to spousal abuse in their offense histories indicate that they also have histories of general criminality. Previous research on the typology of spousal abusers within the Canadian federal system has shown that only 13.5 % would be classified as "spousal only/non-pathological" (Wexler 2000). The program content focuses specifically on skills linked to intimate relationships that would also apply to resolving or avoiding general interpersonal conflicts. There is no emphasis on avoidance of non-violent criminality. The results of this evaluation point to a need for additional interventions for spousal violence perpetrators with histories of non-violent criminality or the incorporation of material in the family violence program that would allow participant to generalize the skills and strategies to address all antisocial behavior.

At this point, the only intermediate variables even weakly related to spousal violence were antisocial personality and a general rating by the facilitators of the extent of the offenders' thinking and beliefs in support of violence against women. It should be noted, however, that we did not look at how change in the pre- and post-measures is related to outcome, which would have been more illustrative of the impact of the program (Bowen et al. 2008). As the base rates increase with longer follow-up periods, a more thorough picture of change in dynamic factors contributing to reduction in intimate partner violence may emerge.

Factors not related to treatment outcome are also noteworthy. Ethnicity was not related to any outcome variables including drop out, measures of recidivism, or almost all the intermediate measures of treatment progress. This suggests that the program as it is designed and delivered is equally effective for different ethnic groups. In addition, program outcome was not related to a diagnosis of Borderline Personality Disorder. Dutton's research had associated problematic attachment patterns related to borderline disorders with spousal violence (Dutton and Haring 1995), and had speculated that this group may have needed a differential

treatment approach. The CSC programs examined in this study, however, appear to have benefited offenders identified with higher scores on borderline symptomology as much as those with lower scores. Facilitators are trained in motivational techniques and to make reasonable adaptations for individual differences. They also provide individual sessions that supplement the group curriculum. These adaptations may have allowed the program to meet the responsivity principle by accommodating the cognitive and emotional styles of program participants with variable personality profiles. Attrition rates were low by the standards of most domestic violence programs. The only variable weakly correlated with dropping out was age. Offenders who dropped out were slightly younger than completers. Younger age is commonly associated with lower completion rates (Jewell and Wormith 2010; Olver et al. 2011). This result for these programs suggests that more up-front motivational work may be required for this sub-group.

Limitations

The design is limited by the measurement of outcome being restricted to official records that have been shown to underestimate actual spousal assault rates given that only a portion of assaults are reported to police (Statistics Canada 2011). For the period during which offenders are in the community under supervision, they are closely monitored and their families and support networks interviewed regularly, thus making it less likely that reoffending would be undetected. For those who were followed past the expiry of their sentences, we recognize the limitations of these official offense records, although the limitations apply to both the treatment and comparison groups.

Perhaps most critically, a methodological weakness of the study is including drop-outs in the comparison group. Within CSC, it is very difficult to identify a non-treatment comparison group of sufficient size for most analyses. Because of the legal mandate to address offenders' criminal behavior through program participation, most offenders participate in a number of correctional programs. Those who do not typically exhibit characteristics that would exclude them from mainstream program participation such as serious mental health problems, inability to understand the language of instruction, or refusal to be involved in the requirements of their correctional plan. However, the majority of offenders in the comparison group in this study did not complete the program for reasons not theoretically linked to outcome. Additionally, 40 % of the comparison group were identified as needing the program, but for administrative reasons (e.g., the program was not offered prior to their parole dates, they were transferred to another institution, they did not speak the language of instruction, there was a conflict with work, etc.) were not able to participate. Among those who started the program but did not

complete it, 30 % left for similar administrative reasons. Although this is a problematic design, we were able to confirm that comparisons between the outcomes for the offenders in the comparison group who left the program and those who were referred but did not attend the program indicated no differences. Further research should apply methodology such as propensity scoring to adjust for possible differences between the treatment and comparison groups that could not be attributed to program treatment effect. Statistical methods like these have been used to improve quasi experimental evaluation design by controlling for potentially significant differences between the comparison and treatment groups (Jones et al. 2004). In future analyses, a longer follow-up period would provide more robust information on the impact of program participation, although previous research has shown the 6 months after program commencement (if they are receiving the program in the community; Gondolf and Jones 2001) to be a critical period for spousal assault reoffending.

Despite disappointing outcomes from many evaluations of programs aiming to reduce intimate partner violence, this study provides converging evidence of the effectiveness of domestic violence programs designed and implemented to meet the criteria set out in the literature on effective corrections. Results indicate that cognitive-behavioral programs that respect the RNR principles and adhere rigorously to issues of program delivery integrity can be effective in reducing partner violence even among samples of highly criminalized offenders.

References

- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychological Association.
- Andrews, D., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Cincinnati, Ohio: Anderson Publishing Co.
- Babcock, J. C., Green, C. E., & Robie, C. (2004). Does batterers' treatment work? A meta-analytic review of domestic violence treatment outcome research. *Clinical Psychology Review*, 23, 1023–1053.
- Babcock, J. C., Canady, B., Graham, K. H., & Schart, L. (2007). The evolution of battering interventions: From the dark ages into the scientific Age. In J. Hamel & T. Nicholls (Eds.), *Family interventions in domestic violence: A handbook of gender-inclusive theory and treatment* (pp. 215–244). New York, NY: Springer Publishing Company.
- Bennett, L. W., Stoops, C., Call, C., & Flett, H. (2007). Program completion and re-arrest in a batterer intervention system. *Research on Social Work Practice*, 17(1), 42–54.
- Blackburn, R., & Fawcett, D. J. (1999). The Antisocial Personality Questionnaire (APQ): An inventory for assessing personality deviation in offenders. *European Journal of Psychological Assessment*, 15, 14–22.
- Bowen, E. (2011). *The rehabilitation of partner-violent men*. Malden, MA: John Wiley & Sons Ltd.
- Bowen, E., Gilchrist, E. A., & Beech, A. R. (2008). Change in treatment has no relationship with subsequent re-offending in U.K. domestic

- violence sample: A preliminary study. *International Journal of Offender Therapy and Comparative Criminology*, 52(5), 598–614.
- Brennan, S. (2009). Violent victimization of Aboriginal women in the Canadian provinces. Juristat. Statistics Canada Catalogue no. 85-002-X. Retrieved from <http://www.statcan.gc.ca/pub/85-002-x/2011001/article/11439-eng.htm>
- Brown, S. L., & Motiuk, L. L. (2005). *The Dynamic Factors Identification and Analysis (DFIA) component of the Offender Intake Assessment (OIA) process: A meta-analytic, psychometric and consultative review. Research report R-164*. Ottawa: Correctional Service of Canada.
- Buttall, F. P., & Carney, M. M. (2006). A large sample evaluation of a court-mandated batterer intervention program: Investigating differential program effect for African American and Caucasian men. *Research on Social Work Practice*, 16(2), 121–131.
- Carney, M. M., & Buttall, F. (2006). Exploring the relevance of interpersonal dependency as a treatment issue in batterer intervention. *Research on Social Work Practice*, 16(3), 276–286.
- Cavanaugh, M. M., & Gelles, R. J. (2005). The utility of male domestic violence offender typologies: New directions for research, policy, and practice. *Journal of Interpersonal Violence*, 20(2), 155–166.
- Compton, W. M., Conway, K. P., Stinson, F. S., Colliver, J. D., & Grant, B. F. (2005). Prevalence, correlates, and comorbidity of DSM-IV antisocial personality syndromes and specific drug use disorders in the united states: Results from the national epidemiologic survey on alcohol and related conditions. *Journal of Clinical Psychiatry*, 66, 677–685.
- Connors, A. D., Mills, J. F., & Gray, A. L. (2012). An evaluation of intimate partner violence intervention with incarcerated offenders. *Journal of Interpersonal Violence*, 27, 1176–1196. doi:10.1177/0886260511424499.
- Correctional Service Canada (2001). Assessment manual for the moderate and high intensity family violence prevention programs. Ottawa, ON: Reintegration Programs Division. Revised 2006.
- Daly, J. E., Power, T. G., & Gondolf, E. W. (2001). Predictors of batterer program attendance. *Journal of Interpersonal Violence*, 16(10), 971–991.
- Davis, R. C., & Taylor, B. G. (1999). Does batterer treatment reduce violence? A synthesis of the literature. *Women and Criminal Justice*, 10(2), 69–94.
- Davis, R., Taylor, B., & Maxwell, C. (1998). *Does batterer treatment reduce violence? A randomized experiment in Brooklyn (Rep. No. 94-IJ-CX-0047)*. Washington, DC: National Institute of Justice/NCJRS.
- Day, A., Chung, D., O'Leary, P., & Carson, E. (2009). Programs for men who perpetrate domestic violence: An examination of the issues underlying the effectiveness of intervention programs. *Journal of Family Violence*, 24(3), 203–212.
- Dunford, F. (2000). The San Diego Navy Experiment: An assessment of interventions for men who assault their wives. *Journal of Consulting and Clinical Psychology*, 68(3), 468–476.
- Dutton, D. G. (1995). *The domestic assault of women*. Vancouver, BC: UBC Press.
- Dutton, D. G., & Corco, K. (2006). Transforming a flawed policy: A call to revive psychology and science in domestic violence research and practice. *Aggression and Violent Behavior*, 11(5), 457–483.
- Dutton, D. G., & Haring, M. (1995). *Perpetrator personality effects on post-separation victim reactions in abusive relationships*. Vancouver, BC: UBC Press.
- Dutton, D. G., & Hart, S. D. (1992). Risk markers for family violence in a federally incarcerated population. *International Journal of Law and Psychiatry*, 15, 101–112.
- Feder, L., & Dugan, L. (2002). A test of the efficacy of court-mandated counseling for domestic violence offenders: The Broward experiment. *Justice Quarterly*, 19(2), 343–375.
- Feder, L., & Wilson, D. B. (2005). A meta-analytic review of court-mandated batterer intervention programs: Can courts affect abusers' behavior? *Journal of Experimental Criminology*, 1(2), 239–262.
- Feder, L., Wilson, D. B., & Austin, S. (2008). Court-mandated interventions for individuals convicted of domestic violence: Campbell Systematic Reviews 2008:12, The Campbell Corporation.
- Ferguson, L. M. (2004). Assessing the Correctional Service of Canada High Intensity Family Violence Prevention Program. Unpublished Master's thesis. Department of Psychology Saskatoon: University of Saskatchewan. <http://ecommons.usask.ca/handle/10388/etd-05272004-060720>. Accessed 20 Jan 2014.
- Gendreau, P. (1996). The principles of effective intervention with offenders. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 117–130). Thousand Oaks, CA: Sage Publications, Inc.
- Gondolf, E. W. (1999). A comparison of reassault rates in four batterer programs: Do court referral, program length and services matter? *Journal of Interpersonal Violence*, 14(1), 41–61.
- Gondolf, E. W., & Jones, A. (2001). The program effect of batterer programs in three cities. *Violence and Victims*, 16(6), 693–704.
- Gordon, J. A., & Moriarty, L. J. (2003). The effects of domestic violence batterer treatment on domestic violence recidivism: The Chesterfield County experience. *Criminal Justice and Behavior*, 30(1), 118–134.
- Hanson, K., & Wallace-Capretta, S. (2000). *Predicting recidivism among male batterers. Research report 2000–06*. Ottawa, ON: Department of the Solicitor General Canada.
- Harris, G. T., Hilton, N. Z., & Rice, M. E. (2011). Explaining the frequency of intimate partner violence by male perpetrators: Do attitude, relationship, and neighborhood variables add to antisociality? *Criminal Justice and Behavior*, 38(4), 309–331.
- Holtzworth-Munroe, A., & Meehan, J. C. (2004). Typologies of men who are maritally violent: Scientific and clinical implications. *Journal of Interpersonal Violence*, 19(12), 369–389.
- Hupka, R. B., & Rusch, P. A. (2001). Interpersonal relationship scale. In B. F. Perlmutter, J. Toulaitos, & G. W. Holden (Eds.), *Handbook of family measurement techniques: Instruments & index* (Vol. 3, pp. 148–150). Thousand Oaks, CA: Sage Publications.
- Jewell, I. M., & Wormith, S. J. (2010). Variables associated with attrition from domestic violence treatment programs targeting male batterers: A meta-analysis. *Criminal Justice and Behavior*, 37(10), 1086–1113.
- Jones, A., D'Agostino, R., Gondolf, E., & Heckert, A. (2004). Assessing the effect of batterer program completion on reassault using propensity scores. *Journal of Interpersonal Violence*, 19(9), 1002–1020.
- Kroner, D. G., & Weekes, J. R. (1996). Balanced Inventory of Desirable Responding: Factor structure, reliability, and validity with an offender sample. *Personality and Individual Differences*, 21, 323–333.
- Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) Guide: Reliability and validity in adult male offenders. *Law and Human Behavior*, 24(1), 101–118.
- Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1999). *Manual for the spousal assault risk assessment guide* (2nd ed.). Toronto: Multi-Health Services.
- Lanyon, R. I., & Carle, A. C. (2007). Internal and external validity scores on the balanced inventory of desirable responding and the paulhus deception scales. *Educational and Psychological Measurement*, 67, 859–876.
- Levesque, D., Gelles, R., & Velicer, W. (2000). Development and validation of a stages of change measure for men in batterer treatment. *Cognitive Therapy and Research*, 24(2), 175–199.
- Maxwell, C. D., Davis, R. C., & Taylor, B. G. (2010). The impact of length of domestic violence treatment on the patterns of subsequent intimate partner violence. *Journal of Experimental Criminology*, 6, 475–497.
- Morrison, S., & Nesius, A. (2003). Findings from an evidence-based review of batterer prevention and intervention programs. Paper presented at the 8th International Conference on Family Violence, San Diego, CA.
- Motiuk, L. L. (1997). Classification for correctional programming: The Offender Intake Assessment (OIA) process. *Forum on Corrections Research*, 9(1), 18–22.

- Motiuk, L. L., & Porporino, F. J. (1991). The prevalence, nature and severity of mental health problems among federal male inmates in Canadian penitentiaries. Report-24, Ottawa: Correctional Service Canada.
- Nafekh, M., & Motiuk, L. L. (2002). The Statistical Information on Recidivism - Revised 1 (SIR- R1) Scale: A Psychometric. Report R-126, Ottawa: Correctional Service Canada.
- Nuffield, J. (1982). *Parole decision making in Canada: Research towards decisions guidelines*. Ottawa: Department of the Solicitor General Canada.
- Oldham, J., Clarkin, J., Applebaum, A., Carr, A., Kernberg, P., Lottermen, A., & Haas, G. (1985). A self-report instrument for Borderline Personality Organization. In T. H. McGlashan (Ed.), *The Borderline: Current empirical research. The progress in psychiatry series* (pp. 3–18). Washington, D.C: American Psychiatric Press, Inc.
- Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2011). A meta-analysis of predictors of offender treatment attrition and its relationship to recidivism. *Journal of Consulting and Clinical Psychology*, 79(1), 6–21.
- Palmer, S. E., Brown, R. A., & Barrera, M. E. (1992). Group treatment program for abusive husbands: Long-term evaluation. *American Journal of Orthopsychiatry*, 62(2), 276–283.
- Paulhus, D. L. (1990). Measurement and control response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measurement of personality and social psychological attitudes* (pp. 37–41). San Diego: Academic Press.
- Shipley, W. C. (1967). Shipley Institute of Living Scale. Los Angeles, CA: Western Psychological Services. Revised by Zachary, R. A. (1991).
- Statistics Canada (2011). Family violence in Canada: A statistical profile. Catalogue no. 85–224-X.
- Stewart, L. A., Gabora, N., Kropp, P. R., & Lee, Z. (2005). *Family violence programming: Treatment outcome for Canadian federally sentenced offenders. Research Report R-174*. Ottawa: Correctional Service Canada.
- Stewart, L. A., Power, J., & Cousineau, C. (2011). *Profile and programming needs of federal offenders with histories of intimate partner violence. Research Report R-265*. Ottawa: Correctional Service Canada.
- Stover, C. S., Meadows, A., & Kaufman, J. (2009). Interventions for intimate partner violence: Review and directions for evidence based practice. *Professional Psychology: Research and Practice*, 40(3), 223–233.
- Stuart, R. B. (2005). Treatment for partner abuse: Time for a paradigm shift. *Professional Psychology: Research and Practice*, 36(3), 254–263.
- Taylor, B. G., Davis, R. C., & Maxwell, C. D. (2001). The effects of a group batterer treatment program: A randomized experiment in Brooklyn. *Justice Quarterly*, 18(1), 171–201.
- Usher, A., & Stewart, L. A. (2011). *Validation of the Generic Program Performance Measure (GPPM). Research Report R-255*. Ottawa: Correctional Service Canada.
- Wexler, A. (2000). Can typologies of male batterers be generalised to populations of federal inmates? Unpublished Masters Thesis. Ottawa: Carleton University. <http://catalogue.library.carleton.ca/search~S9?/aWexler+a/awexler+a/1%2C4%2C4%2CB/frameset&FF=awexler+audrey+f+audrey+francoise+1972&1%2C1%2C>. Accessed 20 Jan 2014.