

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**RISK FOR INTIMATE PARTNER VIOLENCE:
AN INVESTIGATION OF THE PSYCHOMETRIC PROPERTIES OF THE
SPOUSAL ASSAULT RISK ASSESSMENT GUIDE IN A NEW ZEALAND
POPULATION**

A thesis presented in partial fulfilment of the requirements for the
degree of Doctorate of Clinical Psychology

at Massey University
Wellington, New Zealand

Uvonne Callan-Bartkiw

2012

ABSTRACT

International and national studies have consistently shown intimate partner violence is a common phenomenon that cuts across all societies, education and socioeconomic levels, and ethnic and cultural groups. The impact of which includes negative physical and mental health consequences for the victims. Risk assessments may play a role in assisting the management and/or prevention of harm. Assessment of an offender's risk of future violence play a central role in decision making pertaining to that person's sentencing, community release, case management, and public safety concerns. Yet the assessments also need to ensure that the rights of the individual being assessed are not violated by misclassification. One method for addressing this issue is to ensure that risk assessment measures are accurate, that is, the measure is reliable and valid. In New Zealand to date, no intimate partner violence risk assessment tools have been evaluated. The current study, therefore, aims to fill this void by investigating the reliability and validity of the Spousal Assault Risk Assessment (SARA) guide. This was achieved in three parts, using a sample of 43 men recruited from community based stopping violence programmes. Part One evaluated the internal consistency and interrater reliability of the SARA, Part Two evaluated the convergent and discriminant validities, and Part Three, which employed a prospective design with 36 participants from the total sample, evaluated the predictive validity and incremental validity of the dynamic risk factors. The findings indicated that while the internal consistency, and convergent, discriminant, and predictive validates were adequate, the dynamic risk factors did not evidence incremental validity over the static risk factors, and the interrater reliability was variable. In addition, it was found that the source of information provided to the observers impacted on the resulting agreement coefficients. Therefore, before the SARA is implemented as a risk assessment measure in New Zealand methods for improving the interrater reliability and exploration of the usefulness of the dynamic risk factors in reducing risk should be explored.

ACKNOWLEDGEMENTS

There are a number of people who I would like to thank. First, my sincere thanks go to my supervisors. To my primary supervisor Dr Mei Wah Williams, this thesis would not be what it is today without your wonderful enthusiasm, encouragement and advice. To Associate Professor Douglas Boer, your expertise in violence risk assessment was invaluable. Thank you for sharing your knowledge and being patient with me as I developed mine. To Dr Duncan Babbage, your generosity of your time and zest towards this research was greatly valued. Thank you all for agreeing to supervise this research and for your guidance throughout.

Second, my gratitude goes to a number of people who made this study possible. To the staff of the stopping violence programmes, namely the managers and the intake assessors, thank you for both granting me permission to recruit participants at your sites and introducing the research to potential participants with such vigour and enthusiasm that the recruitment process ran more smoothly than it could otherwise have been. To the New Zealand police, especially the steering committee members, thank you for supporting this research and for your guidance, patience, and encouragement. To my research assistant, Morgan, and the independent raters, thank you for the support, time, and effort you gave to this research. To Ken McMaster, I am grateful for the time you gave to guide and support this research. Lastly, to the participants of this research, your willingness, effort, and time given was greatly appreciated, thank you.

Finally, a very special thank you to my partner, Matt, and my family, Tony, Annette, Charmiane and Brooke, this could not have been completed without your unconditional support and encouragement. Thank you for assuring me of my potential throughout this research (and the clinical components of this degree) and for assisting me in fulfilling this goal.

TABLE OF CONTENTS

Introduction and Overview	1
Chapter One: Intimate Partner Violence	3
The Prevalence of Intimate Partner Violence	4
The Health Consequences of Intimate Partner Violence	6
Etiological Frameworks	7
Feminist theory.....	8
Family systems theory.....	8
Intergenerational transmission theories.....	9
The psychology of criminal conduct.....	10
Typologies of people who engage in intimate partner violence.....	11
Current Responses to Intimate Partner Violence	13
Mandatory arrest and risk assessment police policies.....	14
Stopping violence programmes.....	16
Chapter Summary	20
Chapter Two: The Assessment of Risk for Violence	21
Risk Factors.....	21
Static risk factors.....	22
Dynamic risk factors	25
Risk Management	27
Chapter Summary	28
Chapter Three: Considerations in Violence Risk Assessment.....	29
Reliability.....	29
Internal consistency.....	30
Interrater reliability	31
Validity.....	32
Convergent and discriminant validity	33
Predictive validity	34
Incremental validity.....	35
Population validity	36
Chapter Summary	38
Chapter Four: The Spousal Assault Risk Assessment Guide	39
The Spousal Assault Risk Assessment Guide.....	39
Description of the Spousal Assault Risk Assessment guide risk factors	40
Psychometric properties of the Spousal Assault Risk Assessment guide	46
The Current Study	52
Part one: Interrater reliability	53
Part two: Convergent and discriminant validity.....	53
Part three: Predictive and incremental validity	54
Chapter Summary	55
Chapter Five: Method	56
Participants.....	56
Follow up subsample.....	59
Interviewers and observers.....	59

Measures	59
Spousal Assault Risk Assessment guide	60
Domestic Violence Screening Instrument.....	61
Violence Risk Appraisal Guide.....	62
Violent Offender Treatment Programme-Risk Assessment Scale	62
Revised Conflict Tactics Scale.....	63
Procedure	64
Planned Analyses	66
Sample size.....	66
Missing data	67
Measurement levels.....	67
Assumption checks in the total sample	68
Part one: Interrater reliability	68
Part two: Convergent and discriminant validity.....	69
Part three: Predictive and incremental validity	70
Chapter Six: Results	72
Preliminary Analyses	72
Assumption checks.....	72
Descriptive statistics.....	73
Part One: Interrater Reliability.....	74
Part Two: Convergent and Discriminant Validity	81
Convergent validity	81
Discriminant validity.....	85
Part Three: Predictive and Incremental Validity.....	85
Preliminary analyses	86
Predictive validity	87
Incremental validity.....	91
Chapter Seven: Discussion	93
Summary of the Findings.....	93
Part one: Interrater reliability	93
Part two: Convergent and discriminant validity.....	97
Part three: Predictive and incremental validity	99
Implications.....	102
Intimate partner violence risk assessment.....	102
Interrater reliability studies	105
Limitations	106
Future Directions.....	107
Conclusion	108
References.....	110
Appendix A Duluth model wheels.....	132
Appendix B Participant information and consent forms	134
Appendix C Semi-structured interview guide	139
Appendix D Participant assignment for each rater	145
Appendix E Normal probability plots in the total sample	146
Appendix F Normal probability plots in the follow up sub-sample.....	152
Appendix G Distribution of the SARA scores in the follow up sub-sample	154

LIST OF FIGURES

Figure 1	Male and female victims of self-reported intimate partner violence, by different types of violence.....	5
Figure 2	Sample size at each time point in the selection process.....	57
Figure 3	Distribution of the distance from the interviewer’s intimate partner violence subscale score for the observer’s using the audio recording alone method.....	76
Figure 4	Distribution of the distance from the interview’s total score for the three interrater reliability methods.....	78
Figure 5	Distribution of the distance from the interview’s general violence subscale score for the three interrater reliability methods.....	78
Figure 6	Distribution of the distance from the interview’s intimate partner Violence subscale score for the three interrater reliability methods.....	79
Figure 7	Distribution of the distance from the interview’s risk rating for the three interrater reliability methods.....	79
Figure 8	ROC curve of the sensitivity and specificity for the SARA total score and summary risk rating.....	88
Figure 9	Group comparison of the cumulative proportion with no police contact at 270 days, based on the total score (cut off=13).....	90
Figure 10	Group comparison of the cumulative proportion with no police contact at 270 days, based on summary risk rating.....	91

LIST OF FIGURES IN APPENDICES

Figure A1	Power and control wheel.....	132
Figure A2	Equality wheel.....	133
Figure E1	Normal probability plot for the SARA total score.....	146
Figure E2	Normal probability plot for the SARA risk rating.....	146
Figure E3	Normal probability plot for the SARA general violence subscale.....	147
Figure E4	Normal probability plot for the SARA intimate partner violence subscale.....	147
Figure E5	Normal probability plot for the Violence Risk Appraisal Guide.....	148
Figure E6	Normal probability plot for the Domestic Violence Screening Instrument.....	148
Figure E7	Normal probability plot for the Violent Offender Treatment Programme-Risk Assessment Scale.....	149
Figure E8	Normal probability plot for the Revised Conflict Tactics Scale physical assault subscale.....	149
Figure E9	Normal probability plot for the Revised Conflict Tactics Scale psychological aggression subscale.....	150
Figure E10	Normal probability plot for the Revised Conflict Tactics Scale sexual coercion subscale.....	150
Figure E11	Normal probability plot for the Revised Conflict Tactics Scale injury subscale.....	151
Figure E12	Normal probability plot for the Revised Conflict Tactics Scale negotiation subscale.....	151
Figure F1	Normal probability plot for the SARA total score.....	152
Figure F2	Normal probability plot for the SARA risk rating.....	152

Figure F3	Normal probability plot for the SARA static factors	153
Figure F4	Normal probability plot for the SARA dynamic factors.....	153
Figure G1	Distribution of the SARA total score	154
Figure G2	Distribution of the SARA summary risk rating	154
Figure G3	Distribution of the SARA general violence subscale.....	155
Figure G4	Distribution of the SARA intimate partner violence subscale	155

LIST OF TABLES

Table 1	Proposed subtypes of male intimate partner violence perpetrators: How they differ on descriptive dimensions	12
Table 2	SARA predictive validity studies.....	50
Table 3	Summary of demographic characteristics for the total and follow-up participants	58
Table 4	Reliability analyses for the Revised Conflict Tactics Scale subscales	64
Table 5	Descriptive statistics for the measures in the total sample (<i>N</i> =43)	74
Table 6	Intraclass correlation coefficients between the SARA scores using different methods of agreement assessment.....	76
Table 7	Differences in the mean discrepancies of rater 1, rater 2, and rater 3, for the SARA total and subscale score.....	80
Table 8	Differences in the mean discrepancies of rater 1, rater 2, and rater 3, for the SARA risk ratings.....	81
Table 9	Bivariate correlations between SARA scores	82
Table 10	Bivariate correlations between SARA scores and Domestic Violence Screening Instrument	82
Table 11	Bivariate correlations between SARA subscales, Violence Risk Appraisal Guide and Violent Offender Treatment Programme-Risk Assessment Scale	83
Table 12	Bivariate correlations between SARA total, SARA intimate partner violence subscale, SARA risk rating, and Revised Conflict Tactics Scale violent resolution subscales.....	84
Table 13	Bivariate correlations between SARA scores and Revised Conflict Tactics Scale negotiation subscale.....	85
Table 14	Descriptive statistics for the SARA total, risk rating, static factors, and dynamic factors in the follow-up subsample (<i>n</i> =36).....	87

LIST OF TABLES IN APPENDICES

Table D1	Participant assignment for each rater	145
----------	---------------------------------------------	-----

INTRODUCTION AND OVERVIEW

Intimate partner violence is a threat to public health and a significant social issue which, because of the high prevalence rates, has been described as a pressing human rights issue (Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002; Tolan, Gorman-Smith, & Henry, 2006). One method for addressing intimate partner violence involves assessing the risk posed by perpetrators so that management and treatment interventions can be implemented. Despite the overarching aim of risk assessment to minimise harm towards others, risk assessment involves the balancing of public safety and the rights of the person being assessed (Douglas & Kropp, 2002). To determine if a particular measure is addressing this balance the psychometric properties should be evaluated. Currently in New Zealand risk for intimate partner violence is based on either unstructured professional judgement or structured measures that have not been proven reliable and valid within the population. Therefore, it is unknown if New Zealand practice adequately maximises public safety while minimising misclassification. This research proposes to fill this void by exploring the reliability and validity of the Spousal Assault Risk Assessment guide (SARA; Kropp, Hart, Webster, & Eaves, 1994, 1995, 1999), with specific focus on the interrater reliability, and convergent, discriminant, predictive, and incremental validities.

The thesis is divided into seven chapters. Chapters One through Four review the relevant literature. Chapter One provides an overview of the literature on intimate partner violence, including the international and national prevalence rates, health consequences, prominent etiological frameworks, and current New Zealand police and stopping violence programmes responses to intimate partner violence. Chapter Two explores the role of risk factors in current violence risk assessment modalities. Specifically, it focuses on the utility of structured professional judgements in guiding risk management. Chapter Three examines the considerations in reliability and validity investigations of violence risk assessment measures, specifically actuarial and structured professional judgements. Chapter Four explores the risk factors and international reliability and validity studies pertaining to the SARA. This chapter concludes with a discussion of the hypotheses for the current study. Chapter Five provides an overview of the methodology of the current study, including the participants, interviewers and observers, measures, procedure, and planned analyses. Chapter Five presents the results of the current study. Chapter Seven discusses the

results, in the context of the literature and provides implications of the current study, limitations, and recommendations for further studies.

CHAPTER ONE

INTIMATE PARTNER VIOLENCE

Intimate partner violence is a serious problem that cuts across all societies, educational and socioeconomic levels, and ethnic and cultural groups (Hartman, Janes, & Troy, 2009). Many families in New Zealand are affected by family violence (Grant, 2009), with rates of physical and sexual violence perpetrated by intimate partners approximately twice that of violence perpetrated by non-partners (Fanslow & Robinson, 2004). Intimate partner violence has been defined as “any actual, attempted, or threatened physical harm perpetrated by a man or woman against someone with whom he or she has, or has had, an intimate, sexual relationship” (Kropp, Hart, Webster, & Eaves, 1995, p. 1). In addition to physical harm, it is widely recognised that intimate partner violence can take the form of psychological, economic, verbal, sexual, and/or spiritual abuse (Fanslow, 2005; Mowat-Leger, 2001).

While intimate partner violence was historically seen as exclusively being exerted by men towards their female partners, current perspectives acknowledge symmetry in the gender of perpetrators (Fanslow, 2005). Men and women initiate violent behaviour at almost equal rates, with up to 83 percent of all intimate partner violence being mutual (Langhinrichsen-Rohling, Neidig, & Thorn, 1995; Moffitt & Caspi, 1999; Morse, 1995; Straus, 2004). In addition, it has been found that the prevalence rates and dynamics of intimate partner violence in homosexual relationships are similar to those of heterosexual relationships (see Kulkin, Williams, Borne, de la Bretonne, & Laurendine, 2007). Despite this, the following literature review focuses primarily on intimate partner violence perpetrated by males towards females. The reasons for this are twofold: the majority of the literature is asymmetrical (Kulkin et al., 2007), and the current validation study was conducted with a male sample.

The present chapter begins with an overview of the international and domestic prevalence rates of intimate partner violence and the associated physical and mental health consequences experienced by the victims. This is followed by a discussion of the prominent etiological frameworks, each of which, it is found, does not account for the complexity of intimate partner violence. The chapter concludes with a discussion of the current police mandatory arrest and risk assessment policies and the stopping violence programme approaches employed in New Zealand.

The Prevalence of Intimate Partner Violence

International studies into the prevalence of intimate partner violence have consistently shown that it is a common phenomenon (Mowat-Leger, 2001). Breiding, Black and Ryan (2008) surveyed over 70,000 non-institutionalised adults in sixteen American states, Puerto Rico and the Virgin Islands as part of an ongoing surveillance of health behaviours. They found that 26.4 percent ($n = 11,552$) of women and 15.9 percent ($n = 4,175$) of men had reported physical and/or sexual intimate partner violence during their lifetime. The twelve month prevalence rate was 1.4 percent ($n = 588$) of women and 0.7 percent ($n = 166$) of men. In addition they found that, compared to non-Hispanic white women, multicultural women were significantly more likely to report lifetime intimate partner violence, while Asian and Hispanic women were significantly less likely to report intimate partner violence. The findings were similar for men, with the addition of black non-Hispanic men significantly more likely to report lifetime intimate partner violence, and Native Hawaii/Pacific Island men significantly less likely to report intimate partner violence.

Similar studies have reaffirmed the high rates of intimate partner violence. Forty-one percent of women sampled from general practices in London self-reported they had experienced intimate partner violence in their lifetime in a study by Richardson et al. (2002). Walby (2004) reported a twelve month prevalence of intimate partner violence in 3.4 percent ($n = 764$) of women and 2.2 percent ($n = 494$) of men. A study of 3,429 women in the United States found that 14.7 percent of the participants had reported any type of intimate partner violence in the five years prior to being surveyed, and that 45.1 percent of these women had reported more than one type of violence (Thompson et al., 2006).

A recent study in Canada evaluating family violence (Statistics Canada, 2011) provided insight into the prevalence of intimate partner violence severity. Men and women were surveyed about physical, sexual, and psychological violence they had reported over a five year period. The majority of female victims reported being pushed, grabbed, shoved, and slapped (37%) while males reported being kicked, punched, hit, or hit with something (36%). Figure 1 (Statistics Canada, 2011, p. 12) outlines the proportions of people reporting various types of violence within their intimate relationship. A Spanish study by Zorrilla et al. (2009) found that 8.6 percent of the 2,136 women surveyed had reported psychological abuse, 2.4 percent physical violence, and 1.1 percent sexual violence during a twelve month period.

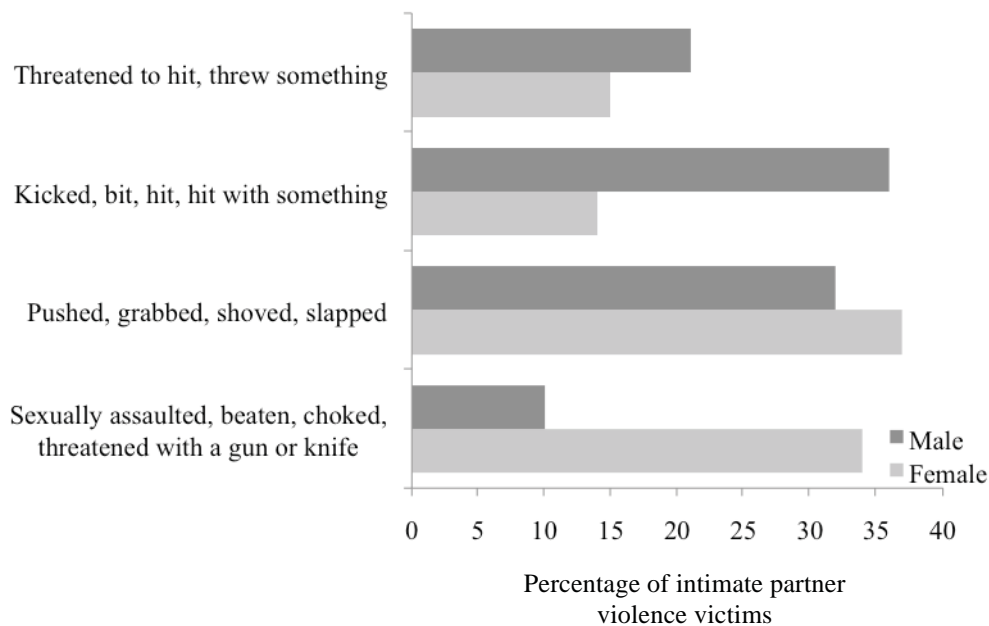


Figure 1. Male and female victims of self-reported intimate partner violence, by different types of violence (Statistics Canada, 2011).

In response to criticism that many of the international studies had been undertaken with westernised samples the World Health Organisation investigated the prevalence rates in ten countries in Asia, Africa, South America, Eastern Europe, and the Pacific (Garcia-Moreno, Jansen, Ellsberg, Heise, & Watts, 2006). Females, aged between 15 and 49, who had been in an intimate relationship ($n = 24,097$) were sampled from fifteen large city and rural sites. The study found lifetime and twelve month prevalence rates of physical and/or sexual intimate partner violence varied widely (between 15 and 71, and 3.4 and 53.7 percent, respectively). Overall, those living rurally reported more intimate partner violence than those living in cities.

Prevalence studies undertaken with New Zealand samples indicate that the rates of intimate partner violence are similar to those of other industrialised countries. In 2003 the Ministry of Justice commissioned a study of victimisation in New Zealand (Morris, Reilly, Berry, & Ransom, 2003). Fifty-three thousand people were randomly sampled throughout the country. The study found that 26.4 percent ($n = 2,526$) of women had reported intimate partner violence in their lifetime. The twelve month prevalence rate was three percent ($n = 1,606$). Fanslow and Robinson (2004) reported physical and/or sexual intimate partner violence in 33 and 39 percent of randomly surveyed women in the Auckland and Waikato districts, respectively.

Studies demonstrate the prevalence of intimate partner violence victimisation and perpetration in Māori populations is consistently higher than that of non-Māori populations. Marie, Fergusson and Boden (2008) sampled 804 participants from the Christchurch birth cohort study who reported being in an intimate relationship lasting more than one month between the ages of 24 and 25 (64% of the original sample). The authors found that, compared to non-Māori, Māori were significantly more likely report having experienced psychological aggression and physical assault in an intimate relationship. In addition, Māori were 2.9 times more likely to report severe physical assault, 3.4 times more likely to have been injured as a result of intimate partner violence, and 5.9 times more likely to fear their partner. Finally, a study by Koziol-McLain, Rameka, Giddings, Fyfe, and Gardiner, (2007) found that 23 percent of the women who attended their Māori health provider general practice clinic reported current intimate partner violence.

The Health Consequences of Intimate Partner Violence

Numerous studies attest to negative physical and mental health consequences for the victims of intimate partner violence. At the extreme it can result in death, through either suicide or homicide. Fanslow (2005) found that women exposed to intimate partner violence are at higher risk for cuts and bruising, fractures, reduced physical functioning, and traumatic brain injury. Women who reported severe physical violence were 1.9 times more likely to be unable to perform usual activities than those who reported no violence (Fanslow & Robinson, 2004). In addition, female victims of intimate partner violence are also likely to have sexual and reproductive health difficulties (Martin, Taft, & Resick, 2007). Common psychological consequences include low self-esteem, depression, anxiety, insomnia, and social dysfunction (Fergusson, Horwood, & Ridder, 2005; Ratner, 1993).

Suicidal ideation and suicide attempts are associated with intimate partner violence. A study of 2,855 women in New Zealand (Fanslow & Robinson, 2004) found one half ($n = 289$) of those who experienced severe violence reported suicidal thoughts, and one fifth ($n = 115$) had attempted suicide. Compared to women who had not reported intimate partner violence, it was found that these women were 7.6 times more likely to have attempted suicide. Of the 299 people who reported moderate violence, 7.5 percent had attempted suicide. It has been suggested that relationship between suicide

and intimate partner violence is mediated by psychological distress, hopelessness, and drug use (Kaslow et al., 1998).

In New Zealand more than one intimate partner is murdered, on average, every three months. Fifty-two percent ($n = 74$) of the 141 homicides that occurred between 2002 and 2006 involved an intimate partner (Martin & Pritchard, 2010). In total 93 people, including intimate partners, perpetrators, children, and new partners died. Two thirds of the perpetrators had had previous contact with the police, with 48 percent having prior documented incidents of intimate partner violence and 24 percent having protection orders imposed. However, given the relatively low base rate of intimate partner violence related homicides, compared to incidents of intimate partner violence, Martin and Pritchard (2010) stated that “any approach to prevention must [...] focus on the reduction or elimination of domestic violence itself, rather than on attempting to identify risks that are specific to lethal violence” (p. 40).

In addition to the impact to the intimate partner in violent environments, there is also evidence of adverse health effects in the children. McCue (2008) reported that children who witnessed intimate partner violence are at risk for both internalising and externalising behavioural and emotional disturbances, including anxiety, depression, a lack of empathy, somatic complaints, aggression, and developmental regression. Furthermore, children are at risk of violence directed towards them. Intimate partner violence often predates violence towards children, with more frequent intimate partner violence associated with an increased likelihood that children are abused (Tjaden & Thoennes, 1998).

Etiological Frameworks

A number of etiological theories have been proposed to explain intimate partner violence. While early theories were dominated by psychiatric explanations that postulated the behaviour was driven by psychological dependency, brain lesions, or sadistic characteristics, later theories have been driven by an increased awareness of the prevalence of intimate partner violence. Psychiatric disturbance alone cannot account for the high rates (La Taillade & Jacobson, 1997). Some of the prominent traditional and contemporary theories that attempt to explain intimate partner violence have been included in the current review (Andrews & Bonta, 2010; Dobash & Dobash, 1977; Dutton, 2003; Lewis & Fremouw, 2001).

Feminist theory

According to feminist perspectives violence against women is socially constructed. Gender inequality in patriarchal dominant societies is taught to both males and females in childhood and results in male advantage in adulthood (Dobash & Dobash, 1977; Mihalic & Elliot, 1997). In order to maintain this advantage males engage in many tactics, one of which is violence (Dobash & Dobash, 1977). McCue (2008) noted that because change requires the family unit to be restructured, the unit is ultimately strengthened. Evidential research has come from descriptive and correlational studies that explored relationships between attitudes held by male participants and self-reported intimate partner violence. For example, Tokar and Fischer (1996) interviewed 106 male university students and found that negative attitudes towards feminism were strongly related to attitudes that accepted violence in dating situations. Similarly, Kantor, Jasinski, and Aldarondo (1994) reported that men who held attitudes accepting physical aggression towards females were 2.2 times more likely to engage in intimate partner violence.

A major criticism of this theory is that it narrowly focuses on male violence towards female victims. Several studies have shown that people in same sex relationships reported intimate partner violence at rates similar to those in heterosexual relationships (Pitts, Smith, Mitchell, & Patel, 2006; Donovan, Hester, Holmes, & McCarry, 2006) and that females are also perpetrators (Dutton, 2007; Dutton & Corvo, 2007). Second, Dutton (2006) questions the model's assumption that patriarchy causes violence. It fails to account for both the high incidence of psychopathology amongst perpetrators and why prevalence is not 100 percent. Third, the societal perspective does not account for individual differences (Andrews & Bonta, 2010). Finally, societal perspectives render individual and group therapy void: men are acting as society dictates (Dutton, 2006).

Family systems theory

Family systems theorists view families as a dynamic organisation made up of interdependent components, which are intricately connected to one another (McCue, 2008). The interactions between family members cause and maintain behaviours, including intimate partner violence, by reinforcing and regulating the system. Specifically, the family's emphasis on intimacy, privacy, and ascribed sex roles contribute to the maintenance and transmission across generations (La Taillade &

Jacobson, 1997). As a result, intimate partner violence accountability is dispersed among the family members (La Taillade & Jacobson, 1997). Despite the fact that family systems theory underlies some current interventions, such as couples counselling, it has been criticised for being too narrow in view. It ignores other empirically supported causal relationships, including negative attitudes towards women, personality style, and societal pressure. Additionally, McCue (2008) noted that the family system theory inherently leads to some responsibility for the violence being placed with the victim.

Intergenerational transmission theories

The intergenerational transmission of violence was first proposed by Curtis (1963) in his paper *Violence Breeds Violence—Perhaps?*, with the author noting his concern that abused children would become violent criminals in adulthood. Within the intimate partner violence field, specifically, much attention has been given to the relationship between experiencing and/or witnessing family of origin violence and perpetrating intimate partner violence as an adult, which is attributed to learning and developmental mechanisms. Males who engage in intimate partner violence are more likely to have been exposed to family of origin violence compared to non-violent males (Delsol & Margolin, 2004; Whitfield, Anda, Dube & Felitti, 2003), males who are dissatisfied in their relationship (Nelson & Wampler, 2000), or males who perpetrate other crimes (Dutton & Hart, 1992b).

The most common theoretical explanation of intergenerational transmission is social learning theory (Delsol & Margolin, 2004). Based on Bandura's (1971, 1973) causal model of aggression, this theory states that intimate partner violence is modelled in childhood and that this, in turn, results in the development of schemas that intimate partner violence is tolerable and acceptable (Lewis & Fremouw, 2001). During childhood males learn that acting violently towards an intimate partner is acceptable and females learn that being victimised is equally normal (McCue, 2008). Violent behaviour is then maintained throughout adulthood if it is positively reinforced in dating or early relationships (Mihalic & Elliott, 1997; Riggs, Caulfield, & Street, 2000) and the person holds schema that the victim deserves physical punishment (Ganley, 1989). Social learning theory has been criticised as being too simplistic; empirical support indicates that transmission rates are highly variable and that most children who witness intimate partner violence do not go on to engage in the behaviour (Dutton, 2007).

In response to the flaws of social learning theory, Dutton (2003) developed the

attachment disruption model. This model states that family of origin violence impacts on a person's developmental environment via three pathways. First, the child is likely to have an insecure attachment style resulting from the victimised parent's experience of shame and trauma. This parental experience leads to them being unable to provide a secure base with which to teach the child adaptive ways of relating to others, which is perceived by the child as rejection. Second, the environment results in the child developing posttraumatic stress disorder symptoms, including feelings of shame, guilt, and being under threat. Third, the child likely develops a borderline personality style, whereby they are angry and impulsive, have an unstable sense of self, and cannot form meaningful and stable relationships. It is argued that the result of these three developmental pathways is a fearful attachment style in adulthood (Dutton, 2003). These adults have a desire for intimate social contact, while fearing rejection and distrusting others. Within this context, and in combination with both intense anger and a desire to control their partner, intimate partner violence occurs (Dutton, 2003).

Using attachment theory as a conceptual framework, Godbout, Dutton, Lussier, and Sabourin (2009) investigated the effects of family of origin violence on both current intimate partner violence and marital adjustment. The sample consisted of 644 adults who were in long-term relationships (mean = 7 years). They found that family of origin violence predicted both the development of an insecure attachment style and current intimate partner violence. In addition, they found that attachment behaviours predicted intimate partner violence, and that both insecure attachment and intimate partner violence resulted in marital dissatisfaction. Despite their findings that attachment style significantly mediated the relationship between family of origin violence and intimate partner violence, the authors noted that early experiences of intimate partner violence might also directly affect later intimate partner violence. Overall, the evidence suggests that this model alone does not account for all intimate partner violence.

The psychology of criminal conduct

Andrews and Bonta (2010) proposed a comprehensive theory of general criminal behaviour that postulates all criminal behaviour is the result of distinct patterns of social and individual factors, which are unique to each person. Criminal activity is driven by the interaction between the person's immediate environment and personal factors. As such, the theory is sensitive to individual differences and to changes in people's behaviour and environment. The primary focus of this theory is to explain

individual difference in criminal behaviour. Despite the individualistic nature of the theory, Andrews and Bonta (2010) postulate that four variables are central to the prediction of criminal behaviour. These variables, which are theoretically based on past meta-analyses include past criminal behaviour, antisocial attitudes, antisocial associates, and antisocial personality. In addition to being predictive variables, Andrews and Bonta (2010) stated that they are causal. That is, they cause criminal behaviour. This general theory can be applied to specific types of criminal behaviour, including sexual offending and intimate partner violence (Scott, 1995). Therefore people with a “history of intimate partner violence, attitudes that condone intimate partner violence, personality disorder traits, and peers that condone such behaviour” are likely to engage in intimate partner violence (Mowat-Leger, 2001, p. 13-14).

Typologies of people who engage in intimate partner violence

It has been acknowledged that people who engage in intimate partner violence are a diverse and heterogeneous group, with different typologies engaging in different forms of violence and having unique factors driving that behaviour (Holtzworth-Munroe & Stuart, 1994; La Taillade & Jacobson, 1997). For example, based on a review of the literature, which utilised clinical samples, Holtzworth-Munroe and Stuart (1994) proposed a conceptual three-fold typology. It included men who are violent only in the family setting and do not exhibit severe pathology (family only), passive aggressive and dependent men who are clingy and extremely controlling in intimate relationships (dysphoric/borderline), and men who exhibit antisocial personality traits and engage in antisocial behaviour both inside and outside the family setting (generally violence/antisocial). The differences between these types on descriptive dimensions are presented in Table 1 (p. 12).

A later testing of the proposed typology by the authors identified four clusters (Holtzworth-Munroe, Meehan, Herron, Rehman, & Stuart, 2000). Intimate partner violent men ($n = 102$) and two comparison non-violent groups (maritally distressed and non-distressed) were recruited from the community. Two of the three types (dysphoric/borderline and generally violent/antisocial) resembled those hypothesised by Holtzworth-Munroe and Stuart (1994), while the third (family only) was less violent and pathological than proposed. The fourth type (low level antisocial) most closely resembled the family only group originally proposed. The authors suggested that the addition of a fourth type reflected the community based population sampled. The

original typology was based on a review of the literature that sampled from clinical populations.

Table 1

Proposed subtypes of male intimate partner violence perpetrators: How they differ on descriptive dimensions (Holtzworth-Munrone & Stuart, 1994).

Dimension	Family-only	Dysphoric/ borderline	Generally violent/ antisocial
Marital violence			
Severity of marital violence	Low	Moderate-high	Moderate-high
Psychological and sexual abuse	Low	Moderate-high	Moderate-high
Generality of violence			
Extrafamilial violence	Low	Low-moderate	High
Criminal behaviour, legal involvement	Low	Low-moderate	High
Psychopathology/personality disorder			
Personality disorder	None or passive/dependent	Borderline or schizoid	Antisocial/psychopathy
Alcohol/drug abuse	Low-moderate	Moderate	High
Depression	Low-moderate	High	Low
Anger	Moderate	High	Moderate

One practical difficulty that has emerged from the typological literature is that the types tend to overlap. That is, rather than either dichotomously endorsing or not endorsing specific traits, the traits exist on a continuum. For example, the generally violent/antisocial group tend to have lower levels of depression than the dysphoric/borderline group, while the family only group falls somewhere in between. This may explain why some authors have had difficulty distinguishing the groups along certain traits. In a comparison of men with a history of intimate partner violence on various measures, including the Millon Clinical Multiaxial Inventory-Second Edition (Millon, 1987), Waltz, Babcock, Jacobson, and Gottman (2000) found that depression was highest for their dysphoric/borderline group, compared to the generally violent and

family only violence groups, though this difference did not reach statistical significance.

It is apparent that no one theory adequately captures the complexity of intimate partner violence; each identifies a subset of etiological variables in isolation (Bell & Naugle, 2008). In addition the feminist, family systems, and intergenerational transmission theories treat the population as homogeneous, which contradicts the heterogeneity found in recent studies (see for a review see Kelly & Johnson, 2008). According to Andrews and Bonta (2010) one of the biggest difficulties in accounting for all forms of violence, including intimate partner violence, is that of individual differences.

Current Responses to Intimate Partner Violence

New Zealand has seen a dramatic transformation in the way intimate partner violence is viewed and responded to in the last 25 years. These changes were brought about by a number of factors, which drew attention to the scale of intimate partner violence directed at women, and research into the effectiveness of different response techniques (Newbold & Cross, 2008). It was the women's movement of the 1960s that first brought to the forefront many issues affecting women, including intimate partner violence. The movement also resulted in the establishment of the first women's refuge in 1973. However, responses remained inadequate, with little legal involvement and a minimalist approach to treatment (Newbold & Cross, 2008). Intimate partner violence was considered neither a crime nor a serious social problem. By the early 1980s this position had started to shift, with two high profile intimate partner violence homicides in New Zealand highlighting that intimate partner violence was not confined to the working class and that the police and the courts had little power to protect potential victims (Newbold & Cross, 2008). In 1982, the introduction of the Domestic Protection Act allowed preventative steps to be taken by the police, by empowering them to hold perpetrators without charge for 24 hours. Laying charges, however, was still at the discretion of the police, with caution urged (Newbold & Cross, 2008).

A further shift occurred with the publication of the *Commissioner's Circular (11)* in 1987 with the introduction of a pro-arrest policy in the New Zealand police (Carswell, 2006). This new policy reflected several international and national studies, which had shown that arresting the perpetrators of intimate partner violence resulted in a decrease in future violence (Ford, 1985; Mugford & Mugford, 1992; Sherman & Berk, 1984). It signalled a change in the way intimate partner violence was handled, it was

now seen as a criminal act in line with other violent crimes (Newbold & Cross, 2008). However, as found in other countries, the policy was implemented haphazardly in New Zealand and discretion was still widely employed by frontline police officers (Newbold & Cross, 2008). In light of this unsatisfactory outcome the Hamilton Abuse Intervention Pilot Project was launched in 1991. The onus of this programme was to implement both an interagency approach, involving the police, courts, and victim support agencies, and community based stopping violence programmes for the perpetrators. In addition, the pro-arrest policy was redefined, allowing officers to make an arrest when evidence of intimate partner violence existed, regardless of whether an official complaint was made.

In 1995 the Domestic Protection Act 1982 was repealed and replaced by the Domestic Violence Act. The new act was seen as a significant step towards addressing intimate partner violence (Newbold & Cross, 2008). It brought into effect the protection order, expanded the definition of “domestic violence” to include psychological abuse, broadened the notion of domestic relationships to include flatmates, couples not living together, and homosexual couples, and mandated that all arrested offenders must be charged. Given the requirements of the new Act and further research suggesting caution was still being used inappropriately, the police policy was updated in 1996. It includes three new initiatives: that children witnessing intimate partner violence may need protection, that incident forms are completed correctly and multi-agency liaison occurs, and that when an arrest is not made the officer consults with his or her supervisor before taking action (Newbold & Cross, 2008). In addition, the focus of the police shifted to address intimate partner violence specifically, rather than the more broadly defined family violence. At present the primary aim of responses in New Zealand is to protect the victims, by preventing and reducing intimate partner violence through a multiple agency response. Of these agencies, the police and stopping violence programmes are two that work directly with the perpetrators. Therefore, their current policies and practices will be discussed.

Mandatory arrest and risk assessment police policies

As discussed, the publication of the *Commissioner's Circular (11)* in 1987 signalled the implementation of pro-arrest policy towards intimate partner violence incidents (Carswell, 2006). It was mandated that arrests be made when there was evidence of either violence or breach of a protection order. However, recent research has found that this policy has been applied haphazardly. Cross and Newbold (2010)

investigated the dynamics behind police officer discretion used in regards to the pro-arrest policy. The study sampled 73 frontline police officers and police managers in Christchurch, New Zealand, over a 24 month period between 2004 and 2006. The authors found that of the 25 incidents police attended where injuries were present, arrests were made in 44 percent of the cases, while 94 percent of the incidents with no injuries ($n = 47$) resulted in no arrest. Further qualitative analyses found that police officers' rely on discretion, with an arrest more likely when the perpetrator presented as angry or agitated upon arrival. One frontline officer remarked that arresting a perpetrator is "all about circumstances [...] it has to be a positive outcome" (Cross & Newbold, 2010, p. 63). However, a conciliatory approach may result in further intimate partner violence because arrest deters future violence (Stark, 2007). Despite their findings Cross and Newbold (2010) noted that changes in policy have been employed since their data collection, including the encouragement of dual arrest and the removal of discretion where the violence is serious or a protection order has been breached. Dual arrest refers to the arrest of all parties involved in the incident. The effectiveness of these changes on reducing intimate partner violence still requires analysis (Cross & Newbold, 2010).

The current police policy includes the compulsory use of risk assessment to assist with monitoring intimate partner violence (Cross & Newbold, 2010). In line with international trends, the New Zealand police developed a family violence risk assessment model that contains three structured risk assessment instruments, each of which is designed to determine the risk of future murder or serious harm (Grant, 2009). The assessments, which were implemented nationwide in 2008, are based on Campbell's Danger Assessment Scale (Campbell, 1986) and a review of the international literature on risk assessment and predictors of intimate partner violence homicide (Grant, 2009). The first measure, the Risk Assessment Questions for Adult Victims, contains three questions on the frequency, seriousness, and safety concerns, for officers to ask victims. The second measure, the Red Flags Risk Factors Assessments, contains 12 risk factors that are designed to predict lethality. The final measure, the Risk and Lethality Assessment Worksheet, is a 31 item checklist of risk markers of serious and lethal violence that are designed to assign risk, ranging from no risk to extreme risk. All of the assessments are completed by frontline police officers at the scene of the incident.

Despite the compulsory nature of this policy, research has shown a discrepancy

between it and operational practice (Grant, 2009). Cross and Newbold (2010) found that the assessment was completed in just over half of the cases surveyed. Many frontline police officers do not use the risk assessment as it was intended because of inadequate training and a lack of understanding of the purpose of risk assessment, generally, and the policy, specifically (Grant, 2009). Until the policy is implemented as intended its efficacy in assessing and managing intimate partner violence remains unclear. Regardless of this, the current risk assessment has not been empirically validated (Grant, 2009) therefore, it is unknown if this measure accurately assesses the risk posed by the perpetrators of intimate partner violence.

Stopping violence programmes

All people convicted of an offence involving family violence (including intimate partner violence) are required, under the Domestic Violence Act 1995, to attend a stopping violence programme. Two of the modalities, the Duluth model and cognitive behavioural therapy, are discussed below. This is followed by a review of the effectiveness of the stopping violence programmes.

The Duluth model was first implemented in Duluth, Minnesota, in 1981 as an experimental programme designed to address the escalation in family violence within the community (Mederos, 1999). It is a psychoeducational approach that views violence in intimate relationships as an enduring pattern of behaviour utilised by men to exert power and control (Pence & Paymar, 1993). The model is a community wide intervention that aims to stop violence by engaging various agencies, including victims' services, men's stopping violence programmes, criminal justice systems, and child protection systems (McMaster & Gregory, 2003). In addition, the model advocates that the agencies work in a collaborative manner to achieve the overall goal, protecting the victims (Domestic Abuse Intervention Programmes, 2011). Within New Zealand this model has influenced both the current interagency approach (discussed above) and the curriculum of the stopping violence programmes.

The Duluth model, in general, is underpinned by five outcome objectives: it is expected that each participant accepts that their violence is used to control their partner, that they understand the cultural and social contexts in which violence is used, that they are willing to change, that there are alternative ways of communicating with an intimate partner, and that the accountability and responsibility for violence lies with the male perpetrator. In line with these objectives, the curriculum of the stopping violence

programme follows eight specific themes, outlined in the *power and control wheel* (see Appendix A). The effects and consequences of each theme are discussed and compared to alternative non-controlling behaviours (on the *equality wheel*, see Appendix A). For example, using coercion and threats to maintain power and control in authoritarian and destructive relationships is replaced with negotiation and fairness, which are behaviours that reflect gender equality and result in egalitarian relationships. Change is achieved through role-plays, vignette discussion, and control logs (Pence & Paymar, 1993).

As noted, the Duluth model has greatly influenced New Zealand stopping violence programmes (McMaster & Gregory, 2003). It was originally piloted in Hamilton as the Hamilton Abuse Intervention Pilot Programme. While the model as a whole was implemented as intended, four significant changes were made to the stopping violence educational programme (Robertson, Busch, Ave, & Balzer 1991). First, Māori groups were set up, with a curriculum that acknowledged the impact of colonisation on the culture. Second, the general programme curriculum was adapted for New Zealand conditions, with locally produced video vignettes and homework sheets reworded. Third, instead of two male group facilitators one man and one woman ran the groups, which Robertson et al. (1991) noted reduced the likelihood of male collusion. Finally, group induction was completed in a group session, rather than individually.

Cognitive behavioural approaches are also widely used (Feder & Wilson, 2005) and have influenced stopping violence programmes (McMaster & Gregory, 2003). These are based on cognitive behaviour therapy, which posits that maladaptive behaviours and negative moods can be improved by challenging dysfunctional thoughts (Beck, 1995). In the intimate partner violence field, specifically, participants learn that violence is a predictable pattern of behaving which can be modified through various cognitive and behavioural techniques (Feder & Wilson, 2005). As a result, the primary focus is violent behaviour at the individual level, rather than patriarchal power and control at the societal level, as in the Duluth model (Babcock, Green, & Robie, 2004). The goal is to modify thinking patterns and behaviours (Feder & Wilson, 2005). This goal is achieved by undermining the beliefs and assumptions that maintain intimate partner violence, such as overevaluations of violence as an effective problem solving technique and negatively biased attributions about partner behaviour (Murphy & Eckhardt, 2005). This is achieved through cognitive restructuring. In addition, emotion regulation techniques aim to reduce anger arousal through relaxation training and other anger management techniques (Murphy & Eckhardt, 2005). Finally, other commonly

used cognitive behaviour techniques include social skills training, such as nonviolent assertiveness, and relapse prevention (Babcock et al., 2004).

Although differences exist between these modalities, most of the current stopping violence programmes incorporate both psychoeducational and cognitive behavioural components, which are presented within a feminist framework (Babcock et al., 2004; Tolman & Edleson, 1995), and delivered in a group format (Babcock et al., 2004). This holds true for New Zealand (McMaster & Gregory, 2003). For example, most programmes that purport to follow the Duluth model address the notion that violence is learned and maintained through reinforcement and most stopping violence cognitive behavioural therapy programmes address patriarchal attitudes in the use of violence towards women (Babcock et al., 2004). Furthermore, the New Zealand stopping violence programme good practice guidelines, outlined by Robertson (1999), recommend that components of both modalities be utilised. Specifically, the guidelines include incorporating cognitive behavioural techniques within an “explicitly feminist analysis of battering” (p. 70). As such the programmes are considered educational, rather than therapeutic.

These individual components and the current stopping violence programmes, which use techniques from both modalities, are not without their critics. The Duluth model has been criticised for its confrontation nature in emphasising the attendee’s negative beliefs—seen as colluding with victim blaming (Dutton, 2007). This, in turn, fails to promote a working alliance and other non-specific conditions, which foster therapeutic change (Murphy & Baxter, 1997). It has been shown that working alliance predicts stopping violence programme outcome, with stronger working alliance leading to a reduction in intimate partner violence (Taft, Murphy, King, Musser, & DeDeyn, 2003). The cognitive behavioural approach has been criticised for allowing perpetrators to use anger as an excuse for intimate partner violence (Murphy & Eckhardt, 2005), and neglecting societal issues of power and control by focusing on the individual perpetrator (Adams, 1988). Overall, the current programmes have been criticised for their one size fits all approach, which does not take into account both individual differences and the research conducted that identifies intimate partner violence perpetrator typologies (see etiological theories subsection of this chapter for a discussion). It appears that the current stopping violence programmes are targeting the violence perpetrated by the family-only type, rather than all of the types (Clements, Holtzworth-Munroe, Gondolf, & Meehan 2002).

The effectiveness of stopping violence programmes. Two meta-analyses have explored the effectiveness of treatment programmes in reducing recidivism (Babcock et al., 2004; Feder & Wilson, 2005). Babcock et al. (2004) examined the findings of 22 intervention efficacy and effectiveness studies. They coded treatment into three groups: Duluth model, cognitive behaviour therapy, and other. The other group was an aggregate of several lesser used therapies, including couples counselling, supportive therapy, relationship enhancement, and unspecified therapies. They found that the overall effect size from experiments, using either police reports or reports from partners to assess recidivism, was small for the Duluth model and other therapies. The effect size of cognitive behaviour therapy could not be calculated due to an insufficient number of studies. The overall effect size from quasi-experiments, using police report to assess recidivism, was moderate to small. Treatment programmes employing the Duluth model showed a moderate effect size, while both cognitive behavioural and other therapies approaches showed small effect sizes. When reports from partners' of the offenders was used to assess recidivism, studies showed the Duluth model had a moderate effect size, while cognitive behavioural therapy had a small effect size. There was no significant difference in the mean effect sizes between the Duluth and cognitive behaviour approaches using either assessment of recidivism. The authors noted that equivalent effect sizes between the interventions likely resulted from the current interventions mixing different theoretical approaches (such as the Hamilton Abuse Intervention Pilot Programme, discussed above). Therefore, this research does not reflect a "head-to-head comparison" between the Duluth model, cognitive behavioural therapy, and other modalities (Babcock et al., 2004, p. 1045).

In a later meta-analysis of the effectiveness of court-mandated stopping violence programmes Feder and Wilson (2005) compared four experimental and six quasi-experimental studies that employed either a psychoeducational or cognitive behavioural approach (or some mix of the two). The effect sizes were calculated by both outcome (official reports and victim reports) and by research design (experimental, quasi-experimental with a no treatment comparison group, and quasi-experimental with treatment dropouts as a comparison group). The authors noted that official reports referred to either re-arrest or conviction. The mean effect size, using re-arrest to assess recidivism, from the experimental studies was small, while the mean effect using victim report was zero. The studies that employed a quasi-experimental with a no treatment comparison design indicated an overall small harmful effect. That is, intimate partner

violence increased following intervention. Finally, the quasi-experimental with treatment dropouts as a comparison studies showed a large, positive mean effect.

A limitation of both meta-analyses is the likely method bias present in the individual studies. Recidivistic outcome measures, including police report, victim report, and perpetrator report, may not adequately reflect recidivism rates because they are dependent on the victim and perpetrators willingness to disclose violence. For example, 12.8 percent of 956 New Zealand women surveyed in a study by Fanslow and Robinson (2010) reported having disclosed their experience of intimate partner violence to the police. By underreporting recidivism the effectiveness of the intervention may have been overestimated and, in turn, the effect size may also have been overestimated. Regardless of the interventions studied, the generally small effect sizes found in both meta-analyses indicate that the stopping violence programme formats may not be targeting the needs of the populations. It has been proposed that programmes begin to differentiate the content to target the needs of the major types of intimate partner violence perpetrators or, further still, tailor treatment to the needs of the individual (Andrews & Bonta, 2010; Kelly & Johnson, 2008).

Chapter Summary

As shown in the current chapter, studying intimate partner violence is important for a number of reasons. First, the prevalence rates and associated physical and mental health consequences indicate that it is a serious problem that impacts on many New Zealanders. Second, we need to better understand why some people engage in intimate partner violence, in order to develop more effective interventions. Research has shown that the current one size fits all approach has limited effectiveness (Babcock et al., 2004; Feder & Wilson, 2005), and may be targeting the violence perpetrated by the family only typology (Clements et al., 2002). Interventions focused on individual differences may help reduce intimate partner violence. Third, high risk perpetrators need to be identified and managed accordingly. While the police currently use a measure to ascertain the risk of homicide and serious harm the measures have not been proven to be reliable and valid therefore, it's utility in guiding both management strategies and interventions is unknown. The following chapters will explore the assessment of risk for violence and considerations in the evaluations of these assessments, culminating in a review of the SARA.

CHAPTER TWO

THE ASSESSMENT OF RISK FOR VIOLENCE

There is little agreement on what constitutes risk (Kropp, 2004). In New Zealand mental health settings risk is defined as the likelihood of an unfavourable event or outcome (Wilson, 1998), and risk for violence implies that the unfavourable event or outcome involves interpersonal violence. The literature defines violence risk as the probability that some form of violence will take place at some time in the future (Hanson & Wallace-Capretta, 2000). However, Kropp (2004) argued that while sufficient for research purposes, this definition is too narrow in practice. Decisions about risk must take into consideration, in addition to likelihood, the imminence, frequency, nature, and seriousness of the projected actions. The inclusion of these considerations reflects the complexity inherent in risk assessment; it is more than just a question of whether future intimate partner violence will or will not occur (Kropp, 2004). This latter definition of risk for violence will be employed in this study.

The first generation of risk assessment, commonly referred to as unstructured clinical judgment, involves the assessor determining risk at their discretion, with no guidelines or constraints (Kropp, 2004). As a result, the decisions are justified according to the expertise of the assessor. The advantage of this approach is that it allows for individual differences to be taken into account. However, clinical judgment has also been widely criticised for its lack of reliability and validity (Douglas & Kropp, 2002), leading Grove and Meehl (1996) to conclude that they are “informal, subjective, [and] impressionistic” (p. 239). Much of the current knowledge base regarding violence risk prediction was accumulated in response to these concerns (Heilburn, 1997; Litwack & Schlesinger, 1999; Monahan & Steadman, 1994). The present chapter discusses this current knowledge base, specifically focusing on static and dynamic risk factors. It is within this context that the later generations of risk assessment, the actuarial and structured professional judgment approaches, are introduced. The chapter concludes with a discussion of risk management.

Risk Factors

A risk factor is "a measurable characterisation of each subject in a specified population that precedes the outcome of interest and which can be used to divide the

population into 2 groups (the high-risk and the low-risk groups that comprise the total population)" (Kraemer et al., 1997, p. 338). For the purpose of this study risk factors will be considered any variable that has been shown to be associated with violence, and which can be individual, contextual, historical, or clinical in nature (Haggard-Grann, 2007). The literature groups risk factors into two categories: static and dynamic.

Static risk factors

A static risk factor describes variables that are fixed, and cannot change over time, such as race, or year of birth (Kraemer et al., 1997). Much of the research to date has explored the relationship between static risk factors and their predictive ability in regards to violence (see Quinsey, Harris, Rice, & Cormier, 2006), with several studies having highlighted their relationship to future violence. For example, the relationship between childhood maltreatment and future offending has been established in offending populations generally (Monahan, 1981; Webster, Dickens, & Addario, 1985), and in intimate partner violence (Barnett & Fagan, 1993; Dutton, Starzomski, & Ryan, 1996) and juvenile offending populations specifically (Smith & Thornberry, 1995; Stewart, Dennison, & Waterson, 2002). In an analysis of the relationship between childhood family violence and intimate partner violence McKinney, Caetano, Ramisetty-Mikler and Nelson (2009) surveyed United States couples ($n = 1,615$) and found that men who reported moderate or severe physical abuse as a child were, respectively, 3.9 and 4.5 times more likely to engage in intimate partner violence than men who did not report childhood abuse. In addition they found that witnessing inter-parental aggression was associated with a slight increase in risk.

Due to the continued empirical support for the reliability and predictive ability of specific static risk factors, some have argued that risk assessments should be solely derived from these variables (Heilbrun, 1997; Monahan et al., 2001). Such methods of risk assessment, those referred to as actuarial, are almost exclusively based on static risk factors (Dvoskin & Heilbrun, 2001), and are designed only to predict future behaviour (as opposed to facilitate behaviour prevention, discussed below; Hart, 2001). The development of the actuarial model represents the second generation of risk assessment. Each risk factor is selected on the basis of its empirical association with violence (Hart, 2001), and each item is then weighted and combined with other risk factors, according to fixed algorithms to yield a decision, or probability of future risk. Much of the research conducted with static risk factors investigates its relationship to official

recidivism events, rather than true reoffending (Beech, Fisher, & Thornton, 2003). As a result, these measures likely underestimate risk.

There are numerous actuarial measures that have been developed to predict the likelihood of future behaviour, including violent behaviour. One of the first measures created for this purpose was the Violence Risk Appraisal Guide (Harris, Rice, & Quinsey, 1993; Quinsey et al., 2006). It contains twelve items that are designed to assess dangerousness in high-risk men, each of which are known predictors of violent behaviour. The Violence Risk Appraisal Guide was developed to assess recidivism in serious offender populations in general. Several studies have shown that it is an adequate predictor of recidivism, with an area under the receiver operating characteristic (ROC) curve ranging from .75 to .90 (Harris & Rice, 2003; Harris, Rice, & Cormier 2002; Kroner & Mills, 2001). However, the predicative capacity of the Violence Risk Appraisal Guide decreases when populations the measure was not developed for are sampled. For example, in their study of the predictive validity of the Violence Risk Appraisal Guide Harris et al. (2003) found the area under the ROC curve to be .65 with men who had been convicted of a sexual offence.

During the time when an “all or nothing” approach was taken towards prediction in mental health and correctional settings, actuarial measures were considered sufficient to perform the task required, that is, predict violence (Dvoskin & Heilbrun, 2001). The “all or nothing” approach refers the involuntary holding of a person until they were predicted to be safe to the community, at which time they were released from institutions. However, in recent years the field of risk assessment has undergone a conceptual shift, from that of violence *prediction* to violence *prevention* (Hart, 1998). That is, from the identification of people considered more at risk of violence than others, at a specified point in time, to a focus on risk management, and the endeavour to decrease a person’s risk. As a result of this, static risk factors and actuarial methods have been questioned. Static risk factors, which inherently cannot be altered, have undergone recent critical review, the results of which highlight their limited utility when considered under the current conceptualisation of risk assessment (see Douglas & Kropp, 2002; Douglas & Skeem, 2005; Haggard-Grann, 2007).

Douglas and Kropp (2002) argued that static risk factors limit any assessment to assigning a risk rating that is unchangeable; that is establishing a person’s *risk status*, rather than their *risk state* (Douglas & Skeem, 2005). Research has shown that risk is not a stable entity; it can change over time (Gendreau & Goggin, 1996). To adequately

assess risk more information should be provided than simply the long-term risk rating provided by actuarial measures. Acute risk factors, which may indicate imminent reoffending, must be identified (Douglas & Skeem, 2005). For example, alcohol intake could increase a person's risk for violence, therefore alcohol consumption would be monitored, and as alcohol intoxication increased, so would that person's level of risk for violence. When only static risk factors are utilised in risk assessment the adjustments necessary to guide treatment and management cannot be assessed and, as a result, risk may increase.

Additionally, static risk factors have been criticised for their limited value in management and treatment planning. Several authors have argued that the purpose of risk assessment is to guide management and treatment, and to reduce the future likelihood of violence (Douglas & Kropp, 2002; Douglas & Skeem, 2005; Haggard-Grann, 2007). Static risk factors cannot be targeted for treatment because they are historical in nature. For example, item one of the Violence Risk Appraisal Guide, which asks whether the person being assessed lived with both biological parents until the age of sixteen, cannot be changed. Therefore, actuarial assessments are capable only of providing guidelines for the level of external control required (for example, parole and release decisions), rather than decisions about the interventions that can be implemented to reduce risk (Philipse, Koeter, van der Staak, & van den Brink, 2005). The capacity actuarial assessments have to facilitate the prevention violence is limited (Haggard-Grann, 2007).

Despite these limitations static risk factors remain important in aspects of risk assessment, namely determining the intensity of an intervention (Douglas & Kropp, 2002; Haggard-Grann, 2007) and estimating long-term risk (Hart, Webster, & Douglas, 2001). However, static risk factors are no longer adequate for risk assessment, when used without dynamic, or changeable, risk factors (discussed further in the next session). The focus has shifted from establishing the risk status, to establishing the *risk state*, or the "propensity to become involved in violence at a given time, based on particular changes in biological, psychological, and social variables in his or her life" (Douglas & Skeem, 2005, p. 349). The latter is dependent on dynamic risk factors. Therefore, the risk status determines who is high risk, and should receive the often limited resources, while the risk state provides information on both who is at high risk and the interventions that the person would benefit from, in order to reduce their risk status.

Dynamic risk factors

According to Kraemer et al. (1997) dynamic risk factors are variables associated with an outcome that can change, either spontaneously or through intervention. Until recently dynamic risk factors were inadequately researched—they were considered to provide little supplementation to the predictability of actuarial measures and to be unreliable (Gendreau & Goggin, 1996). However, as a result of the conceptual shift in risk assessment thinking research investigating dynamic risk factors has begun to be conducted and measures that incorporate these factors are being developed. Dynamic risk factors have been hypothetically categorised into stable and acute risk factors, those that can change gradually and rapidly, respectively. Douglas and Skeem (2005) noted that while no measures currently exist that were developed specifically to assess risk state, it appears that the “third generation” of risk assessment method, structured professional judgment, may inadvertently be doing so.

Structured professional judgment measures are guidelines that contain the minimum number of empirically derived static and dynamic risk factors to be considered in any risk assessment. However, instead of weighting the factors according to algorithms, as in actuarial assessments, the rater uses clinical judgment to assign a risk rating that is explicitly linked with the ideas around management and treatment requirements for the person being assessed. Therefore, structured professional judgement measures reflect theoretical, empirical, and clinical knowledge about the construct being assessed, a “bridg[ing] the gap” between the two previous generations (unstructured clinical judgment and actuarial assessments; Kropp, 2004, p. 683).

Structured professional judgement measures have been developed for use with general offending (for example, Historical Clinical Risk Management-20, Webster, Douglas, Eaves, & Hart, 1997; Short Term Assessment of Risk and Treatability, Webster, Martin, Brink, Nicholls, & Middleton, 2004), sexual offending (for example, Sexual Violence Risk-20, Boer, Hart, Kropp, & Webster, 1997; Risk for Sexual Violence Protocol, Hart et al., 2003) and intimate partner violence (SARA, Kropp et al., 1999) populations. All of these measures are designed around the premise of facilitating violence prevention, not violence prediction (Douglas & Kropp, 2002). That is, they endeavour to prevent violence by both generating risk predictions and guiding the development of a management plan.

In light of the potential utility of dynamic risk factors, researches have begun to investigate their predictive ability. For instance, Dempster and Hart (2002) found that

the inclusion of dynamic psychosocial risk factors, such as substance abuse, and relationship and employment difficulties, improved the prediction of sexual violence recidivism. Furthermore, in a meta-analysis of risk factors of general violence recidivism undertaken by Gendreau, Little, and Goggin (1996), the correlation coefficients between dynamic risk factors and recidivism were equal to or greater than the correlation coefficients between static risk factors and recidivism. Although these studies are useful in highlighting the potential of dynamic risk factors in predicting recidivism, they do not demonstrate the utility of the dynamic risk factors—that is, whether they are changeable and whether any change is related to a change in risk for violence. To determine if dynamic risk factors facilitate the prevention of violence they need to be studied using longitudinal research designs, rather than the discussed cross-sectional research designs.

Given this limitation, studies employing longitudinal designs, with dual and multiple time points, have been conducted. Brown (2002) conducted a three wave prospective study with adult male offenders in Canada examining the predictive ability of dynamic risk factors. The risk factors were assessed at three time points: pre-release, one month post release, and three months post release. For the participants who were not recidivists seven dynamic risk factors (employment, single/unsupportive partner, negative affect, perceived problem level, substance abuse, social support, and expected positive consequences of crime) changed over the follow-up period, with the changes leading to a reduction in risk. In turn, these changes significantly predicted an absence of further violence. In a later longitudinal study, Mulvey et al. (2006) assessed dynamic risk factors every week for six months in a sample of 135 psychiatric patients. They found that drinking, drug use, and violence occurred in acute periods of psychiatric symptom activity, and that alcohol “yesterday” predicted violence “today”.

In addition to the research investigating the predictive utility of individual dynamic risk factors, studies have been undertaken investigating the predictive accuracy of actuarial and structured professional judgment approaches for intimate partner violence populations. A recent meta-analysis of the accuracy of various assessment approaches (intimate partner violence actuarial scales, other actuarial scales, intimate partner violence structured professional judgment scales, and victim judgment) to predict intimate partner violence recidivism was undertaken by Hanson, Helmus, and Bourgon (2007). While the mean effect sizes of the approaches ranged from small to large, no significant difference was found between the approaches, indicating that all of

the approaches demonstrated similar predictive accuracy.

Risk Management

The identification of dynamic risk factors allows for the management of risk, through decisions about which tactics and strategies to implement. Risk management has been defined as "the task of constructing social and physical environments that, in combination with knowledge of the individual's assets and liabilities, will likely lead to substantial reduction in violence potential" (Webster, Douglas, Belfrage, & Link, 2000, p. 128). Hart (2008) argued that risk management strategies should be developed according to three principles, each of which is derived from the risk assessment procedure (Douglas & Kropp, 2002). First, the level of management should reflect the risk level. For example, specialist one to one counselling interventions, which are more costly and time intensive than group based programmes, should be offered to high risk perpetrators because a reduction in the likelihood, severity, and frequency of the violence is likely to be more meaningful than that of a low risk offender. Low risk offenders are less likely to recidivate. Second, risk management should reflect the relevant risk factors derived from the risk assessment procedure. For example, if alcohol intoxication is identified as a risk factor then it should be addressed in the management strategy. Third, risk management should be individualised to the person, and be tailored and adjusted to their specific needs. That is, the combination of strategies employed should reflect individual differences.

Studies have recently begun exploring the efficacy of interventions driven by third generation risk assessment. In a meta-analysis, Andrews and Bonta (2006) found that the average mean effect size for interventions that targeted identified risk factors for violence, specifically antisocial cognitions, antisocial associates, or work/school problems was small. In comparison, the mean effect size for interventions that did not target these risk factors was almost zero. The interventions targeting identified risk factors clearly had a greater impact on outcome. Within the intimate partner violence field, specifically, studies have examined the relationship between risk assessment, the number of risk management strategies implemented, and intimate partner violence recidivism. For example, Belfrage et al. (2011) investigated the use of the SARA by police officers in Sweden to facilitate the prevention of recidivism. They found that the number of risk management strategies mediated the relationship between risk assessment and intimate partner violence recidivism. A limitation of this study was

that the authors did not investigate the effects of the specific strategies. Nevertheless, these studies indicate that targeting identified dynamic risk factors may facilitate the prevention of further violence.

Chapter Summary

The current chapter provided an overview of risk assessment for violence. In general, assessments of an offender's risk of future violence play a central role in decision making pertaining to that person's sentencing, community release, case management, and public safety concerns (Andrews & Bonta, 2010; Hoge & Andrews, 1996). While static risk factors have shown predictive accuracy and utility in determining the intensity of any intervention, they are no longer adequate for risk assessment when used alone. Static risk factors do not assess risk state. In light of this, Douglas and Skeem (2005) propose that structured professional judgment approaches, which incorporate both static and dynamic risk factors, may provide more utility than actuarial approaches because they assess risk state. The assessment of risk state should guide risk management, through intervention and management goals and strategies that will lead to risk reduction (Heilbrun, 1997). The specific actuarial and structured professional judgment measures used with intimate partner violence populations are discussed in chapter four, after an overview of the psychometric considerations in violence risk assessment.

CHAPTER THREE

CONSIDERATIONS IN VIOLENCE RISK ASSESSMENTS

While the overarching aim of violence risk assessment, and subsequently risk management, is to avoid or minimise harm towards others, risk assessment involves balancing the safety of the community with the rights of the person being assessed (Douglas & Kropp, 2002). Further confounding this melee is the reality is that no assessment procedure is completely error free. In violence risk assessment, the results of such errors will either be that potential victims may experience undue harm or that misclassified perpetrators may experience an unjust loss of liberty. One way of addressing this is to evaluate the psychometric properties of the risk assessment being employed. That is, the measure needs to be accurate. Evidence for this comes from evaluating reliability and validity.

The selection of which type of reliability and validity to assess is critical in evaluation procedures; there needs to be a congruence between the type of reliability and validity established and the intended use of the measure (Douglas & Kropp, 2002; Groth-Marnet, 2009). For structured professional judgement measures this should include establishing reliability, the validity of the proposed construct, and the ability of the measure to predict and facilitate the prevention of violence. The present chapter reviews the reliability and validity literature, with specific focus on internal consistency, interrater reliability, and convergent, discriminant, predictive, and incremental validities. This chapter is specific to the evaluation of risk assessment and, accordingly, the reliabilities and validities discussed will be in relation to this context. International reliability and validity studies pertaining to the SARA, specifically, are discussed in the following chapter. The chapter concludes with a discussion of the specific characteristics of New Zealand offender populations that affect the generalisability of the current risk assessment evidence base and the shrinkage of predictive validity when violence risk assessments are implemented with indigenous offender samples.

Reliability

The reliability of a test refers to the degree of “stability, consistency, predictably, and accuracy” within the test scores (Groth-Marnet, 2009, p. 11). According to classical test theory, the general premise underlying reliability is that all test scores are comprised of true scores and measurement error. The latter are

fluctuations that occur either randomly or systematically. While some fluctuation is expected, large fluctuations can reduce the certainty that the score is accurate and, in turn, negate validity. The reliability of test scores can be established through various methods, including internal consistency and interrater reliability, as well as temporal stability and alternative forms (Kaplan & Saccuzzo, 2005). The following sections discuss internal consistency and interrater reliability.

Internal consistency

Classical test theory assumes that the items in a test are randomly selected from a universe of potential items, which together assess the whole construct (Streiner, 2003). Therefore, the test items should be related to one another, and evaluating the internal consistency provides information on the relationship between the items. However, several authors argued that internal consistency is only relevant when the measure assesses an underlying singular construct, which most risk assessments do not (Douglas & Kropp, 2002; Douglas, Skeem, & Nicholson, 2011; Mills, Kroner, & Morgan, 2011). Recidivism is an observable outcome, not a construct. In addition, the individual risk factors are independent variables, rather than indicators of a singular latent construct. Violence risk recidivism (the construct) does not cause the test items (risk factors such as witnessed or was the victim of violence as a child) in the same manner as, for example, anxiety causes an increased heart rate. Therefore, structural reliability analyses do not match with the characterisation of current structured professional judgement risk assessments.

Most commonly evidence comes from the Cronbach's alpha statistic (Streiner, 2003), which estimates the extent to which responses on the items correlate with the other items. A Cronbach's alpha above .70 is desirable as it reflects adequate internal consistency of the test structure and, subsequently, that the items are measuring the same construct. For risk assessment the risk factors do not need to correlate with the other items; the relationship between recidivism and risk factors can be measured. Ideally the risk factors should be independent of each other and each have a strong relationship to recidivism. However, doing so minimises Cronbach's alpha. Internal consistency is relevant when a test purports to measure an underlying construct, including past violent behaviours. These are multi-item measures of individual risk factors. Therefore, while internal consistency is important it is not required in evaluating the reliability of third generation risk assessment measures.

Interrater reliability

Homogeneity amongst raters can be divided into two categories. Interrater reliability is concerned with the *relative consistency* in test scores from multiple judges—that is, the consistency of each judge in scoring, according to his or her individualised definition of the test (LeBreton & Senter, 2008; Stemler, 2004). On the other hand, interrater agreement is concerned with the *consensus* in test scores from multiple judges, or the degree to which they share a common understanding of the construct. Several authors argued that interrater reliability is the most important measure of reliability for violence risk assessment (Baird, 2009; Douglas & Kropp, 2002; Mills et al., 2011). Regardless of the model employed, all violence risk assessments depend on some degree of subjective assessment and clinical inference. If there is no consensus between raters on the level of risk or risk factors present then the test is of limited use; the risk ratings will not be reliable or consistent.

While multiple methods are used to examine interrater reliability and interrater agreement individually, intraclass correlation coefficients establish both simultaneously. Intraclass correlation coefficients assess reliability by comparing the variability of the ratings of the same subject to the total variation across all ratings and all subjects. Correlation coefficients between .70 and .85 are desired (LeBreton & Senter, 2008). Disagreement on a given case can be caused by various factors, including rater association, rater bias, and rater distribution (Uebersax, 2010). Rater association concerns whether the raters have a shared understanding of the trait being assessed (Fleiss, 1973). In terms of structured professional judgements, raters may vary in the weight they give to the individual risk factors, or use different algorithms to combine the information presented. The result of rater association is correlations of less than 1.0. Intraclass correlation coefficients take into account this association (Bartko & Carpenter, 1976). Rater bias refers to the tendency of raters to systematically make ratings higher or lower than the other raters. Comparing the rating means given by each of the raters, across the cases, can assess bias. Intraclass correlation coefficients are sensitive to bias in the rater means (Uebersax, 2010). Finally, rater distribution refers to cases where one rater has a noticeably different distribution in ratings, compared to all of the raters combined (Uebersax, 2010). Such rater distributions are typically identified through visual analysis of graphed distributions of ratings (Uebersax, 2010).

Validity

While validity was originally defined as showing that a test “measures what [it] purports to measure” (Kelley, 1927, p. 14; Cattell, 1946), this concept has undergone many evolutions throughout the twentieth century. These have included whether the empirical relationships between test scores match some defined theoretical relations (Cronbach & Meehl, 1955) and whether the individual components of validity evidence are justified scientifically, as well as socially and ethically (Messick, 1980; 1993; 1998). However, Borsboom, Mellenbergh, and van Heerden (2004) criticise these revisions for complicating a basic, unitary concept and failing to “offer a simple, clear, and workable alternative” (p. 1061). They stated that a test measures a construct if it can be shown that the construct exists and that changes in the construct results in changes in the test outcome.

That validity can be considered a unitary concept does not imply there is one source of evidence (Sartori & Pasini, 2007). Validity can be divided into two parts: test construction and clinical practice (Groth-Marnet, 2009). Initially, the validity of test construction relied upon criterion-related validity and was demonstrated by correlating a test score with some criterion, which represented a desired construct (Cronbach & Meehl, 1955). However, finding objective criteria that were universally accepted as an indicator of the construct proved difficult (Groth-Marnet, 2009). As a result additional forms were developed, specifically construct validity (Cronbach & Meehl, 1955). According to the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1999) test standards psychometric evaluations should establish the criterion-related (concurrent and predictive), content, and construct (convergent and discriminant) validities of any test that will be used.

The evidence that a test has demonstrated validity usually comes from previously conducted validation studies. If a test is shown to predict intimate partner violence then one can say that it can be ethically used. However, research findings that are valid in one setting may not be valid in another. They may not generalise to settings outside those in which they were originally tested. In the case of New Zealand, several characteristics (for example, ethnicity and culture) likely impact on the population as a whole and create a uniqueness that differentiates it from the populations of other countries. Thus the psychometric properties of violence risk assessments undertaken with Canadian samples cannot be assumed to hold in New Zealand samples. Validity

studies with New Zealand populations are (ideally¹) required in order to interpret the test scores (such as risk ratings) with confidence. The following sections discuss the sources of validity that will be used in this study—specifically convergent and discriminant validity, predictive validity, and incremental validity.

Convergent and discriminant validity

Convergent validity refers to the degree to which a test measures the proposed construct (Cone & Foster, 2006), while discriminant validity is the degree to which a test does not measure constructs that are proposed to be independent (Cone & Foster, 2006). Evidence is commonly established by correlating test scores on the measure being validated with scores on other measures of similar or different constructs. For example, a person's score on a violence risk assessment should be closely correlated to their score on alternate violence risk assessment measures, if both tests are measuring the same construct. On the other hand, correlations with scores on measures of differing constructs should be small, indicating no relationship between the constructs of violence risk and the alternative.

There are a number of issues with establishing convergent and discriminant validity using correlational methods. First, variance that results from using the same type of measurement (for example, ratings by others and self-report) can artificially inflate relationships, which would otherwise be attributed to a shared construct (Campbell & Fiske, 1959). Second, the strength of the relationship is dependent on the psychometric strength of the comparison tests (American Educational Research Association et al., 1999). If large measurement error exists in the comparison test then it becomes unclear as to whether the relationship between the variables is due to the construct of interest or due, in fact, to the measurement error. Several methods have been proposed to overcome these limitations (see Foster & Cone, 1995; Straus & Smith, 2009), including using multiple measurement classes and well validated comparison tests.

¹ There are, of course, many psychometric measures in widespread use in New Zealand for which local standardisation have not been conducted. However, particularly in the area of risk assessment, such an approach would appear to be unjustifiable.

Predictive validity

Predictive validity refers to the extent to which test scores predict the scores on an outcome criterion. This validity is considered appropriate for tests that aim to classify people (Groth-Marnet, 2009), such as risk assessments. Unlike concurrent validity, in predictive validity the criterion score is collected after the test score, at some point in the future. This is achieved through longitudinal research designs, with criterion data collected prospectively. There are several methods for evaluating predictive validity, including the percent correctly classified, correlation coefficients, and odds ratio (Cohen, 1969; Rice & Harris, 1995). An alternative method is the area under the ROC curve analysis. ROC curves assess the trade off between sensitivity and specificity (Rice & Harris, 1995). Sensitivity is the true positive rate, or the degree to which positive cases are identified as positive by the test. Specificity is the true negative rate, or the number of negative cases that are confirmed as negative by the test. Ideally both sensitivity and specificity should be 100 percent. However, prioritising one inevitably negatively impacts on the other. While Mossman (1994) argued that ROC curves are the most useful measure of predictive validity because such analyses are not dependent on either the base rate of the recidivistic behaviour or the proportion of cases predicted to be violent, others (e.g. Mills et al., 2011) argued that ROC curves ignore the impact of base rates. Even with relatively high area under the ROC curve values, predictions about behaviours with low base rates will increase the number of false positives; that is the number of positive cases that are incorrectly identified.

Of central importance in evaluating the predictive validity of a test is the outcome criterion used. Douglas, Otto, and Borum (2003) stated that it should be representative of the attribute being tested, be a reliable measure of that attribute, and not be influenced by external factors. Evaluations of risk assessment measures often utilise official records, rather than self-reports, to measure the criterion recidivism (Quinsey et al., 2006). Despite criticism that official records often underestimate the prevalence of the criterion, they are considered less influenced by the limitations of self-report. These include high levels of participants leaving the study during the follow-up period and responding to the outcome questions in a socially desirable manner (such as stating no violence has occurred, when in fact it has). In addition, official record is less influenced by systematic positive treatment effect bias where the lower recidivism rate is falsely attributed to the effect under investigation. Official records include measures

of convictions, charges and police contact. Police contact, specifically, includes all incidents that the police attended, regardless of whether charges resulted.

Wiggins (1993) and others (for example, Anastasi, 1997) have argued that conclusions drawn from predictive validity are specific to the context, including the test being assessed, the goal of the assessment, and the assessment process. Furthermore, the generalisability of the findings is dependent on the base rate of the behaviour being predicted. If the base rates in two or more situations are substantially different then the predictive validity of a test in one sample cannot be assumed in the other (Hunsley & Meyer, 2003). For example, given the base rate of self-reported violence is higher than the base rate of police contact for violence (Babcock et al., 2004), the validity of any risk assessment to predict self-reported violence cannot be generalised to the same measure using police contact for violence as the outcome criterion.

According to test standards published by the American Educational Research Association et al. (1999) psychometric evaluations should match the test's intended purpose. Therefore, structured professional judgement risk assessments should be shown to both predict recidivism and facilitate the prevention of recidivism (Douglas & Kropp, 2002). That is, the relationship between risk decisions made in professional judgements for both recidivism and the facilitation of violence prevention should be demonstrated. Most traditional studies of predictive validity assess the ability of the measure to predict recidivism, as opposed to facilitate the prevention of recidivism. Studies need to evaluate the relative effects of risk management tactics, which are guided by the risk assessment, in preventing further violent recidivism. While establishing the facilitation of violence prevention is outside the scope of the current study, it is still an important principle of assessing the validity of instruments.

Incremental validity

While construct and predictive validities establish the validity of a test in terms of its construction, they provide limited direction in clinical practice (Groth-Marnet, 2009). This is especially true when there are multiple tests that purport to evaluate the same construct. Incremental validity provides information on a test's ability to enhance the "prediction of a criterion beyond what can be predicted with other data" (Hunsley, 2003, p. 443) and, therefore, enhance clinical utility by guiding assessment processes. That is, it establishes the value of adding either new tests, or parts of tests, to the existing assessment protocols, in light of the economic and psychological cost that

comes with larger test batteries (Hunsley, 2003). A common method for evaluating incremental validity is to examine the unique and shared variance of the test's prediction in relation to the criterion through regression analyses (Haynes & Lench, 2003; Hunsley & Meyer, 2003). Many of the limitations in incremental validity are similar to that of predictive validity, including the requirements of a "gold standard" outcome criterion and the context specificity.

Population validity

Validation studies conducted with one population may not generalise to another population due to characteristics unique to specific populations. Specifically, differences exist both between Māori and non-Māori populations in New Zealand and between indigenous samples and the Northern American validation samples used for most measures. Due to current research focusing solely on indigenous populations this section focuses predominantly on Māori peoples. However, it is worth noting that New Zealand has high levels of Pasifika and Asian peoples as well. The cultural differences between these populations and those the SARA has been previously validated on may also impact on item endorsement and interpretation.

While most counties report an over representation of indigenous people in offender and intimate partner violence populations, the rate in New Zealand is higher than the majority. As previously discussed Māori are overrepresented as victims of intimate partner violence (see chapter 1, p. 6). This holds true for perpetrators as well. Although Māori make up 15 percent of the general population, 50 percent of both the general offender population and men sentenced for a "male assaults female" offence were Māori (New Zealand Department of Corrections, 2007; Doone, 2000). In comparison, approximately 19 percent of offenders in Canada and 24 percent of offenders in Australia identified as aboriginal (Bartels, 2010; Correctional Service Canada, 2006). Balzer and colleagues (Balzer, Haimona, Henare, & Matchitt, 1997) argued that the high prevalence of intimate partner violence within Māori should be understood within the context of colonisation, which has resulted in a loss of traditional beliefs, values and identity, as well as the breakdown of whānau (extended family) and hapū (sub-tribe) structures within which family violence was traditionally resolved.

At the risk factor level, Maynard and colleagues (Maynard, Coebergh, Anstiss, Bakker, & Huriwai, 1999) state that culturally specific risk factors exist and the rate of risk factor endorsement differs between indigenous and non-indigenous samples. A

study by the New Zealand Department of Corrections found that for Māori offenders there exist culture related risk factors, which are above and beyond those endorsed by the population as a whole (Maynard et al., 1999). These were a lack of cultural identity, lack of a sense of group membership, negative self-image, relationship with whānau, and presence or absence of whakawhānaunga (whānau-like relationships). It has been argued that the first three of these risk factors may be applicable for indigenous populations as a whole, while the later two are unique to Māori (Rugge, 2006).

Internationally, studies have shown that some generally accepted risk factors do not differentiate indigenous and non-indigenous offenders. Employment status, changes in primary address, and motivation to change /willingness to accept responsibility for the offence have been shown to not differentiate indigenous offenders that reoffend from those that do not reoffend (Broadhurst & Maller, 1990; Wright, Clear, & Dickson, 1984), while family/marital problems and school/work problems were not found to predict recidivism in indigenous offenders (Bonta, 1989; Bonta, LaPrairie, & Wallace-Capretta, 1997). That indigenous populations endorse some risk factors, regardless of their recidivism status, likely holds true for Māori as well. It is likely that these risk factors do not take into account societal and cultural differences. For example, according to Statistics New Zealand Māori had a higher rate of unemployment in 2012 than New Zealand Europeans (13% versus 5%; Bascand, 2012). Therefore, Māori are likely to score higher on this item regardless of recidivism status. For New Zealand, overall, half of the men charged in an intimate partner violence offence are likely to score higher than North American samples on existing violence risk assessment measures while some risk factors specific to Māori are not being assessed. This is not to say that being Māori is a causative factor in offending. It is likely that cultural isolation exacerbates underlying risk factors, which, in turn, contributes to offending behaviour (Doone, 2000).

In terms of risk assessment measures, studies have shown that the predictive validity of sexual violence risk assessment measures shrink when used outside North American populations they were validated on (see Allan, Dawson, & Allan, 2006; Helmus, Babchishin, & Blais, 2012; Langstrom, 2004; Ward & Dockerill, 1999). Langstrom (2004) found that in samples of Nordic, non-Nordic European, and African Asian Swedes the Rapid Risk Assessment for Sexual Offence Recidivism and Static-99 accurately predicted recidivism for the Nordic and non-Nordic European offenders, but both measures did not differentiate African Asian recidivists from non-recidivists. A

study of 538 sexual offenders in Western Australia found the predictive accuracy of the Violent Offender Treatment Programme Risk Assessment Scale, a measure of violent recidivism, was lower for the Aboriginal sample, compared to the non-Aboriginal sample for both sexual recidivism and non sexual recidivism (Ward & Dockerill, 1999). A recent study in Canada evaluating the predictive validity of the STABLE-2007 found that the measure did not significantly predict recidivism for the indigenous offenders, while it did for the non-indigenous offenders (Helmus, Babchishin, Blais, 2012).

It has been argued that differences in the predictive validity of these measures results from differences in the ethnic make up of North American and other populations. That is, factors unique to indigenous populations, such as culturally specific risk factors discussed above, which international risk assessment measures does not include, impact on the ability of risk assessment measures to accurately predict recidivism (Maynard et al., 1999). Given that approximately half of men sentenced for an intimate partner violence offence in New Zealand are Māori, the validity of risk assessment measures with this population needs to be investigated to evaluate if shrinkage impacts on the validity.

Chapter Summary

The accuracy of violence risk assessment has implications for both the safety of the community and the rights of the person. Evaluation of the reliability and validity of the assessment under investigation is therefore important. By doing so, the extent to which the scores are stable over multiple trials and the measure assesses what it intends to is established. For structured professional judgments, which were shown in the previous chapter to be the most useful type of risk assessment, evaluations should include establishing the interrater reliability, and convergent, discriminant, predictive, and incremental validities. Evaluations of internal consistency are not required because risk assessments do not measure a single underlying construct. While most of this evidence comes from research that has been previously conducted, the findings may not generalise to populations that differ from the original sample both because the exclusion of culturally specific risk factors likely leads to shrinkage in the predictive validity. As a result, validation studies need to be undertaken when measures are used within New Zealand, in order to interpret the scores with confidence. The following chapter outlines the SARA, specifically describing the risk factors and the existing reliability and validity literature.

CHAPTER FOUR

THE SPOUSAL ASSAULT RISK ASSESSMENT GUIDE

Monahan and Steadman (1994) asserted that unique risk predictor factors exist for the sub-types of violence, including intimate partner violence. There is a need for specific intimate partner violence risk assessment measures, which are based on these factors, as opposed to measures that predict the likelihood of general violence. While there are a number of actuarial risk assessments for intimate partner violence (including those used by the New Zealand police), the SARA was the only structured professional judgement measure that can be used by people from multiple professional backgrounds available when the current study began². Given the literature review on risk factors in chapter two, it was decided that a structured professional judgement measure would provide the most utility. Structured professional judgement measures guide risk management, which allows for intimate partner violence to be predicted and prevention to be facilitated. In addition, the authors noted that all professionals who have experience in both risk assessment and intimate partner violence can administer the SARA (Kropp et al., 1999). Therefore, the SARA was selected for evaluation.

The current chapter outlines the SARA, with specific focus on the rationale of each of the risk factors. The brief description of the scoring information is supplemented by the discussion of the SARA in the following chapter. This is followed by a review of the international reliability and validity studies. While a small number of studies have been conducted, it becomes apparent that voids exist in the current reliability and validity literature. The chapter then concludes by outlining the present study and the hypotheses.

The Spousal Assault Risk Assessment Guide

The SARA was developed in Canada, in the context of a growing acknowledgement of the economic and social toll of intimate partner violence (Kropp & Gibas, 2010). It is a structured professional judgement measure designed to predict intimate partner violence recidivism and guide intervention strategies. The measure was developed to provide a comprehensive list of empirically supported and clinically useful risk factors, which allowed for both the heterogeneity of perpetrators and the specificity

² The alternative structured professional judgement measure, the B-SAFER, is a brief version of the SARA designed for use only by police officers (Kropp & Hart, 2004).

of intimate partner violence, compared to general violence. In the development of the SARA the authors undertook a review of the clinical and empirical literature on risk for intimate partner violence. From this 20 static and dynamic risk factors were identified, each of which had demonstrated it either differentiated people who engaged in intimate partner violence from those who did not, or was associated with risk for intimate partner violence recidivism. The authors noted that a moderate level of specificity was aimed for and, as a result, risk factors that assess specific or detached behavioural acts were excluded (Kropp et al., 1995). Instead, risk factors at the trait, characteristic, or incident level were included. The risk factors reflect the minimum variables that should be considered when evaluating risk for intimate partner violence. To date no formal revision has been undertaken with the 20 risk factors, despite the authors noting that this would be completed as the research on risk for intimate partner violence accumulated (Kropp et al., 1999). The scoring information and individual risk factors will be discussed in the following section.

Description of the Spousal Assault Risk Assessment guide risk factors

The SARA is comprised of 20 static and dynamic risk factors that are rated on a three point scale of severity, from zero to two. Each risk factor is scored according to descriptive criteria unique to each, although each requires a degree of subjective professional judgement. The individual risk factor scores are added together to give a total score, with a possible range of 0 to 40. To arrive at an overall judgement of risk, the assessor considers the 20 risk factors (and any other relevant factors) in determining whether the person being assessed is at low, moderate, or high risk of recidivism (Kropp et al., 1999). This is referred to as the summary risk rating. Because the SARA is a structured professional judgement measure there are no fixed algorithms or guidelines for combining the risk factors into the summary risk rating. In addition, the levels of risk are not defined. Instead, the assessor uses their clinical discretion.

The risk factors are also grouped into two parts, referred to in the literature simply as part one and part two. However, for the purpose of this research, to avoid confusion with the thesis structure, part one was renamed the *general violence* subscale, and part two renamed the *intimate partner violence* subscale. These two subscales are discussed in the following sections.

General violence subscale. The general violence subscale, which consists of 10 risk factors, assesses risk for general violence. The possible score range is from 0 to 20.

The first three risk factors in the subscale are related to criminal history. These risk factors are past assault of family members (excluding past or present intimate partners), past assault of strangers or acquaintances, and past violation of bail, court orders, probation and parole. The remaining seven risk factors are related to psychosocial adjustment. These risk factors are recent relationship problems, recent employment problems, victim and/or witness to family violence as a child or adolescent, recent substance abuse/dependence, recent suicidal or homicidal ideation/intent, recent psychotic and/or manic symptoms, and personality disorder with anger, impulsivity, or behavioural instability. The literature outlining the relationship between each risk factor and intimate partner violence is discussed below.

Studies have shown that having a criminal history is associated with an increased risk for recidivistic violence, in general, and intimate partner violence, specifically (Gondolf, 1998; Gondolf & White, 2001; Hilton & Harris, 2005; Hilton, Harris, Rice, Lang, & Cormier, 2004; Jones & Gondolf, 2001). Research has shown that people with a history of assault, unrelated to intimate partner violence, are at an increased risk for violent recidivism (Monahan, 1981; Webster et al., 1985). Despite this, Campbell et al. (2003) found prior *arrest* for any violent act to be a protective factor against intimate partner homicide, specifically. The authors argued that their finding was mediated by coordinated community responses that were implemented following intimate partner violence arrest. Therefore, while prior arrest may be a protective factor where community responses exist, prior assault where no response is implemented may increase the risk of intimate partner violence and intimate partner homicide. In addition, research has found that people who engage in both intimate partner and non-intimate partner violence engage in more frequent and severe physical (Hamberger, Lohr, Bonge, & Tolin, 1996; Holtzworth-Munroe & Stuart, 1994) and psychological intimate partner violence (Boyle, O'Leary, Rosenbaum, & Hassett-Walker, 2008) than those whose violence is solely directed at their intimate partner.

Violating bail, court orders, probation and parole is assessed through three separate SARA risk factors, in relation to general criminal behaviour, past intimate partner violence, and during the current event. While the latter two risk factors are found in the intimate partner violence subscale, the literature pertaining to all three will be discussed here. This includes research that has shown that violation of any prohibiting order is a predictor of recidivism. Compared to people who do not violate probation conditions, people who do violate probation conditions have higher rates of

any general recidivism (Hart, Kropp, & Hare, 1988; Hilton et al., 2004; Quinsey et al., 2006), and violent recidivism, specifically (Harris et al., 1993).

The remaining seven risk factors pertain to psychosocial adjustment. Research has shown that social maladjustment, as assessed through relationship and employment difficulties, is associated with an increase in the severity and frequency of intimate partner violence, and risk for intimate partner homicide (Campbell et al., 2003; Cattaneo & Goodman, 2005; Hilton & Harris, 2005; Williams & Houghton, 2004). The SARA defines relationship difficulties as separation of the parties or extreme conflict regarding the relationship. The perceived threat of abandonment, resulting from relationship termination, has been associated with an increased risk for both intimate partner violence (Holtzworth-Munroe & Anglin, 1991) and intimate partner homicide (Wilson, Daly, & Wright, 1993). For example, separated women were three times more likely to be victimised than divorced women, and 25 times more likely to be victimised than married women (Tjaden & Thoennes, 1998). In addition, Campbell et al. (2003) found that employment difficulties were significantly correlated with future intimate partner violence.

While independent studies support the notion that trauma reported during childhood and adolescence, especially being the victim of and/or witnessing family violence, is related to recidivism in intimate partner violence (Aldarondo & Sugarman, 1996; Dutton & Hart, 1992a, 1992b; Dutton et al., 1996), meta-analyses of these studies do not support a direct relationship (Bennett Cantaneo & Goodman, 2005; Riggs, Caulfield, & Street, 2000). The authors suggest that mediating factors may be regulating intergenerational transmission (Riggs et al., 2000). Furthermore, research has found that trauma in childhood may be causally related to the onset of intimate partner violence (Aldarondo & Sugarman, 1996; Widom, 1989). The experience of family violence in childhood discriminates men who engage in intimate partner violence, from those who do not, but does not discriminate men currently engaging in intimate partner violence from those who had ceased the behaviour.

The last four risk factors related to psychosocial adjustment reflect research highlighting a link between specific mental health disorders, symptoms, and behaviours, and intimate partner violence. This relationship is considered either directly causal or mediated by factors associated with mental health disorders, such as poor coping skills and increased interpersonal stress (Kropp et al., 1995). Several studies have shown that alcohol abuse or dependence and, to a lesser extent, drug use, are related to both general

violence (Harris et al., 1993; Monahan, 1981) and intimate partner violence (Field, Caetano, & Nelson, 2004; Hanson & Wallace-Capretta, 2000). For example, Murphy and O'Farrell (1994) found that 66 percent of a sample of married men seeking treatment for alcohol abuse had engaged in intimate partner violence. In comparison, studies have estimated that 26 percent of the general population report intimate partner violence (Morris et al., 2003). The specific theories linking alcohol abuse and intimate partner violence suggest that the substance either disinhibits the abuser, or increases conflict. However, these relationships have not been well researched (Riggs et al., 2000).

It has been demonstrated in the literature that engaging in prior potentially fatal behaviour is related to homicide. For example, Glass et al. (2008) found that prior non-fatal strangulation was associated with a six fold increase in subsequent attempted homicide, and a seven fold increase in completed homicide. In relation to potentially fatal self harming behaviour, however, the association is less established. While people who murder an intimate partner often report experiencing suicidal ideation prior to the offence (Kropp et al., 1999), several authors have found that suicidal behaviour is either not or inversely related with intimate partner violence (Campbell et al., 2003; Hilton et al., 2004). Therefore, the literature indicates that prior potentially fatal behaviour is related to intimate partner homicide, but not intimate partner violence.

There is little consensus in the literature about the relationship between mental illness and intimate partner violence; this is especially true for psychosis and schizophrenia. Several studies found a relationship between schizophrenia and violent behaviour (see, Monahan, 1992), from which the authors concluded that schizophrenia was a risk factor for violence. More recently, however, studies have refined these findings and allowed for more specific conclusions. In a population survey of 38,132 adults in the United States, Mojtabai (2006) found that the greater the number of psychotic-like experiences the more likely intimate partner violence was to occur and that different psychotic-like experiences were associated with intimate partner violence to differing degrees. When compared to the participants who did not report psychotic-like experiences, the participants who reported experiencing 'thoughts inserted', 'reference ideations', and 'paranoid ideations' were more than five times more likely to report intimate partner violence. In addition, studies have found that current psychotic and/or manic symptoms increased the short-term risk for violence (Link & Steueve, 1994; Schwartz, Reynolds, Austin, & Petersen, 2003).

The relationship between personality disorder and violence is less contentious, with general agreement that people who engage in intimate partner violence are more likely to have a diagnosable personality disorder (Dutton et al., 1996; O’Leary, Malone, & Tyree, 1994; Schumacher, Feldbau-Kohn, Slep, & Heyman, 2001). This includes, specifically, psychopathic, antisocial, borderline, narcissistic, or histrionic personality structures. A meta-analysis conducted by Schumacher et al. (2001) found that of the four studies reporting on personality profiles (Beasley & Stoltenberg, 1992; Hamberger & Hastings, 1991; Hastings & Hamberger, 1994; Murphy, Meyer, & O’Leary, 1993) using the Millon Clinical Multiaxial Inventory (Millon, 1983) and Millon Clinical Multiaxial Inventory-Second Edition (Millon, 1987), all found that men with a history of intimate partner violence had elevations on the narcissistic and aggressive subscales. These studies also reported elevations on borderline (Beasley & Stoltenberg, 1992; Hamberger & Hastings, 1991; Murphy et al., 1993), avoidant (Hamberger & Hastings, 1991; Hastings & Hamberger, 1994; Murphy et al., 1993), antisocial (Beasley & Stoltenberg, 1992; Murphy et al., 1993), schizotypal (Beasley & Stoltenberg, 1992; Murphy et al., 1993), negativistic (Hamberger & Hastings, 1991; Hastings & Hamberger, 1994), gregarious (Hamberger & Hastings, 1991; Hastings & Hamberger, 1994), and paranoid subscales (Hastings & Hamberger, 1994; Murphy et al., 1993)³.

Intimate partner violence subscale. The intimate partner violence subscale assesses risk unique to intimate partner violence. It consists of 10 risk factors, with a possible score range on 0 to 20. The first seven risk factors in the subscale are related to past intimate partner violence. These risk factors are past physical assault, past sexual assault/sexual jealousy, past use of weapons and/or credible threats of death, recent escalation in frequency or severity of violence, past violation of “no contact” orders, extreme minimisation or denial of spousal assault history, and attitudes that support or condone intimate partner violence. The literature outlining the relationship between these seven risk factors and intimate partner violence is discussed below. The remaining three risk factors are related to the alleged or current incident. These risk factors are severe and/or sexual assault, use of weapons and/or credible death threats, and violation of “no contact” orders. Given that the literature highlighting the relationship between these three risk factors and intimate partner violence is discussed elsewhere in this section, the reader is directed there (pp. 39-40).

³ Only the elevations found by two or more of the studies are reported.

The first four of the seven risk factors related to past intimate partner violence relate specifically to the extent and nature of the violence. Past physical intimate partner violence is a robust predictive risk factor of future intimate partner violence (Campbell et al., 2003, Dutton & Kropp, 2000; Hilton & Harris, 2005; Saunders & Browne, 2000). In a prospective study of 541 couples Leonard and Senchak (1996) found that a history of intimate partner violence significantly predicted future intimate partner violence. In addition, people with a history of sexual assault or sexual jealousy have an increased risk for physical intimate partner violence (Campbell et al., 2003). This was highlighted in a meta-analysis, which found that the relationship between forced sex and intimate partner violence had a moderate effect size (Stith, Smith, Penn, Ward, & Tritt, 2004). Past use of weapons and credible threats of serious harm or death is associated with an increased risk for intimate partner violence (Campbell et al., 2003; Dutton & Kropp, 2000). In a study of risk factors for intimate partner homicide, Campbell et al. (2003) found that previous threats with a weapon and threats to kill were associated with a 4.1 and 2.6 increase in future intimate partner violence. Finally, an escalation in the frequency of intimate partner violence is associated with both the reoccurrence of intimate partner violence (Mahoney, Williams, & West, 2001; Weisz, Tolamn, & Saunders, 2000) and the imminence of this reoccurrence (Kropp et al., 1999).

The remaining three risk factors related to past intimate partner violence assess the attitudes and behaviours that accompany intimate partner violence. Some common examples of the topics that are minimised or denied include past physical assaults, lack of responsibility for intimate partner violence, and the health consequences for the victims. Several authors have found that perpetrators of serious and persistent intimate partner violence exhibit minimisation or denial of antisocial behaviour (Dutton, 1995; Hanson & Wallace-Capretta, 2000; Hilton & Harris, 2005). In addition, Hanson and Wallace-Capretta (2000) found that clinician ratings of minimisation were significantly associated with intimate partner violence recidivism.

Attitudes that condone aggressive behaviour is correlated with intimate partner violence, with common attitudes including those of patriarchy, misogyny, and that violence can resolve conflict and enforce control (Campbell et al., 2003; Hanson & Wallace-Capretta, 2000). While attitudes do not discriminate people who do and do not use intimate partner violence, they have been shown to be a significant predictor (Ateah, Secco, & Woodgate, 2003). For example, Kantor et al. (1994) found that attitudes accepting intimate partner violence increase the risk of that behaviour by 2.2 times. In

addition, meta-analytic research of perpetrator risk factors found that attitudes condoning violence had an overall statistically significant moderate effect size and that traditional sex-role ideology had a statically significant but small effect size (Stith et al., 2004).

Case specific considerations. The risk factors discussed above are considered the minimum that should be assessed to guide risk assessment in intimate partner violence populations (Kropp et al., 1999). Given that research on risk factors is extensive and ongoing, the authors suggest that case specific factors, not already listed, should also be assessed. For example, a history of stalking is a risk factor for intimate partner violence, with 68 percent of victims of attempted or actual intimate partner homicides reporting being stalked in the twelve months before the incident (McFarlane, Campbell, & Watson, 2002). Other examples of risk factors that are considered relevant to the nature and severity of risk include a history of torturing, maiming, or sexual sadism (Kropp et al., 1999).

Psychometric properties of the Spousal Assault Risk Assessment guide

Studies into the psychometric properties of the SARA have been well established in Northern Hemisphere populations, namely in Canada, Sweden, Spain, and the Unites States. A review of the reliability and validity studies will be presented below, with specific focus on the interrater reliability, and construct, concurrent, and predictive validates.

Interrater reliability. The interrater reliability of the SARA using file information has been explored in two studies. The total score reliability was moderate to high, with intraclass correlation coefficients of .85 ($n = 18$, Grann & Wedin, 2002) and .84 ($n = 86$, Kropp & Hart, 2000). Both studies found intraclass correlation coefficients were lower for the general violence subscale (.74 in the Grann & Wedin, 2002, study, and .68 in the Kropp & Hart, 2000, study) than the intimate partner violence subscale (.88 in the Grann & Wedin, 2002, study, and .87 in the Kropp & Hart, 2000, study). The summary risk rating reliability was lower than that of the total score (.63; Kropp & Hart, 2000), and below the recommended standard of .70 (LeBreton & Senter, 2008). The inadequate level of agreement amongst the raters could reflect either the subjectivity of the summary risk ratings, or the methodology of the study. While the original assessor interviewed the participants, the observer (independent rater) had access only to their prison file.

Using audio recorded interviews, Mowat-Leger (2001) investigated the reliability of the SARA based on 16 cases. The total, general violence subscale, and intimate partner violence subscale scores showed high reliability, all being above .90. Helmus and Bourgon (2011) argued that the difference between the results of Mowat-Leger's (2001) study and the other studies (Grann & Wedin, 2002; Kropp & Hart, 2000) likely reflects the different methods employed. That is, higher reliability is more likely when the information provided is of sufficient quality and there is similarity in the methods of the interviewer and the observer. Despite Grann and Wedin (2002) and Kropp and Hart's (2000) assertion that the files provided to the observers' contained detailed information about each participant (including the summaries of several prior interviews, psychological and psychiatric reports, and the results of previous risk assessments for general violence), a prison file collected for general purposes is different from conducting an interview for risk for intimate partner violence. It may be that the ratings were affected by the quality of the information provided. On that other hand, given that an audio recording of the interview is more similar to conducting an interview in person, compared to reviewing a file, the higher reliability in Mowat-Leger's (2001) study may reflect the reliability of the information.

Construct validity. Two studies have explored the ability of the SARA scores to differentiate between different populations of violent and non-violent men. Kropp and Hart (2000) compared the scores of inmates with ($n = 638$) and without ($n = 372$) a documented history of intimate partner violence, and found that the inmates with a history of intimate partner violence had significantly higher total score, general violence subscale, and intimate partner violence subscale scores, compared to the inmates without a history of intimate partner violence. In an extension of the Kropp and Hart (2000) study, Mowat-Leger (2001) examined whether the SARA scores differentiated between four groups, family only violent ($n = 37$), stranger only violent ($n = 35$), generally violent ($n = 41$), and non-violent ($n = 41$). The total score for the family only violent group was not significantly different than that of the general violent group, however, both were significantly higher than the stranger only violent group which, in turn, was significantly higher than the non-violent groups. Therefore, the family only violent and general violent groups were assessed as higher risk of intimate partner violence. Despite these encouraging findings, the author did not examine the discriminatory power of the summary risk ratings.

Convergent and discriminant validity. Studies establishing the convergent and discriminant validity of the SARA have found moderate to large correlations between the measure and various alternative measures of risk for both general violence and intimate partner violence. In a validation study of 1,465 men arrested for intimate partner violence offences within a nine month period in Colorado, Williams and Houghton (2004) found large correlations between the Domestic Violence Screening Instrument and both the SARA total score (.54) and summary risk ratings (.57). Wong and Hisashima (2008) found a comparable correlation (.54) between the Domestic Violence Screening Instrument and the SARA total score with a sample of 196 Hawaiian men on probation for IPV. Moderate to large correlations have also been found between the SARA total score and the Violence Risk Appraisal Guide (.33, Grann & Wedin, 2002), the Ontario Domestic Assault Risk Assessment (.60, Hilton et al., 2004), the Psychopathy Checklist-Screening Version (.43, Kroop & Hart, 2000; .57, Mowat-Leger, 2001), the Psychopathy Checklist-Revised (.59; Grann & Wedin, 2002), and the historical items of the Historical Clinical, Risk Management-20 (.46, Grann & Wedin, 2002). Therefore, the SARA appears to be measuring the proposed construct.

As would be expected, studies have demonstrated that the SARA general violence subscale has a stronger relationship with measures of risk for general violence, compared to the intimate partner violence subscale. The general violence subscale purports to measure the same construct as these general risk assessment measures. In Kropp and Hart's (2000) study, which sampled 39 men incarcerated for intimate partner violence, the correlations between the Violence Risk Appraisal Guide and the SARA were .50 for the general violence subscale and .08 for the intimate partner violence subscale—the expected pattern of relationships. Similarly, Grann and Wedin (2002), who also sampled men incarcerated for intimate partner violence, found correlations with the same measures of .49 for the general violence subscale and -.01 for the intimate partner violence subscale. This relationship has also been shown with alternative general violence risk assessment measures. For example, both Kropp and Hart (2000) and Mowat-Leger (2001) found that the correlations between the SARA and the Psychopathy Checklist-Screening Version were moderate to large for the general violence subscale, and moderate for the intimate partner violence subscale, while Grann and Wedin (2002) found that correlations with the historical factors of the Historical checklist, Risk Management-20 were large for the general violence subscale, and small for the intimate partner violence subscale. The SARA general violence subscale appears

to show convergent validity with measures of risk for general violence, while the intimate partner violence subscale shows both convergent and discriminant validities, depending on the alternative measure utilised.

In addition to establishing the concurrent validity with alternative risk assessments, the convergent validity of the SARA with past intimate partner violence behaviours has been explored. Mowat-Leger (2001) found the relationships between the SARA total score and the physical abuse and emotional abuse subscales of the Abusive Behaviour Inventory were in the large (.59) and moderate (.36) ranges, respectively. In a validation study of a brief version of the SARA, designed for use by police officers, Au et al. (2008) compared the measure to the Revised Conflict Tactics Scale. They found moderate correlations with the psychological aggression (.36) and physical assault (.38) subscales, small correlations with the sexual coercion (.03) and injury (.19) subscales, and a small negative correlation with the negotiation subscale (-.12). The latter subscale assesses non-violent conflict resolution behaviours. These findings indicated that there is a relationship between past physical and emotional abuse and risk for intimate partner violence, while the relationship between risk for intimate partner violence, as assessed by the SARA, and past sexual coercion, injuries caused, and non-violent conflict resolution behaviours requires investigation.

Predictive validity. The predictive accuracy of the SARA has received more attention than other forms of validity, with 11 studies having been published. As shown in Table 2 (p. 50), for intimate partner violence recidivism, five studies have explored the predictive validity of the SARA summary risk ratings. The area under the ROC curve ranged from .56 to .87. However, Helmus and Bourgon (2011) reported that the average weighted area under the ROC curve (.67; 95% CI .63 to .71) was difficult to interpret because of the significant variability in the values across the studies. After removing two outlier studies (.87, Andrés-P, López, & Álvarez, 2008; .56, Kropp, 2003), the average weighted area under the ROC curve was .72 (95% CI .66 to .78) and the resulting variability in the remaining three studies was non-significant. While the multiple outliers result in difficulty interpreting the predictive validity of the SARA summary risk ratings, the studies likely reflected the true variability in the accuracy of the subjective assessments (Helmus & Bourgon, 2011). That is, some assessors assigned risk more accurately than others. Eight studies examined the predictive validity of the SARA total scores, with the average weighted area under the ROC curve ranging from .63 (95% CI .60) to .65, (Helmus & Bourgon, 2011). The area under the ROC curve for

the individual studies ranged from .59 to .77. Unlike the summary risk ratings, the variability in the predictive validity across the studies was not significant (Helmus & Bourgon, 2011). In summary, the effect sizes of the predictive validity for both the SARA summary risk ratings and the total score ranged from small to large.

Table 2

SARA predictive validity studies (Helmus & Bourgon, 2011).

Study	Sample size	Recid. source	Recid. definition	Follow up <i>M</i> (months)	Total score (area under to ROC curve)			Summary risk ratings (area under the ROC curve)		
					IPV recid.	Violent recid.	Any recid.	IPV recid.	Violent recid.	Any recid.
Andrés-Pueyo et al. (2008)	102	-	-	12	.77	-	-	.87	-	-
Gibas et al. (2008)	108	-	-	Minimum 6	.70	.58	-	.76	.55	-
Glackman (2004)	2044	Criminal records	New contact	24	-	-	-	-	.66	.65
Grann and Wedin (2002)	83	Criminal records	Convictions	24	.63	-	-	-	-	-
Heckert and Gondolf (2004)	499	Partner report	Repeat assault	15	.64	-	-	-	-	-
Hilton et al. (2008)	649	Police	Reported incident	60	.59	-	-	-	-	-
Kropp (2003)	429	-	-	-	-	-	-	.56	-	-
Kropp and Hart (2000)	102	-	Charges	-	.60	-	-	.70	-	-
Reeves et al. (2008)	251	Partner report	Arrest	10	-	-	-	.69	-	-
Williams and Houghton (2004)	434	Criminal records	Arrest	18	.65	-	.70	-	-	-
Wong and Hisashima (2008)	196	-	Arrest	3	.62	.64	-	-	-	-

Note: Recid.= recidivism
 IPV= Intimate partner violence
 ROC= Receiver operating characteristic

Despite these findings, methodological and sample features vary considerably across the studies. In addition, three of the studies did not report the source of, or define the recidivism variable. Of the remaining eight studies, three sampled participants from community stopping violence programmes. Heckert and Gondolf (2004) retrospectively assessed 840 men from four cities in the United States using data from an evaluation of stopping violence programmes. As a result of the SARA being assessed from file information, only 50 percent of the items were reliably applied (Kropp & Gibas, 2010). For the total score, the area under the ROC curve for partner reported repeat assault over 15 months was .64. Kropp and Hart (2000) assessed 102 men directed by the Canadian courts to attend a stopping violence programme using correctional files, which contained summaries of previous interviews, risk assessments for general violence, and psychological reports. The authors followed the participants for 11 years, and reported the area under the ROC curve for new intimate partner violence criminal charges was .60 for the SARA total score, and .70 for the SARA summary risk ratings. Finally, Reeves, Kropp, and Cairns (2008) interviewed 251 men who attended a stopping violence programme in Canada, and found the area under the ROC curve for partner reported intimate partner violence recidivism over 10 months was .69 for the SARA summary risk ratings. A prospective study with a community sample that employed both participant interviews to rate the SARA and official records as the outcome variable has yet to be undertaken. The assessment of risk from participant's interviews reflects the recommended method of assessment, while official records, such as police contact, are less likely to be influenced by positive responding bias.

Other statistical techniques have also been employed to assess the predictive validity of the SARA. Using group comparison techniques, Kropp and Hart (2000) found that the median SARA risk ratings of the participants who went on to be charged with an intimate partner violence offence ($n = 50$) were significantly higher than the median risk rating of the participants who did not recidivate ($n = 52$). The mean SARA total scores did not significantly differentiate the groups. In a Hawaiian based study, Wong and Hisashima (2008) found that 32 percent of the high risk group and 17 percent of the combined low/moderate risk group were arrested for intimate partner violence offences over three months following the date of assessment. The association between the SARA scores and recidivism was significant. Finally, in a study that followed 108 men for at least six months following release from prison, Gibas, Kropp, Hart, and Stewart (2008) found that risk rating and intimate partner violence recidivism was

strongly related. Eight, 17, and 31 percent of the low, moderate, and high risk groups, respectively, were involved in further incidents.

Given that decisions about risk should take into consideration the likelihood, imminence, frequency, nature, and seriousness of the projected actions (Kropp, 2004), the predictive validity of the SARA in relation to these outcome criteria requires investigation. To date, two studies have extended the likelihood studies discussed. Wong and Hisashima (2008) explored the ability of the SARA total score to predict the time to intimate partner violence arrest in a sample of 196 Hawaiian men who were released from a correctional facility. The authors dichotomised the total score according to previously established cut off scores (“low/moderate” risk = eight and under and “high” risk = nine and above), and found that the participants assigned high risk ratings had a significantly shorter time to arrest than the participants assigned low/moderate risk ratings. The majority of the participants were “high” risk; the mean SARA score was 10.4. Second, Hilton, Harris, Rice, Houghton, and Eke (2007) included an investigation of the relationship between the SARA total score and number of recidivistic incidents in their validation study of 649 men arrested for an intimate partner violence offence. Recidivism was defined as reported incidents from police, corrections, and criminal records over 60 months. The correlation between the two variables was small in magnitude (.22), indicating a small association between the SARA total score and the number of recidivistic incidents. However, both of these studies are limited in that they did not investigate the accuracy of the SARA summary risk ratings and, therefore, the risk assessment was actuarial in nature and not reflective of the conceptual framework of the measure.

The Current Study

The remainder of this chapter outlines the current study, specifically the rationale and hypotheses. The primary purpose of the current study is to evaluate the reliability and validity of the SARA with a New Zealand sample. This will be achieved in three parts: first, the interrater reliability, second, the convergent and discriminant validity, and third, the predictive accuracy and incremental validity.

Part One: Interrater reliability

Part One evaluates the interrater reliability of the SARA. As shown in the literature review, previous research (Grann & Wedin, 2002; Kropp & Hart, 2000;

Mowat-Leger, 2001) has found the SARA has adequate interrater reliability when using file based reviews and high interrater reliability when using audio recording reviews. The current research will explore whether these differences in the methodology will impact on the interrater reliability coefficient. This will be achieved by comparing three methods: interview notes, audio recorded interview, and combined interview notes and audio recorded interview. The hypotheses related to Part One are outlined below.

- 1a. The SARA will have adequate interrater agreement.
- 1b. There will be an incremental increase in interrater reliability depending on the method employed, with the combined audio recording and interview method evidencing higher reliability than the audio recording method alone which, in turn, will evidence higher reliability than the interview notes alone method.

Part Two: Convergent and discriminant validity

Part Two will employ a within-participants design to evaluate the convergent and discriminant validity of the SARA. The hypotheses related to Part Two are outlined below. The magnitudes of the relationships were recommended by Cohen (1988).

Convergent validity

- 2a. Risk for intimate partner violence (as measured by the Domestic Violence Screening Instrument; Williams & Houghton, 2004) will have a strong positive relationship with the SARA total score, SARA general violence subscale and intimate partner violence subscale scores, and SARA summary risk rating, as evidenced by correlations above .5.
- 2b. Risk for general violence (as measured by the Violence Risk Appraisal Guide; Quinsey et al., 2006 and Violent Offender Treatment Programme–Risk Assessment Scale; Ward & Dockerill, 1999) will have a strong positive relationship with the SARA general violence subscale score, as evidenced by correlations above .5.
- 2c. Risk for general violence (as measured by the Violence Risk Appraisal Guide and Violent Offender Treatment Programme–Risk Assessment Scale) will have a moderate positive relationship with the SARA intimate partner violence subscale score, as evidenced by correlations between .3 and .49.

- 2d. Intimate partner violent behaviours (as measured by the Revised Conflict Tactics Scale physical assault, psychological aggression, sexual coercion, and injury subscales; Straus, Hamby, Boney-McCoy, & Sugarman, 1996) will have a moderate positive relationship with the SARA total score, SARA general violence subscale and intimate partner violence subscale scores, and SARA summary risk rating, as evidenced by correlations between .3 and .49.

Discriminant validity

- 2e. Non-violent intimate partner conflict resolution behaviours (as measured by the Revised Conflict Tactics Scale negotiation subscale) will have a small negative relationship with the SARA total score, SARA general violence subscale and intimate partner violence subscale scores, and SARA summary risk rating, as evidenced by correlations between -.1 and -.29.

Part Three: Predictive and incremental validity

Part Three will evaluate the predictive validity of the SARA total score and summary risk rating with a New Zealand sample using a prospective design. In addition, the ability of the SARA total score and summary risk ratings to predict time to recidivism will be explored. This will be achieved by examining police contact for intimate partner violence, which occurred in the 270 days (approximately 9 months) from the date of assessment, as the outcome criterion. Finally, the ability of the SARA dynamic risk factors to incrementally enhance the predictive validity of the static risk factors will be explored. The hypotheses related to Part Three are outlined below.

Predictive validity

- 3a. The mean SARA total scores for the police contact participants will be statistically higher than those of the no police contact participants.
- 3b. The median SARA summary risk rating scores for the police contact participants will be statistically higher than those of the no police contact participants.
- 3c. The SARA total and summary risk rating scores will have significant predictive accuracy for police contact during the follow-up period.
- 3d. Participants with higher scores on the SARA will have shorter time to police contact during the follow-up period.

Incremental validity

- 3e. The dynamic risk factors will not statistically increase the predictive efficacy of the static risk factors in the prediction of intimate partner violence recidivism at 270 days.

Chapter Summary

The SARA was selected for evaluation because it is a structured professional judgement measure that can be administered by people with a number of professional affiliations. As was shown, the literature was generally supportive of the inclusion of each of the risk factors as empirically valid and clinically useful. In addition, the studies conducted with international samples indicate that the SARA is reliable and valid. However, these studies are limited and voids exist in the literature as a whole. These include that the impact of method on interrater reliability coefficients, the relationship between the SARA scores and non violent behaviours used in conflict resolution, the ability of the total score and summary risk rating to predict recidivism in a community sample using official records as the outcome variable, and the incremental validity of the dynamic risk factors. The chapter then outlined the hypotheses for the present study. Some areas, particularly the convergent validity and predictive validity, aim to replicate the previous studies, while others explore new areas such as the incremental increase in interrater reliability achieved by different methodologies, the discriminant validity, ability of the SARA to predict time to police contact and the incremental validity of the dynamic risk factors. The following three chapters outline the methodology and results of the current study, and discuss the results as they pertain to the literature and the practice and research of risk assessment for intimate partner violence.

CHAPTER FIVE

METHOD

The previous chapters outlined the literature pertaining to risk assessment for intimate partner violence, culminating in the rationale and hypotheses for the three parts of the present study. Part One investigates the interrater reliability of the SARA, Part Two investigates the convergent and discriminant validity to examine the underlying construct of the SARA, and Part Three investigates the predictive validity of the SARA, generally, and the incremental validity of the SARA dynamic risk factors, specifically. The present chapter outlines the methodology, including the study participants who made up the total sample and subsample, psychometric materials, procedure used to collect psychometric and recidivism data, and an overview of the data analyses employed.

Participants

A flow diagram of the sample size at each time point is presented in Figure 2. The participants were sampled from a population of men ($n = 137$) who were waiting to commence a stopping violence programme over a period of 18 months. The stopping violence programmes were provided by community based non-governmental agencies, which accepted referrals from the courts, other agencies, and the participants themselves. Eighty-four (61%) of these men attended an intake assessment for entry to the programme and, of these, 77 (92%) were approached to participate in the study. The remaining 7 (8%) men were excluded from the study because they were referred by the programme facilitators to other agencies due to the significant role of substance use and/or mental illness in their presentation. Confidentiality of the participants excluded at this point was retained, and no information was provided to the researcher. Of the 77 who were approached to participate in the study, 48 (62%) agreed to participate, two (3%) did not meet criteria, and 27 (35%) declined to participate. One of the participants was excluded from the study because he was under 18 years of age, and the other because he did not speak English (non-English speaking groups were run at one of the programmes). Of the 48 men who agreed to participate in the study, six (13%) men reported no history of intimate partner violence. However, these men were included until their history could be verified through offence histories attained from the New Zealand Police at a later date. One of the six men had a previous charge of *male assault*

female, where the victim was an intimate partner and, as a result, was included in the study sample. The remaining five men were excluded. The total sample consisted of 43 participants (56% of those who were approached to participate). This participant rate was similar to that reported by Hetherington (2009). From the total sample 36 participants (84%) were followed up over a period of 270 days, making up the follow-up subsample.

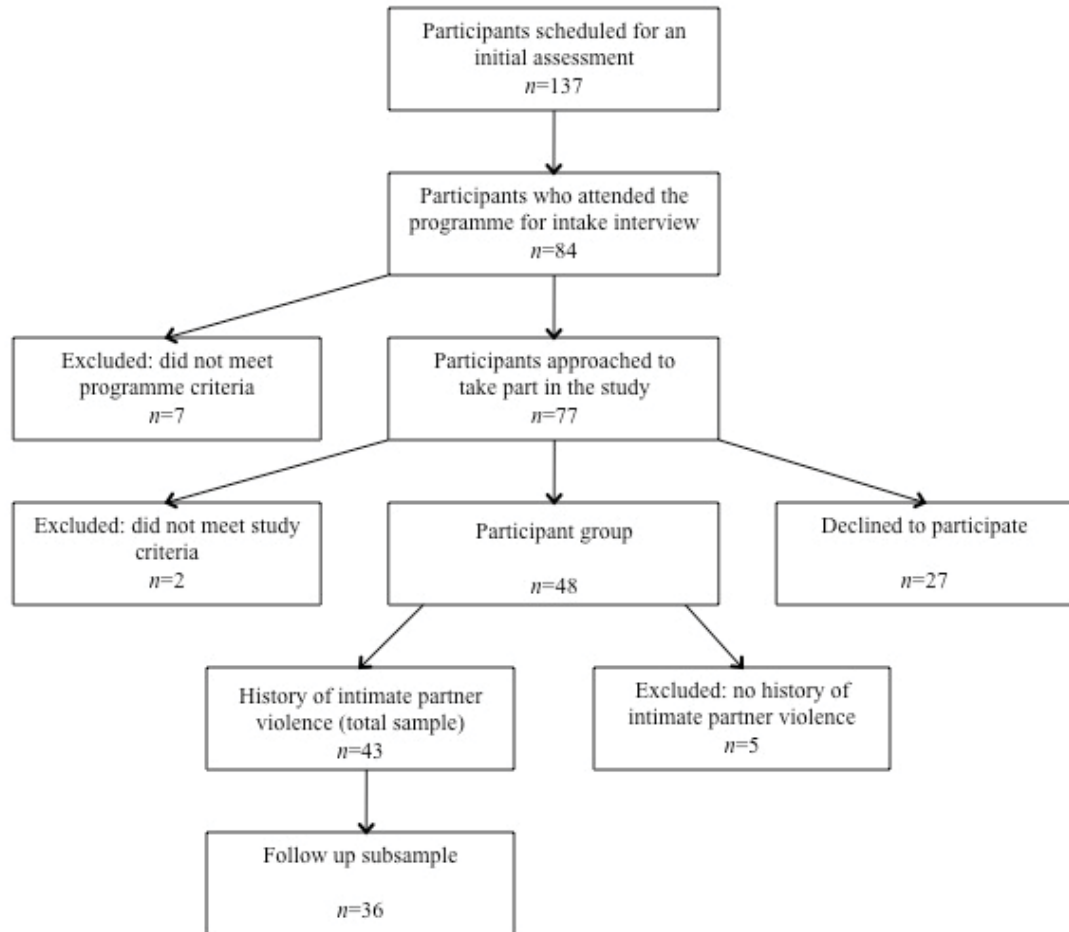


Figure 2. Sample size at each time point in the selection process.

The participants in the total sample were aged between 19 and 51 years ($M = 31.70$, $SD = 9.51$; see Table 3). The majority identified themselves as either New Zealand Māori (40%) or Pacific Island (33%). The remaining ethnicities were New Zealand European (14%), Indian (9%), South African (2%), and British (2%). Fifty-eight percent of the participants were living with an intimate partner at the time of assessment, 30 percent were separated or divorced, and 11 percent were living apart. Of the 47 percent of men who were employed 95 percent were full-time and 5 percent part-time. Forty-two percent of the men unemployed, while the remaining 12 percent

studying full-time. Most of the participants (72%) were court-referred to the stopping violence programme, followed by self-referral (26%), and referred by another agency (2%; Child Youth and Family).

Table 3

Summary of demographic characteristics for the total and follow-up participants.

Demographic	Total Participants <i>N</i> = 43	Followed Up Participants <i>n</i> = 36
AGE		
Range	19-51	19-51
Mean (<i>SD</i>)	31.70 (9.51)	31.64 (9.53)
ETHNICITY		
New Zealand Māori	17 (40%)	12 (33%)
Pacific Island	14 (33%)	12 (33%)
New Zealand European	6 (14%)	6 (17%)
Indian	4 (9%)	4 (11%)
African	1 (2%)	1 (3%)
British	1 (2%)	1 (3%)
MARITAL STATUS		
Living together	25 (58%)	21 (58%)
Living apart	5 (12%)	5 (14%)
Divorced/separated	13 (30%)	10 (28%)
EMPLOYMENT STATUS		
Full-time work	19 (44%)	17 (47%)
Part-time work	1 (2%)	1 (3%)
Unemployed	18 (42%)	13 (36%)
Studying	5 (12%)	5 (14%)
PROGRAMME REFERRAL SOURCE		
Self-referral	11 (26%)	10 (28%)
Court ordered	31 (72%)	25 (69%)
Referred by other agency	1 (2%)	1 (3%)

Note: the total percentages may not equal 100 due to rounding

Follow-up subsample

As shown in Figure 2, 36 people from the total sample of 43 participants (84%) consented to their New Zealand Police files being reviewed and, subsequently, were followed up for a period of 270 days after the assessment date. The average age of the follow-up subsample was 31.64 years ($SD = 9.53$, range 19-51 years; see Table 3). Most identified as either New Zealand Māori (33%) or Pacific Island (33%), were court-referred (69%), living with an intimate partner (58%), and employed in a full-time position (47%). The total sample and follow-up subsample groups were not significantly different on any of the demographic characteristics. The participants who were lost to follow-up tended to be New Zealand Māori, living with an intimate partner, unemployed and court-referred.

Interviewers and observers

Before the data collection began, the researcher, research assistant and observers received formal training in administering and scoring the SARA, provided by one of the supervisors. The research assistant was a Masters student studying psychology, two of the observers' were Doctor of Clinical Psychology candidates who had completed their final clinical internships (that is, were in the fourth year of their doctorate), and one observer was a PhD candidate studying psychology. In addition, the latter observer also had experience administering the SARA in Canada, where she worked in the correctional field.

The training covered the theoretical basis of risk assessment, actuarial and structured professional judgement measures, and an introduction to the SARA, which incorporated practical components. In addition, the research assistant was trained in conducting the semi-structured interview. This involved two role plays with the researcher, twice observing the researcher, and having the researcher observe three of the research assistant's interviews. The audio recorded interviews were then reviewed to ensure the interview was conducted accurately during data collection. Feedback was provided if and when it was necessary.

Measures

The psychometric measures assessed three constructs. First, risk for intimate partner violence was assessed using the SARA (Kropp et al., 1994, 1995, 1999) and Domestic Violence Screening Instrument (Williams & Houghton, 2004). Second, risk

for general violence we assessed using the Violence Risk Appraisal Guide (Quinsey et al., 2006) and Violent Offender Treatment Programme-Risk Assessment Scale (Ward & Dockerill, 1999). Finally, both violent and non-violent conflict resolution behaviours were assessed using the Revised Conflict Tactics Scale (Straus et al., 1996). A summary of each measure is provided in the following section.

Spousal Assault Risk Assessment guide

The SARA (Kropp et al., 1994, 1995, 1999) is a clinical checklist of static and dynamic risk factors for intimate partner violence. The measure consists of 20 risk factors, each of which has demonstrated clinical and empirical validity (Kropp et al., 1999). As a moderate level of specificity was aimed for, the factors are at the trait, characteristic, or incident level, as opposed to specific or detached behavioural acts. The 20 risk factors are divided into two parts; the general violence subscale (items 1-10) assesses risk for general violence while the intimate partner violence subscale (items 11 to 20) assesses risk for intimate partner violence. The first subscale includes criminal history and psychosocial adjustment, while the second subscale includes intimate partner violence and current offence. In addition, the SARA allows for other considerations, which are not directly assessed by the 20 risk factors, to be incorporated.

From the 20 risk factors three scores are derived. The first is an aggregated score of the positively endorsed risk factors, which are derived from the rating of each of the 20 risk factors based on information from the assessment. The 20 risk factors are rated on a three point scale of severity (0 = absent, 1 = sub threshold, 2 = present), based on descriptive criteria unique to each. The risk factor scores are then summed to give a total score ranging from 0 to 40.

Investigating the items deemed to be critical factors derives the second summary score. Critical factors are those that, given the circumstances in the case at hand, are sufficient on their own to compel the assessor to conclude that the person poses an imminent risk of harm, for example 'threatens to kill'. Kropp et al. (1999) argued that they are included because risk, as perceived by the assessor, is not a simple linear function of the number of risk factors present in a case: it is conceivable that an assessor could judge a person to be at high risk for violence on the basis of a single critical item. Critical items are rated on a two point scale of 0 (absent) or 1 (present).

Assessing the risk posed by the person to others derives the third summary score. The summary risk rating is rated on a three point scale of severity, where a score

of one indicates low risk, two indicates moderate risk, and three indicates high risk. The authors noted that the risk rating refers to the imminence of harm to both an intimate partner and others, however they do not define the levels of risk (Kropp et al., 1999). The summary risk ratings are derived from professional judgement, where the interviewer relies on clinical discretion to assign a risk rating. In general, and especially in the absence of critical factors, risk can be expected to increase with the number of factors that are coded “present” (Kroop et al., 1999). The total score, subscale scores, individual risk factors (dynamic and static), and risk rating were employed in this study.

Psychometric evaluations of the SARA have shown it has adequate interrater reliability, while assessments of the concurrent, construct, and predictive validities have been encouraging. See the previous chapter for a specific discussion of these psychometric properties.

Domestic Violence Screening Instrument

The Domestic Violence Screening Instrument (Williams & Houghton, 2004) is a 12 item measure that assesses risk for intimate partner violence. Each item is scored on either a three (0-2) or four (0-3) point scale of severity, with stronger weight assigned to items that were considered to have stronger associations with intimate partner violence (Williams & Houghton, 2004). The item scores are added together to give a total score, with a possible range of 0 to 30. A score of nine or more represents high risk, seven to eight moderate risk, and six or below low risk. Unlike the SARA, the Domestic Violence Screening Instrument was designed to be scored from file material and was validated without interviewing the participants. In the current study the interview notes were utilised to code the measure in line with the validation study.

Reliability assessments of the measure have been positive. Alpha coefficients for the total scale have been reported at .71 (Williams & Houghton, 2004). The construct validity has been established with small to moderate correlations between the Domestic Violence Screening Instrument and measures of general violence (see Hilton et al., 2007; Williams & Houghton, 2004). The area under the ROC curve indicated that the Domestic Violence Screening Instrument significantly predicted intimate partner violence recidivism (.61, Hilton et al., 2007; .68, Williams & Houghton, 2004).

Violence Risk Appraisal Guide

The Violence Risk Appraisal Guide (Quinsey et al., 2006) is an actuarial measure developed to assess dangerousness in male offenders, namely general violence. The measure consists of 12 items that are weighted according to fixed algorithms to yield a total score, ranging from -26 to 38. The severity of risk is derived through three broad scoring bands: -26 to -8 low, -7 to 13 moderate, and 14 to 38 high. Each item is an empirically derived predictor of violent behaviour, with three related to the person's development, five to current and past offences, and the remaining four to the diagnosis of personality disorder, diagnosis of schizophrenia, marital status, and failure of prior conditional release. To score item 12 in the current study, the weighted score of the Hare Psychopathy Checklist-Revised total was replaced with the Child and Adolescent Taxon Scale, as recommended by the authors (Quinsey et al., 2006). The latter is a nine item measure of childhood and adolescent adjustment (for scoring information see Quinsey et al., 2006).

Interrater reliability for the Violence Risk Appraisal Guide total score and individual items has been established, ranging from .80 to .90 (Rice & Harris, 1997; Rice, Harris, & Cormier, 1992; Rice, Harris, & Quinsey, 1990). Several studies have shown that it is an adequate predictor of recidivism, with area under the ROC curve ranging from .75 to .90 (Douglas, Yeomans, & Boer, 2005; Harris & Rice, 2003; Harris, Rice, & Cormier 2002; Harris et al., 2003; Kroner & Mills, 2001). In relation to intimate partner violence recidivism specifically, the average area under the ROC curve was .75 (Grann & Wedin, 2002).

Violent Offender Treatment Programme-Risk Assessment Scale

The Violent Offender Treatment Programme-Risk Assessment Scale (Ward & Dockerill, 1999) is a brief actuarial screening instrument of risk for general violence, which determines treatment eligibility. The measure consists of seven items. Items one to five relate to previous and current criminal offences, and items six and seven to drug and alcohol misuse. The individual items are weighted unevenly to give a total score, with a range of 2 to 30. The total score above the cut-off of 11 determines the presence of high risk. While the authors identified a cut-off of 15 as the point that minimised misclassification for 36- and 60-month intervals, the use of this cut-off results in "extremely poor sensitivity" (Ward & Dockerill, 1999, p. 135). At a cut-off of 11 the sensitivity is raised to an acceptable level, but the rate of false positives also increases.

Little has been published about the Violent Offender Treatment Programme-Risk Assessment Scale; however, interrater reliability was .82 (Ward & Dockerill, 1999) and the correlation with the Violence Risk Appraisal Guide was .50 (Douglas et al., 2005). Predictive accuracy for violent recidivism has been established between .61 (Douglas et al., 2005) and .76 (Ward & Dockerill, 1999).

Revised Conflict Tactics Scale

The Revised Conflict Tactics Scale (Straus et al., 1996) is a measure of interpersonal conflict behaviours commonly used in intimate relationships. The behaviours are scored in reference to the last year. It consists of 78 items that are scored to produce five scales: physical assault, psychological aggression, sexual coercion, injury, and negotiation. Each item is scored on a seven point scale from zero to seven (has never happened, happened once in the last year, two times in the last year, three to five times, six to 10 times, 11 to 20 times, more than 20 times in the last year, or, alternatively, has happened but not in the last year), with each coded for both the participant and their partner's behaviour. The physical assault scale contains 12 items, with a possible range for both the participant and their partner of 0 to 300, the psychological aggression scale eight items (range 0 to 200), sexual coercion scale seven items (range 0 to 175), injury scale six items (range 0 to 150), and negotiation scale six items (range 0 to 150).

Psychometric assessments of the measure have found moderate to high test-retest reliabilities. Vega and O'Leary (2007) reported correlations of .76 for the physical assault, .69 for the psychological aggression, .30 for the sexual coercion, .70 for the injury, and .60 for the negotiation scales. The internal consistency of the measure has been found to be strong. Alpha coefficients for the five scales ranged from .79 for the psychological aggression scale to .95 for the injury scale (Straus et al., 1996). The Revised Conflict Tactics Scale had positive small to moderate correlations with a brief police version of the SARA was positive for all the scales except negotiation, which was small and negative (Vega & O'Leary, 2007).

In the current study, Cronbach's alpha and inter-item correlation coefficients were used to evaluate a series of reliability analyses for the Revised Conflict Tactics Scale subscales (see Table 4). These results suggest adequate internal consistency, and are above the recommended minimum value of .70 (Streiner, 2003) or the optimal

range for the inter-item correlations for subscales that have less than 10 items (.2 to .4; Briggs & Cheek (1986).

Table 4

Reliability analyses for the Revised Conflict Tactics Scale subscales.

Revised Conflict Tactics Scale subscale	<i>N</i> (cases)	<i>N</i> (items)	α
Physical assault	42	12	.96 ^a
Psychological aggression	43	8	.40 ^b
Sexual coercion	40	7	.56 ^b
Injury	43	6	.25 ^b
Negotiation	43	6	.22 ^b

Note: ^a Cronbach's alpha

^b Inter-item correlation coefficients

Procedure

Ethical approval to conduct this study was obtained from the Massey University Human Ethics Committee: Northern (09/05). In addition, consent to approach potential participants was granted by the managers of three stopping violence programmes, two in Auckland and one in Wellington, and consent to access data from the New Zealand police database was gained from the Research and Evaluation Steering Committee within that organisation.

Potential participants were invited to take part in the study by intake clinicians at the three stopping violence programmes. However, ultimately all of the participants included in this study were recruited from one of the stopping violence programmes in Auckland, as all potential participants from the Wellington programme and the other Auckland programme were excluded during the screening process. Of the participants invite to take part in this study ($n=77$), 95 percent were recruited from the one stopping violence programme the sample resulted from. The exclusion criteria for the current study were being under the age of 18 and not speaking English. The inclusion criterion was having a history of intimate partner violence.

After participating in an initial interview for entry into the stopping violence programmes, the men were asked by programme staff if they would participate in an independent study evaluating the SARA with a New Zealand population. They were instructed that the researcher (or the research assistant) was on site to meet with them

and provide more detail about the study if they were interested. If the potential participants did express interest they were introduced to the researcher; if not interested, the researcher was instructed and the decline was recorded. Interested potential participants were next provided with both verbal and written information outlining the purpose of the study (see Appendix B).

Prior to taking part in the study the participants signed a consent form (see Appendix B). In addition, all of the participants were asked to consent to two further procedures: the release of information held by the New Zealand Police (see Appendix B) for the follow-up part of the study and the audio recording of their interview for the purpose of investigating the interrater reliability of the SARA (see Appendix B). Participation in these additional procedures was voluntary, and did not impact on their participation in the main study. In total 36 participants consented to the release of information held by the New Zealand police, and eight (19%) consented to their interview being recorded. Having consented to participate, the participants either took part in the study immediately at that point, or organised an appointment for a later date. In either instance the interview was undertaken in a room provided by the stopping violence programme, during business hours. The researcher completed 26 (60%) of the interviews and one research assistant conducted the remainder. Data was collected from each participant using a semi-structured interview (see Appendix C, based on Mowat-Leger, 2001) and a self-report questionnaire, which took between one and two hours to complete. The areas covered in the interview included demographic information (age, ethnicity, marital status, employment status, and referral source), past and current offences, school and work history, current mental health, family history, current and historical substance use, intimate relationship history, and general questions regarding their attitude towards criminality and intimate partner violence. After completing the interview the participants filled in the Revised Conflict Tactics Scale. Each participant was then given a \$20 Motor Trade Association voucher to compensate for his time.

Data was collected from the New Zealand police national headquarters, drawn from their electronic database to provide collateral information for the six participants who reported no history of intimate partner violence and inform the follow-up part of the study. First, the six files for the participants who reported no history of intimate partner violence were located. A review of their intimate partner violence police contact histories, before the date of assessment, highlighted a previous conviction of *male assaults female* for one of the participants. Because of this, the participant was included

in the study. The remaining five participants were excluded from the study and their data was removed. Second, the 36 files for the participants who agreed to the release of information held by the New Zealand Police were located. For each of these participants any police incidents after the date of assessment were recorded. Specifically, this included the incidence code, date, and whether or not the victim was an intimate partner. Hemphill, Hare, and Wong (1998) suggested that in order to increase recidivism base rates and improve the power of the analyses broad definitions of violence should be utilised. As such 'domestic dispute', where the victim was identified as an intimate partner in the summary, was included. All names were replaced with the identification numbers assigned to each participant before the data was removed from the premises. This data was then computed into two categories (any incidents of police contact for an intimate partner violence offence and time in days to first police contact for an intimate partner violence offence).

Finally, to assess interrater reliability three independent observers used data collected from eight participants, who consented to their interview being audio recorded, to carry out independent SARA scoring. Each observer used one of the three data methods (interview notes alone, audio recording alone, or both interview notes and audio recording) to complete the SARA scoring sheets for each of the eight participants. The observers were considered interchangeable, so that not all of observers scored the SARA for all of the participants using all of the methods. That is, observer identity was not considered a variable. This meant that each observer rated each participant once. See Appendix D for a table outlining the participant number and data source assigned to each observer.

Planned Analyses

Sample size

The current sample size, which reflected the number of potential participants that eventuated during the research period, impacted on the study in two ways. First, it meant that some analyses that required a minimum number of participants could not be conducted, including a factor analysis and discriminant function analysis. Second, it meant that the chance of detecting even a large population effect size at an alpha of .05 was less than 80 percent for the group difference analyses (Cohen, 1992). That is, the power was insufficient. Despite this the group difference analyses were still conducted. Therefore, the significant results will reflect true differences in the samples, while it will

be unclear whether non-significant findings were accurate, or whether they represent a Type II error. That is, there may have been genuine differences in the underlying population, which the current study would fail to find.

Missing data

The statistical analyses were conducted using the statistical software package Predictive Analytics SoftWare statistics 18.0 (SPSS Inc., 2009). The Revised Conflict Tactics Scale physical assault and sexual coercion subscales contained one (2%) and two (5%) cases with missing data, respectively. Given that McCarroll et al. (2000) found that people who omitted questions on the Revised Conflict Tactics Scale physical assault subscale were likely to have engaged in the behaviour but chose to not report it, the data in the current study is likely not missing at random. That is, the participants intentionally did not answer all of the questions. While Scheffer (2002) argued that multiple imputation is the only effective method of addressing data that is not missing at random, the author of the Revised Conflict Tactics Scale (Straus, 2004) recommended that the data not be replaced. Therefore, rather than multiple imputation, pairwise deletion was employed in the current study to exclude cases for these specific analyses.

Measurement levels

The level of data guides which statistical analysis is appropriate. In the current research all of the scales are ordinal. That is, the data had order, but the intervals between scale points may be uneven. For example, in terms of the SARA summary risk rating, a person assessed as high risk (risk ratings = 3) reflects a greater risk of recidivism than both a moderate risk (risk ratings = 2) and low risk (risk ratings = 1). However, the distance between high risk and low risk cannot be assumed to be exactly twice that of the distance between a moderate risk and a low risk.

It is recommended that nonparametric statistics be used when the data level is ordinal (Tabachnick & Fidell, 2007). In practice, however, parametric statistics are often used, and have been previously with measures similar to those used in the study. For example, three previous studies (Grann & Wedin, 2002; Kropp & Hart, 2000; Mowat-Leger, 2001) of the SARA utilised intraclass correlation coefficients to analyse the interrater reliability, despite this being a parametric technique. According to Tabachnick and Fidell (2007) ordinal variables can be treated as interval when three conditions are met: the underlying scale is thought to be continuous but is measured

with an ordinal scale, the number of values on the scale exceeds seven, and no other assumptions of the analysis being used are violated. Therefore, it was planned that if the data did not violate other assumptions specific to the analyses (see below), the SARA and other measures (Domestic Violence Screening Instrument, Violent Offender Treatment Programme-Risk Assessment Scale, and Violence Risk Appraisal Guide, and Revised Conflict Tactics Scale physical assault, psychological aggression and sexual coercion subscales total and subscale) data would be treated as interval scales. That is, parametric analyses would be utilised. The SARA summary risk ratings and Revised Conflict Tactics Scale injury and negotiation subscales were treated differently because they violated the aforementioned conditions. That is, the number of values on these scales was less than seven. Therefore, they were treated as ordinal and non-parametric analyses utilised, unless there was clear direction in the literature that parametric analyses could be used.

Assumption checks in the total sample

Preliminary analyses would be performed to ensure there was no violation of assumptions of normality, linearity, homoscedasticity, outliers, and multicollinearity in the total sample data (SARA scores, Domestic Violence Screening Instrument, Violent Offender Treatment Programme-Risk Assessment Scale, and Violence Risk Appraisal Guide, and Revised Conflict Tactics Scale subscales). In addition, reliability analyses would be carried out with the Revised Conflict Tactics Scale subscales, which are multi item scales of singular risk factors. For the subscales with ten or more items Cronbach's alpha would be employed. However, given that Cronbach's alpha is sensitive to the length of the measures mean inter-item correlations would be reported for shorter measures (fewer than 10 items; Pallant, 2007).

Part One: Interrater reliability

It was planned that in order to explore the interrater reliability of the SARA scores (total score, general violence and intimate partner violence subscales, and summary risk rating) intraclass correlation coefficients would be computed to determine the level of agreement between the interviewer (in-person) and the observers' (independent raters) ratings. Intraclass correlation coefficients would be used, rather than alternative methods, for three reasons. First, it is the most widely used method with interval level data (Tinsley & Weiss, 1975). Second, the coefficients provide reliability

estimates that account for both chance and systematic differences between observers (making them more appropriate than Pearson product moment correlations, Spearman's rank-order correlations, or percent agreement; Shrout & Fleiss, 1979). Finally, because intraclass correlation coefficients make no assumptions about observer means, it is preferred when the sample size is small (<15; Walter, Eliasziw, & Donner, 1998). In addition, it is the method that was employed by Grann and Wedin (2002), Kropp and Hart (2000), and Mowat-Leger (2001) in previous studies of the interrater reliability of the SARA.

Despite the ordinal nature of the SARA summary risk rating, intraclass correlation coefficients would be used with this data. Uebersax (2010) noted that both Kappa measures of agreement and intraclass correlation coefficients can be performed with ordered category data, however intraclass correlation coefficients would allow the differences between the methods to be assessed for significance. For all of the analyses a two-way mixed absolute agreement model would be used, with all the judges rating each participant. To determine if the interrater reliability of each method (interview notes alone, audio recording alone, combined interview notes and audio recording) increased incrementally the correlation coefficients would be converted into z scores, before the significant differences would be assessed using z tests.

If correlation coefficients below .70 were found post hoc analyses would be conducted to test if the disagreement was caused by rater bias. Systematic under- or over-rating of the participants by the observers' would be explored both graphically and statistically, using a between subjects ANOVA across the three observers.

Part Two: Convergent and discriminant validity

It was planned that correlation coefficient matrixes would be used to investigate the strength and direction of the relationship between the SARA scores and several criteria, including risk for general violence, risk for intimate partner violence, violent conflict resolution behaviours, and non-violent conflict resolution behaviours. Specifically, parametric (Pearsons product moment) and non-parametric (Spearman's rank-order) correlations would be computed between the SARA scores and the Domestic Violence Screening Instrument, the Violence Risk Appraisal Guide, the Violent Offender Treatment Programme-Risk Assessment Scale, and the Revised Conflict Tactics Scale subscales.

Part Three: Predictive and incremental validity

The predictive validity of the SARA total score and summary risk rating would be investigated using two external criteria: any police contact for an intimate partner violence offence in the follow-up period, and time to first intimate partner violence police contact. In addition, the incremental validity of the SARA dynamic factors, compared to the static factors, would also be investigated. The subsample would provide the data for both of these lines of exploration.

Assumption checks in the follow-up sample. It was planned that preliminary analyses would be performed to ensure there was no violation of the assumptions of normality, homogeneity of variance, univariate or multivariate outliers, linearity, or absence of multicollinearity.

Predictive validity. It was planned that an independent samples t-test and the non-parametric alternative Mann Whitney U test, would be performed to compare the total score means and summary risk rating medians of the two groups (police contact and no police contact). Following this ROC curve analysis was performed to provide information as to the predictive validity of the SARA total score and summary risk rating. These analyses would also indicate the rate of true positives versus false positives.

In relation to the second criterion, time to police contact for an intimate partner violence offence during the follow-up period, it was planned that Kaplan-Meier survival analysis would be used. Two analyses would be conducted, one for the total factors (two categories, high risk and low risk) and one for the summary risk rating factors (three categories, high risk, moderate risk, and low risk). The cut off point used to dichotomise the total score would be derived from the ROC curve analysis, with equal weight given to the sensitivity and specificity. While it is in society's interest to maximise the true positive rate, it is also important to minimise the false positive rate to protect the rights of offenders. For each analysis the factors would then be compared using Log-rank tests of equality. This would assess the equality of the function between different categories of the factor variable.

Incremental validity. It was planned that multiple runs of exact logistic regression analyses would be computed to investigate the ability of the SARA dynamic risk factors to incrementally enhance the predictive validity of the static risk. The first run would predict police contact at 270 days from the static risk factors. The second run would predict police contact at 270 days from both the static and the dynamic risk

factors. The difference between the two models would be evaluated using the goodness-of-fit χ^2 process. As this process would use the loglikelihood statistic from the unconditional regression output, the results would be an approximation. Exact logistic regression is computed with small sample sizes, which allows for multivariate analysis of a dichotomous dependent variable without the large sample size requirements of unconditional logistic regression (Mehta & Patel, 1995).

CHAPTER SIX

RESULTS

The current chapter begins with preliminary analyses, including assumption checks and descriptive statistics for the data obtained from the total sample, and internal consistency of the Revised Conflict Tactics subscales. The chapter is then divided into three parts. Part One explores the interrater reliability of the SARA. Specifically, this part examines whether differences in the methodology impact on the reliability coefficient. Part Two explores the convergent and divergent validity of the SARA using several external criteria. Finally, Part Three begins with further preliminary analyses for the data obtained from the follow-up subsample. This is followed by an exploration of the predictive and incremental validity. Specifically, the predictive validity analyses examine the ability of the SARA to predict both police contact for an intimate partner violence offence during the follow-up period and time to first police contact for an intimate partner violence offence, while the incremental validity analyses examine the ability of the dynamic risk factors to enhance to predictive validity of the static risk factors.

Preliminary Analyses

Assumption checks

Normality of data was established for the SARA total score, general violence and intimate partner violence subscale scores, and summary risk rating, using the Kolmogorov-Smirnov statistic (.16, .13, .19, and .30, respectively) and normal probability plots. The Kolmogorov-Smirnov statistic total score, intimate partner violence subscale score, and summary risk rating suggested that the variables were significantly non-normal (all $p < .01$), and slightly positively skewed. However, inspection of the normal probability plots (see Appendix E) for these variables suggested that the deviation was minimal. Scatterplots of the standardised residuals were centred on the zero point and rectangularly distributed.

Evaluation of distribution of total scores on the other risk assessments (Violence Risk Appraisal Guide, Violent Offender Treatment Programme-Risk Assessment Scale and Domestic Violence Screening Instrument), using the Kolmogorov-Smirnov statistic, showed that the data for the Violence Risk Appraisal Guide (.13) were normally distributed and the Violent Offender Treatment Programme-Risk Assessment Scale and

Domestic Violence Screening Instrument (.14 and .19, respectively) were significantly non-normal (both $p < .05$), with positive skew. However, inspection of the normal probability plots (see Appendix E) and scatterplots suggested that the deviation was minimal.

Normality of the Revised Conflict Tactics Scale negotiation (.08) subscale score was established using the Kolmogorov-Smirnov statistic, while the physical assault (.39), psychological aggression (.16), sexual coercion (.35), and injury (.32) subscales were significantly non-normal (all $p < .01$). Inspection of the normal probability (see Appendix E) and scatterplots for these non-normal variables confirmed deviation from normality. All of the non-normal subscales had positive skew and kurtosis, indicating the scores were clustered around lower values and had little variance. For the non-normally distributed variables both parametric (Pearson product-moment correlation coefficients) and non-parametric (Spearman's rank-order correlations coefficients) tests were conducted, with the intention that if the results were similar and statistically significant then the parametric results would be reported, as they are more powerful and are robust to violations of their assumptions (Field, 2009).

Boxplots revealed outliers for one case on the Violence Risk Appraisal Guide, one case on the SARA total score, two cases on the Revised Conflict Tactics Scale psychological aggression subscale, and two cases on the Revised Conflict Tactics Scale sexual coercion subscale. In total there were five extreme outliers, where the values were more than three standard deviations from the mean, on the physical assault ($n = 1$), psychological aggression ($n = 1$), sexual coercion ($n = 1$), and injury ($n = 2$) subscales of the Revised Conflict Tactics Scale. Extreme cases, where the standard scores are more than 3.29 standard deviations from the mean, can inflate or deflate correlation coefficients between variables, and influence assumptions of normality (Tabachnick & Fidell, 2007). Following Tabachnick and Fidell (2007), these cases had their values changed to within three standard deviations for the statistical techniques that explored the relationships amongst these variables, performed in relation to Part Two.

Descriptive statistics

The mean, standard deviation, standard error and range for the SARA (total score, general violence and intimate partner violence subscale scores, and summary risk rating), Domestic Violence Screening Instrument, Violence Risk Appraisal Guide, Violent Offender Treatment Programme–Risk Assessment Scale, and Revised Conflict

Tactics Scale (physical assault, psychological aggression, sexual coercion, injury and negotiation subscales) scores for the total study sample are reported in Table 5. The SARA summary risk rating mean was in the low risk range. The mean for the Domestic Violence Screening Instrument total score indicated high risk, while the means for the Violence Risk Appraisal Guide score and Violent Offender Treatment Programme-Risk Assessment Scale score indicated moderate and low/moderate risk, respectively.

Table 5

Descriptive statistics for the measures in the total sample (N = 43).

Measure	Mean	Std. Deviation	Std. Error	Range
Spousal Assault Risk Assessment guide				
Total score	14.14	7.87	1.20	4-36
General violence subscale	7.36	4.04	0.62	2-18
Intimate partner violence subscale	6.83	4.14	0.64	2-18
Summary risk rating	1.84	0.65	0.10	1-3
Domestic Violence Screening Instrument	9.81	4.20	0.64	4-18
Violence Risk Appraisal Guide	2.49	7.10	1.08	-11-21
Violent Offender Treatment Programme-Risk Assessment Scale	8.42	4.19	0.64	2-18
Revised Conflict Tactics Scale				
Physical assault ^a	9.55	21.12	3.26	0-127
Psychological aggression	41.37	36.89	5.63	0-161
Sexual coercion ^b	8.98	19.48	3.04	0-90
Injury	2.16	3.95	0.60	0-20
Negotiation	77.44	35.28	5.38	16-250

Note: ^a n = 42

^b n = 40

Part One: Interrater reliability

Consensus estimates of interrater reliability were carried out using intraclass correlation coefficients for the SARA scores using three sources of information (interview notes alone, audio recording alone, and combined audio recording and interview notes; see Table 6). According to Montgomery, Graham, Evans, and Fahey

(2002) no universal guideline for describing intraclass correlation coefficients has been accepted in the literature. Despite this, those authors developed a scale (Montgomery et al., 2002), which was employed in the current study. It includes: < .20 'slight agreement'; .21 to .40 'fair agreement'; .41 to .60 'moderate agreement'; .61 to .80 'substantial agreement'; > .80 'almost perfect agreement'. Overall, interrater reliability increased from the interview notes alone to the audio recording alone, and then decreased from the audio recording method to the combined interview notes and audio recording method.

As seen in Table 6, the SARA total score agreement was highest when the observers had access to the audio recording alone, compared to interview notes alone and combined interview notes and audio recording methods. Similarly, the SARA general violence and intimate partner violence subscale agreement was highest with access to the audio recording alone. The latter was .40. However, on closer inspection of the raw data it could be seen that one observer in the intimate partner violence subscale audio recording alone method had a different rating pattern than the other observers (see Figure 3 for a descriptive display). When the intraclass correlation coefficient was recalculated without this particular observer's rating, the interrater reliability increased to .82. Finally, the interrater reliability of the summary risk rating for all three methods was in the slight agreement to fair agreement ranges, with the highest agreement with access to the audio recording alone. Overall, three of the intraclass correlation coefficients were in the substantial agreement or almost perfect agreement ranges. These included the total score, and general violence subscale coefficient from the audio recording alone method, and the general violence subscale from the combined interview notes and audio recording method.

Table 6

Intraclass correlation coefficients between the SARA scores using different methods of agreement assessment.

Method	SARA			Summary risk rating
	Total score	GV	IPV	
Notes	.29	.30	.13	.18
Audio	.84	.92	.82 ^a	.28
Combined	.36	.79	.02	.10

Note: SARA = Spousal Assault Risk Assessment guide
 GV = General Violence subscale
 IPV = Intimate Partner Violence subscale
 Notes = Interview notes alone
 Audio = Audio recording alone
 Combined = Combined interview notes and audio recording
^a = Anomalous rating removed

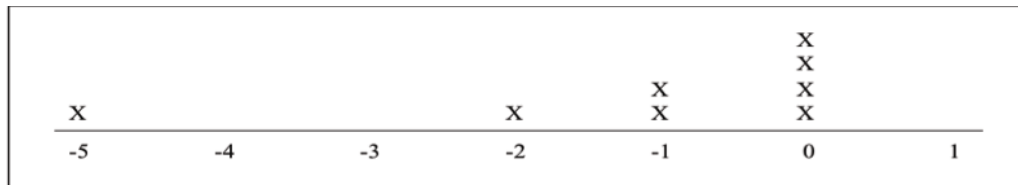


Figure 3. Distribution of the distance from the interviewers' intimate partner violence subscale score for the observers using the audio recording alone method.

Incremental increase in interrater reliability according to the method employed. To determine if the interrater reliability significantly increased as the observers' had access to more information z-tests were performed. Significant differences for the general violence subscale were found between audio recording alone method and both the interview notes alone, $z = 3.92, p < .01$, and the interview notes and audio recording combined, $z = 2.66, p < .01$, methods. The audio recording method produced significantly *better* agreement between the interviewer and the observers than the other two methods for the SARA general violence subscale, even though they had access to less information than ratings based on both audio recording and interview notes. No other significant differences were found. These results indicated that the agreement was significantly higher when the observers used audio recording alone to score the general violence subscale, compared to other two methods.

Rater bias. Post hoc analyses of rater bias was undertaken to determine if any observer systematically under- or over-rated the SARA, compared to the interviewer's (in-person) ratings. First, descriptive plots of the distribution of the distance from the original score for the three observers' for each of the SARA scores (total score, general violence and intimate partner violence subscales, and summary risk rating) was examined. As seen in Figure 4 (and subsequent Figures 5 through 7) the mean difference for the total score for all of the observers' was negative (on average the observers scored lower than the interviewer), with the mean differences for observer one and three lower than that of observer two. The general violence plot showed that the mean differences for observer one and three were negative, while the mean difference for observer two was 0 (no discrepancy between the two ratings). Observer three appeared to have an outlier in their scoring, with one rating 4 points below the remainder. The intimate partner violence plot showed that the mean differences for all of the observers' were negative. Observer one and three scored more cases at distance from the interviewer, while four of the eight ratings made by observer two matched the interviewer's ratings. The summary risk rating plot showed that observer one and three scored five of the eight participants one summary risk rating below that of the interviewer and observer two scored one of the eight participants at the same distance. Overall, subsequent ratings based on all secondary sources tended to result in lower risk ratings than the interviewers.

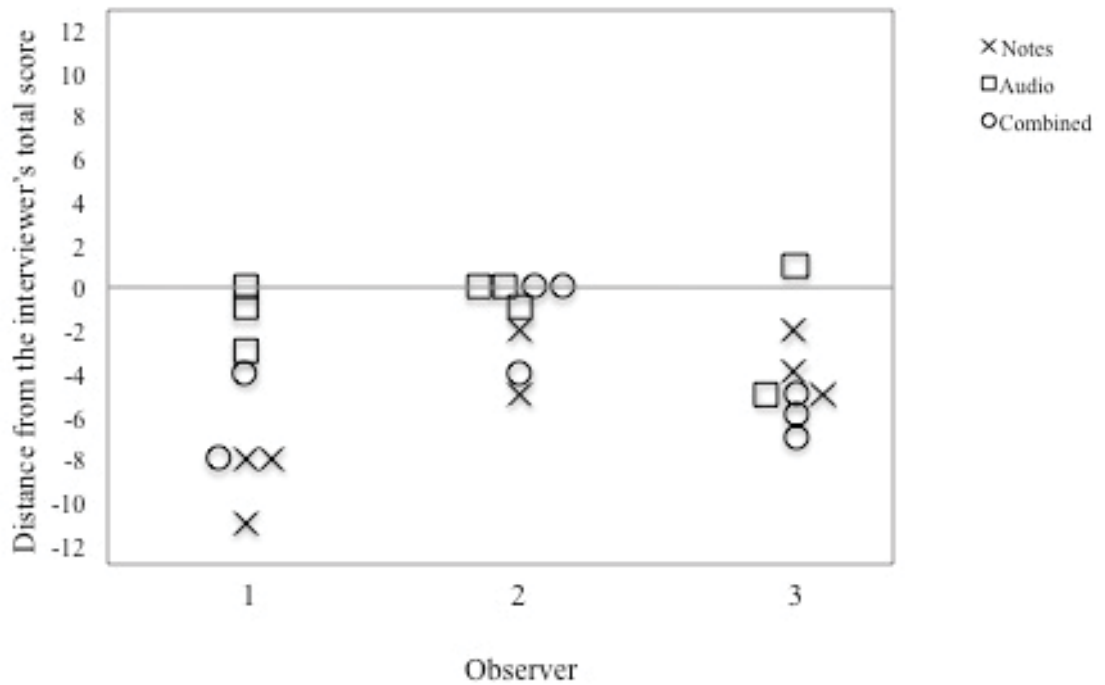


Figure 4. Distribution of the distance from the interviewer's total score for the three interrater reliability methods.

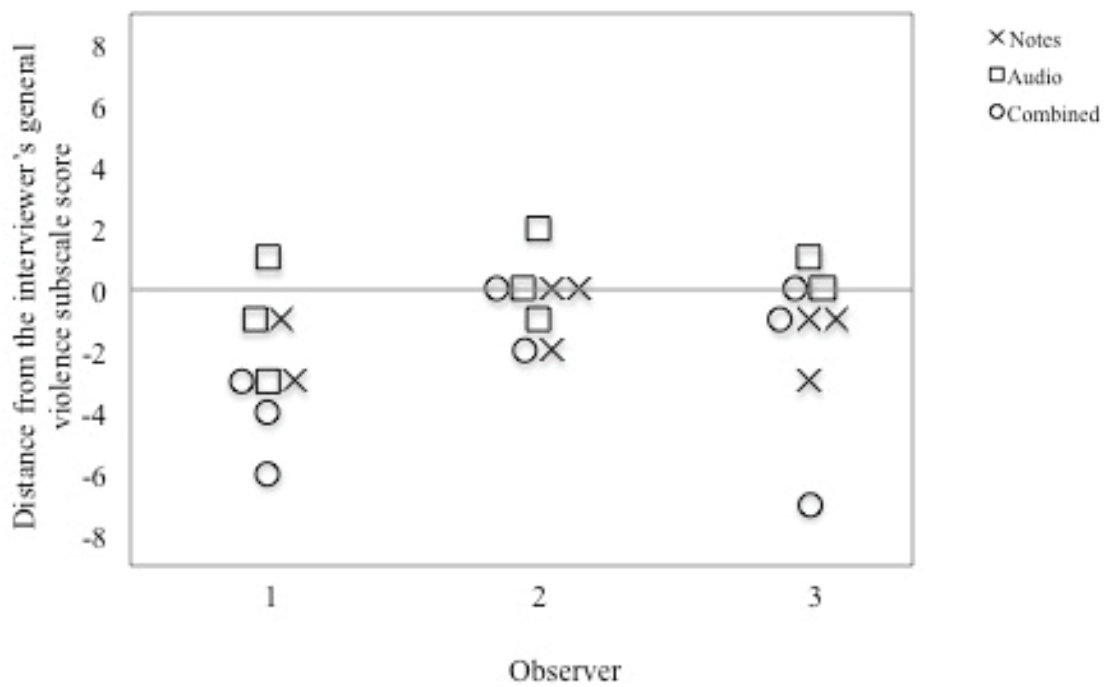


Figure 5. Distribution of the distance from the interviewer's general violence subscale score for the three interrater reliability methods.

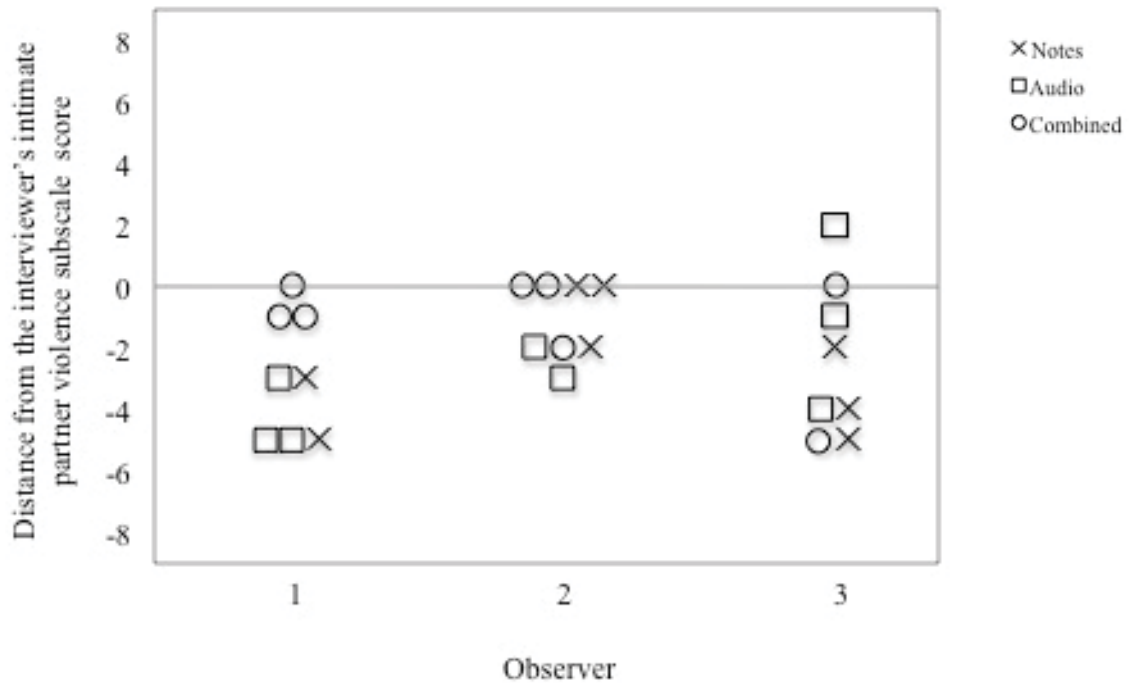


Figure 6. Distribution of the distance from the interviewer's intimate partner violence subscale score for the three interrater reliability methods.

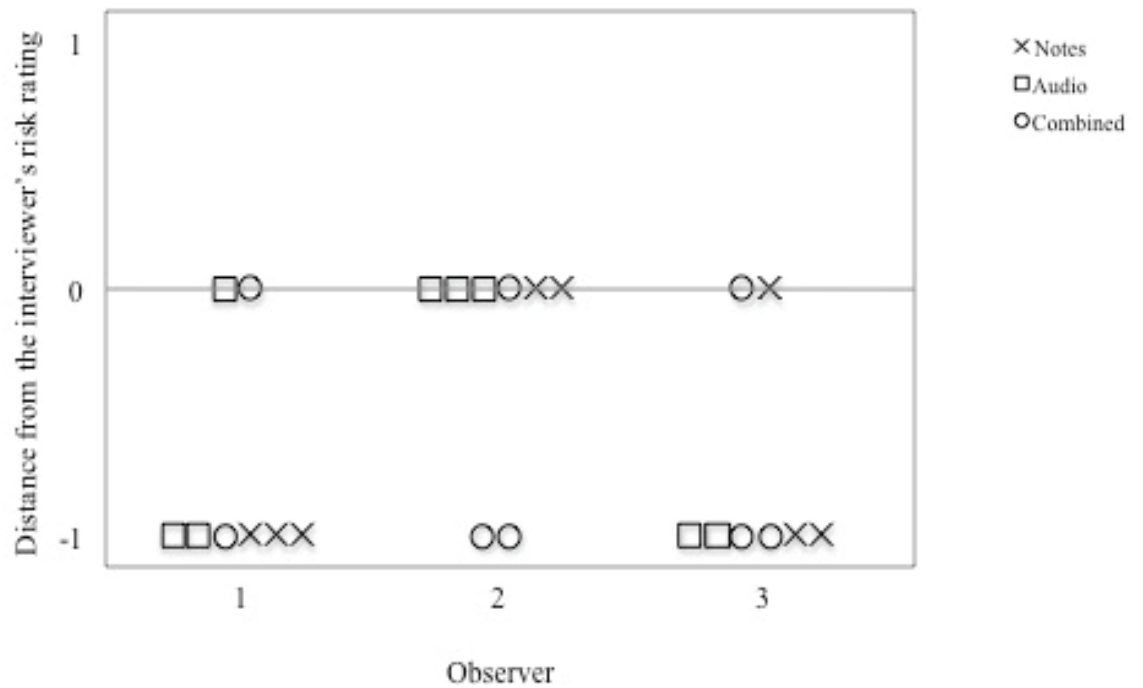


Figure 7. Distribution of the distance from the interviewer's summary risk rating for the three interrater reliability methods.

Second, three one-way analyses of variance were conducted to explore the impact of rater bias (the mean discrepancy between the interviewer and the observers) on the observer's total score, and general violence and intimate partner violence subscale scores (see Table 7). One Friedman test, the non-parametric alternative, was computed for the summary risk rating (see Table 8). The effect size thresholds for all comparative analyses were .10 for a small effect, .30 for a moderate effect, and .50 for a large effect (Cohen, 1988, 1992). There was a statistically significant difference in SARA total score mean discrepancy scores for the three observers, $F(2, 21) = 3.64, p < .05$. The effect size, calculated using eta squared, was small. Post-hoc comparisons using Tukey HSD test indicated that the mean discrepancy score for observer one ($M = -5.38, SD = 3.93$) was significantly different from observer two ($M = -1.50, SD = 2.00$). The mean discrepancy score for observer three did not significantly differ from either observer one or observer two. No significant differences were found in SARA general violence and intimate partner violence subscales mean discrepancy scores for the three observers.

Table 7

Differences in the mean discrepancies of observer 1, observer 2, and observer 3, for the SARA total and subscale scores.

SARA	Observer 1		Observer 2		Observer 3		<i>F</i>	<i>p</i>	<i>r</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Total	-5.38	3.93	-1.50	2.00	-4.13	2.53	3.64	.05	.26
GV	-2.38	2.13	-0.38	1.30	-1.50	2.51	1.93	.17	.15
IPV	-3.00	2.07	-1.13	1.25	-2.25	2.49	1.77	.19	.14

Note: SARA = Spousal Assault Risk Assessment guide

Total = Total score

GV = General Violence subscale

IPV = Intimate Partner Violence subscale

The results of the Friedman test indicated that there was a statistically significant difference in summary risk ratings across the three observers, $\chi^2(2, n = 8) = 13.87, p < .01$. Inspection of the median values showed that observer two had less discrepancy with the interviewer than both observer one and observer three. Interestingly, observer

one had the most experience administering the SARA, indicating that previous experience may not have increased agreement with the interviewer.

Table 8

Differences in the mean discrepancies of observer 1, observer 2, and observer 3, for the SARA summary risk ratings.

SARA	Observer 1 Median	Observer 2 Median	Observer 3 Median	χ^2	p
Summary risk rating	-1.00	0.00	-1.00	13.87	.01

Note: SARA = Spousal Assault Risk Assessment guide

Part Two: Convergent and discriminant validity

Convergent validity

Correlations between the SARA total score, subscale scores, and summary risk rating. First, Pearson product-moment and Spearman rank order correlation coefficients were used to examine the relationships between the SARA (total score, general violence and intimate partner violence subscale scores, and summary risk rating; see Table 9). The magnitudes for the correlations as recommended by Cohen (1988) will be used with small (.10 to .29), medium (.30 to .49), and large (.50 to 1.0). All of the correlations between the SARA scores were significant ($p < .01$). The correlations between the summary risk rating score and the total score, general violence subscale, and intimate partner violence subscale lay in the large range. Large significant correlations were found between the total score and the general violence and intimate partner violence subscales. Correlations between the SARA general violence subscale and the SARA intimate partner violence subscale scores were large and significant. The high inter-correlation between these two subscales and the similar correlations between them and the SARA total score may reflect that they are measuring the same construct or highly related constructs.

Table 9

Bivariate correlations between SARA scores.

SARA	Total score	GV	IPV	Summary risk rating
Total score	-			
GV	.97*	-		
IPV	.96*	.87*	-	
Summary risk rating	.63*	.53*	.71*	-

Note: * $p < .01$

SARA = Spousal Assault Risk Assessment guide

GV = General Violence subscale

IPV = Intimate Partner Violence subscale

Correlations between the SARA, risk for intimate partner violence, and risk for general violence. Table 10 presents Pearson product-moment and Spearman's rank order correlation coefficients between the SARA scores and risk for intimate partner violence (as measured by the Domestic Violence Screening Instrument). All of the correlations were significant ($p < .01$).

Table 10

Bivariate correlations between SARA scores and Domestic Violence Screening Instrument.

SARA	DVSI
Total score	.74*
GV	.70*
IPV	.73*
Summary risk rating	.52*

Note: * $p < .01$

SARA = Spousal Assault Risk Assessment guide

GV = General Violence subscale

IPV = Intimate Partner Violence subscale

DVSI = Domestic Violence Screening Instrument

A large statistically significant correlation was present between the SARA total score and the Domestic Violence Screening Instrument. Significant correlations between the SARA general violence and intimate partner violence subscales and the

Domestic Violence Screening Instrument were large. The correlation between the SARA summary risk rating and the Domestic Violence Screening Instrument indicated a statistically significant large relationship.

Table 11 presents Pearson product-moment correlation coefficients between the SARA general violence subscale, SARA intimate partner violence subscale and risk for general violence (as measured by the Violence Risk Appraisal Guide, and the Violent Offender Treatment Programme-Risk Assessment Scale).

Table 11

Bivariate correlations between SARA subscales, Violence Risk Appraisal Guide and Violent Offender Treatment Programme-Risk Assessment Scale.

	SARA GV	SARA IPV	VRAG	VORAS
SARA GV	-			
SARA IPV	.87*	-		
VRAG	.65*	.53*	-	
VORAS	.57*	.50*	.69*	-

*Note: *p < .01*

SARA GV = Spousal Assault Risk Assessment guide General Violence subscale

SARA IPV = Spousal Assault Risk Assessment guide Intimate Partner Violence subscale

VRAG = Violence Risk Appraisal Guide

VORAS = Violent Offender Treatment Programme-Risk Assessment Scale

Large significant correlations were present between the SARA general violence subscale and the Violence Risk Appraisal Guide, and SARA general violence subscale and the Violent Offender Treatment Programme-Risk Assessment Scale. Correlations between the SARA intimate partner violence subscale and both the Violence Risk Appraisal Guide and Violent Offender Treatment Programme-Risk Assessment Scale were also significant and lay in the large range.

The relationship between the SARA scores and risk for both intimate partner violence and general violence indicated that the SARA assessed the proposed constructs, with stronger correlations between the SARA scores and Domestic Violence Screening Instrument reflecting a closer relationship with risk for intimate partner violence. Compared to the SARA intimate partner violence subscale, the general violence subscale had stronger correlations with risk for general violence indicating that this subscale was more closely aligned with the proposed construct.

Correlations between the SARA and violent conflict resolution behaviours.

Table 12 presents Spearman’s rank order correlation coefficients between the SARA total score, general violence and intimate partner violence subscales and summary risk rating and violent conflict resolution behaviours (as measured by the Revised Conflict Tactics Scale physical assault, psychological aggression, sexual coercion and injury subscale scores). A moderate significant relationship was found between the intimate partner violence subscale and injury subscale. A moderate but non-significant relationship was found between the summary risk ratings and the injury subscale, while the remaining associations with the physical assault, psychological aggression and injury subscales were small but non-significant. The non-significant relationships between the SARA scores and the sexual coercion subscale indicated no relationship between these variables.

Table 12

Bivariate correlations between SARA total score, SARA intimate partner violence subscale, SARA summary risk rating, and Revised Conflict Tactics Scale violent resolution subscales.

SARA	CTS2			
	Physical assault	Psych aggression	Sexual coercion	Injury
Total score	.19 ^a	.23	-.02 ^b	.29
GV	.19 ^a	.24	.02 ^b	.23
IPV	.18 ^a	.17	-.05 ^b	.37*
Summary risk rating	.17 ^a	.23	.04 ^b	.31

Note: * $p < .05$

^a $n = 42$

^b $n = 40$

SARA = Spousal Assault Risk Assessment Guide

GV = General Violence subscale

IPV = Intimate Partner Violence subscale

CTS2 = Revised Conflict Tactics Scale

Psych Aggression = Psychological Aggression

Discriminant validity

Correlations between the SARA and non-violent conflict resolution behaviours. Table 13 presents Pearson product-moment correlations between the SARA total score, general violence subscale and intimate partner violence subscale, and the non-violent conflict resolution behaviours (as measured by the Revised Conflict Tactics Scale physical negotiation subscale). All of these correlations were non-significant and in the positive direction. In addition, Table 13 presents Spearman's rank order correlations between the SARA summary risk rating and Revised Conflict Tactics Scale negotiation subscale. In the current study no significant relationship was found, although it should also be noted that the current study did not have power to detect an effect size of -0.12, as evidenced by the observed relationship of .15 not being found to be statistically significant. Given the observed effect size was in the opposite direction to that hypothesised, however, a simple lack of power would not explain the results. The findings provide support for the possibility that the SARA is not assessing the non-violent conflict resolution behaviours construct.

Table 13

Bivariate correlations between SARA scores and Revised Conflict Tactics Scale negotiation subscale.

SARA	CTS2 Negotiation
Total score	.02
GV	.03
IPV	.01
Summary risk rating	.15

Note: SARA = Spousal Assault Risk Assessment Guide
GV = General Violence subscale
IPV = Intimate Partner Violence subscale
CTS2 = Revised Conflict Tactics Scale

Part Three: Predictive and incremental validity

The predictive validity of the SARA total score and summary risk ratings was explored in relation to two outcome criteria: any police contact for an intimate partner violence offence during the follow-up period (270 days), and number of days to police contact for an intimate partner violence offence. In addition, the incremental validity of

the SARA dynamic risk factors to improve the predictive validity of the SARA static risk factors was explored.

Preliminary analyses

Assumption checks. The follow-up subsample SARA data (total score, summary risk rating, static factors, and dynamic factors; $n = 36$) were inspected to ensure no violations of the statistical assumptions of normality, homogeneity of variance, univariate and multivariate outliers, linearity, and absence of multicollinearity. Normality was examined using the Kolmogorov-Smirnov statistic and normal probability plots. The dynamic factors had a normal distribution (Kolmogorov-Smirnov = .13). The total score (Kolmogorov-Smirnov = .16), summary risk rating (Kolmogorov-Smirnov = .31), and static factors (Kolmogorov-Smirnov = .18) were significantly positively skewed (all $p < .05$). Inspection of the normal probability plots for the total score, summary risk rating, and static factors variables found minimal deviation from normality (see Appendix F). Boxplots revealed two outliers on the SARA total score, two on the static factors, and one on the dynamic factors. The standardised scores of each case were less than 3.29 and therefore the cases and their values were retained (see Tabachnick & Fidell, 2007). No other violations of the assumptions were found.

SARA descriptive statistics. The mean, standard deviation, and range for the SARA total score, summary risk rating, static factors, and dynamic factors for the follow-up subsample are reported in Table 14 (See Appendix G for the distribution figures). The mean for the summary risk rating was in the low risk range. Compared to the total sample, the follow-up sample scores were not significantly different. According to the summary risk ratings, which were assigned by the assessor, 10 cases were categorised as low risk, 21 as moderate risk, and five cases as high risk.

Table 14

Descriptive statistics for the SARA total score, summary risk rating, static factors, and dynamic factors in the follow-up subsample (n = 36).

SARA	Mean	Std. Deviation	Range
Total score	13.78	7.43	5-36
Summary risk rating	1.86	0.64	1-3
Static factors	7.89	4.73	2-22
Dynamic factors	5.89	3.34	0-14

Note: SARA = Spousal Assault Risk Assessment Guide

Recidivism descriptive statistics. Intimate partner violence police contact events were yielded for 12 (33%) of the 36 cases followed up over 270 days (approximately nine months). Four (33%) of the recidivistic participants had intimate partner violence police contact within three months of assessment, three (25%) between three and six months of assessment, and five (42%) between six and nine months of assessment.

Predictive validity

SARA total and summary risk rating scores differences between the police contact and no police contact groups. Prior to evaluating the predictive accuracy of the SARA total score and summary risk rating, analyses were performed to evaluate if the SARA scores were different for the participants who had police contact for an intimate partner violence and those who had no police contact for an intimate partner violence offence in the follow-up period. An independent samples t-test was conducted to compare the SARA total score for the police contact ($m = 18.00$, $sd = 9.31$) and no police contact ($m = 11.67$, $sd = 5.40$) groups from the current sample. There was a significant difference between these groups, $t(34) = -2.18$, $p < .05$). The effect size was small ($\eta^2 = .12$). A Mann Whitney U test found the police contact group (median = 2) also had significantly different SARA summary risk ratings than the no police contact group (median = 2; $U = 223.50$, $z = 3.03$, $p < .01$). The effect size was large ($r = .51$). As expected, the police contact group had significantly higher SARA total scores and summary risk ratings than the no police contact group.

Accuracy of the SARA total score and summary risk rating to predict intimate partner violence police contact. ROC curve analysis was used to provide an indication of the relative accuracy of the total score and summary risk rating in predicting intimate partner violence police contact to 270 days. Areas under the curve can range from 0 (perfect negative prediction) to .50 (chance) to 1.00 (perfect positive prediction). An area under the curve of .56 corresponds to a small effect, .64 corresponds to a moderate effect, and .71 corresponds to a large effect (Rice & Harris, 2005). The ROC curves for the total score and summary risk rating are shown in Figure 8. The area under the ROC curve analyses indicated that both scores significantly predicted police contact. The area under the curve of the summary risk rating was .78, $p < .01$, 95% CI [.61 to .95], while the area under the curve of the total score was .72, $p < .05$, 95% CI [.53 to .90]. Overall, the effect sizes for each ROC curve analysis were in the large range.

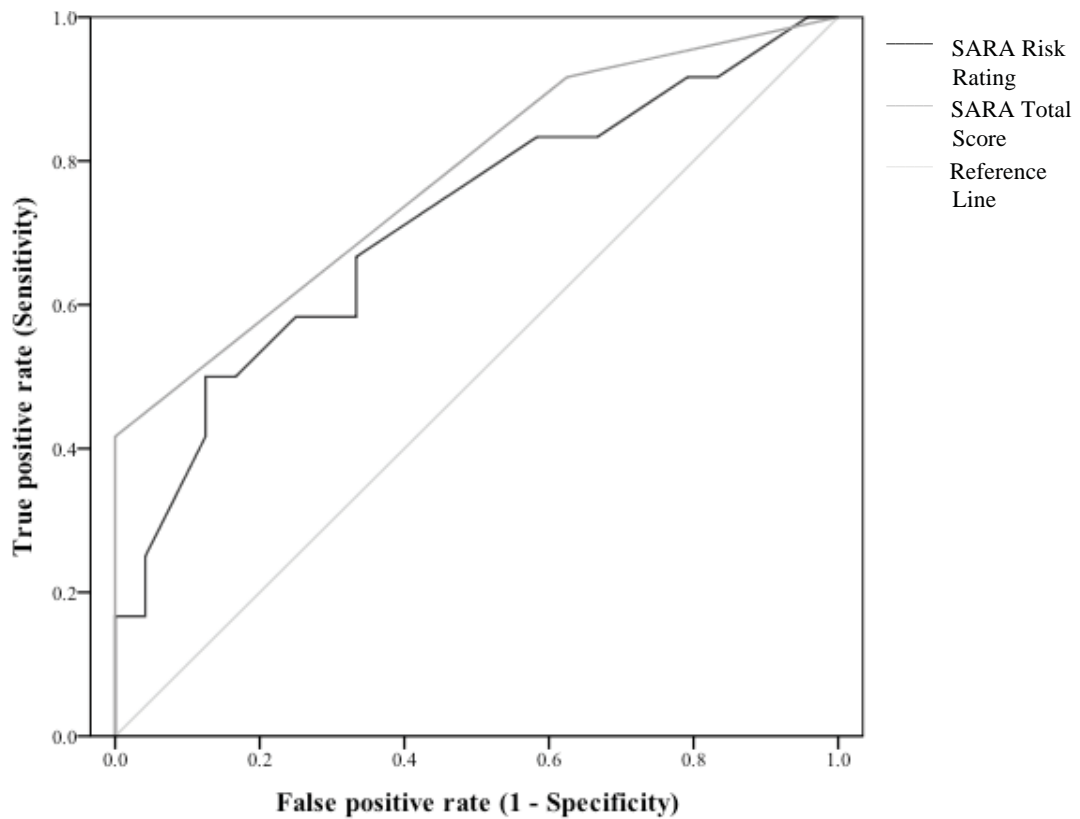


Figure 8. ROC curve of the sensitivity and specificity for the SARA total score and summary risk rating.

Accuracy of the SARA total score and summary risk rating to predict time to police contact. Between-group survival analysis (Kaplan-Meier product-limit model) was used to examine the relationship between police contact over 270 days and group membership (total score and summary risk rating). Group membership for the total score was computed using a cut off score of 13, where cases with scores 12 and below indicated 'low' total score, and cases with scores 13 and above indicated 'high' total score (see Appendix H for true positive and true negative rates at different cut off levels). Kaplan-Meier survival analysis estimate survival function when there are censored cases (cases that remained free of intimate partner violence police contact). The analysis determined the survival function plot directly from the failure times for recidivism. That is, the proportion of cases that were censored, as a function of the time, in days, since their assessment date.

Figure 9 shows the rate of police contact over time for the two total score groups. The top line indicates the rate of police contact for cases categorised as 'low' total score and indicates a low percentage and rate of police contact. In contrast, the bottom line (rate of police contact for high total score cases) accounted for more of the police contact. Log rank test was used to compare time to police contact in the two groups and found that police contact rates between the high total score and low total score groups were not significantly different, $\chi^2(1) = 3.49, p = .06$.

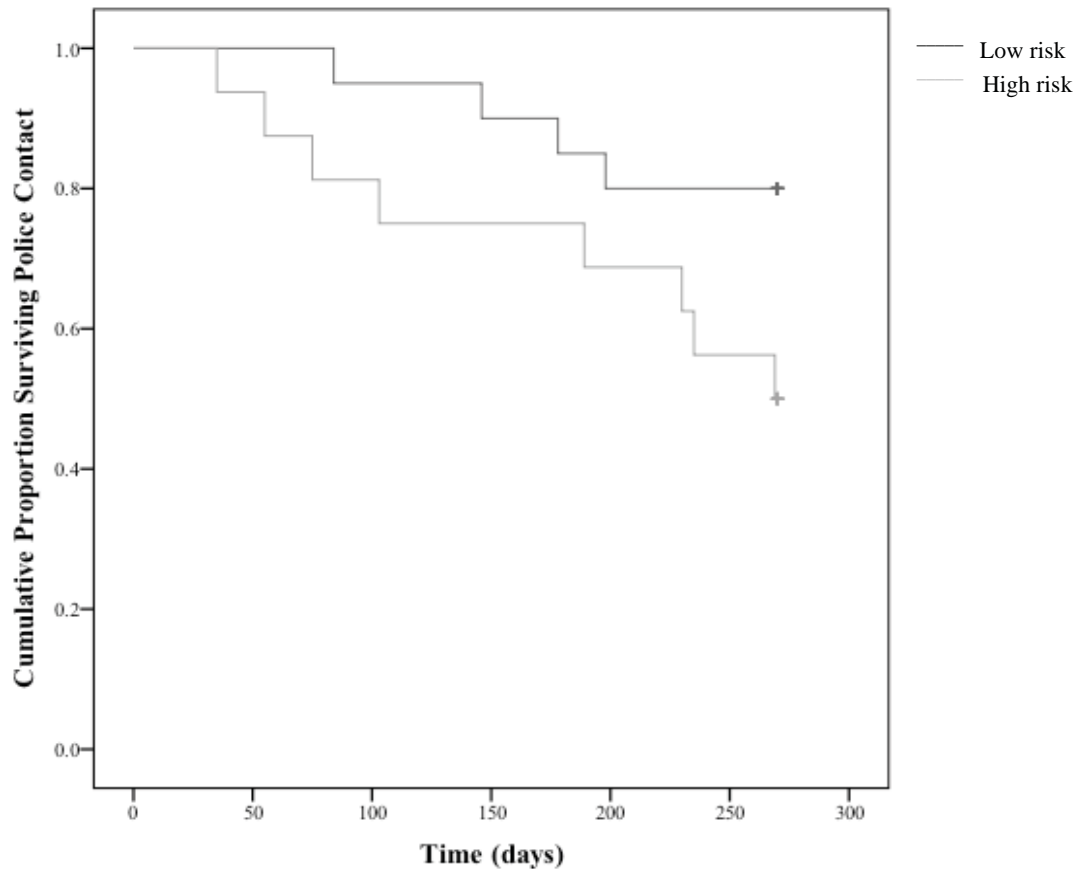


Figure 9. Group comparison of cumulative proportion with no police contact at 270 days, based on total score (cut off = 13).

Figure 10 shows the rate of police contact over time for the three risk rating groups (low, moderate, and high). The top line represents the rate of police contact for cases categorised as low risk and indicates a low percentage and rate of police contact, with no contacts for the first 200 days. The middle line represents the rate of police contact for moderate risk cases and indicates a relatively stable percentage of failure over the first 200 days, with no failure (police contact) after this point. Finally, the bottom line represents the rate of police contact for high risk cases, and accounted for the majority of police contact. Log rank tests found significant differences between police contact of the high risk group and both the low risk group, $\chi^2(1) = 12.71, p < .01$, and moderate risk group, $\chi^2(1) = 7.94, p < .01$. No significant difference was found between the low risk and moderate risk groups, $\chi^2(1) = 1.36, p = .24$.

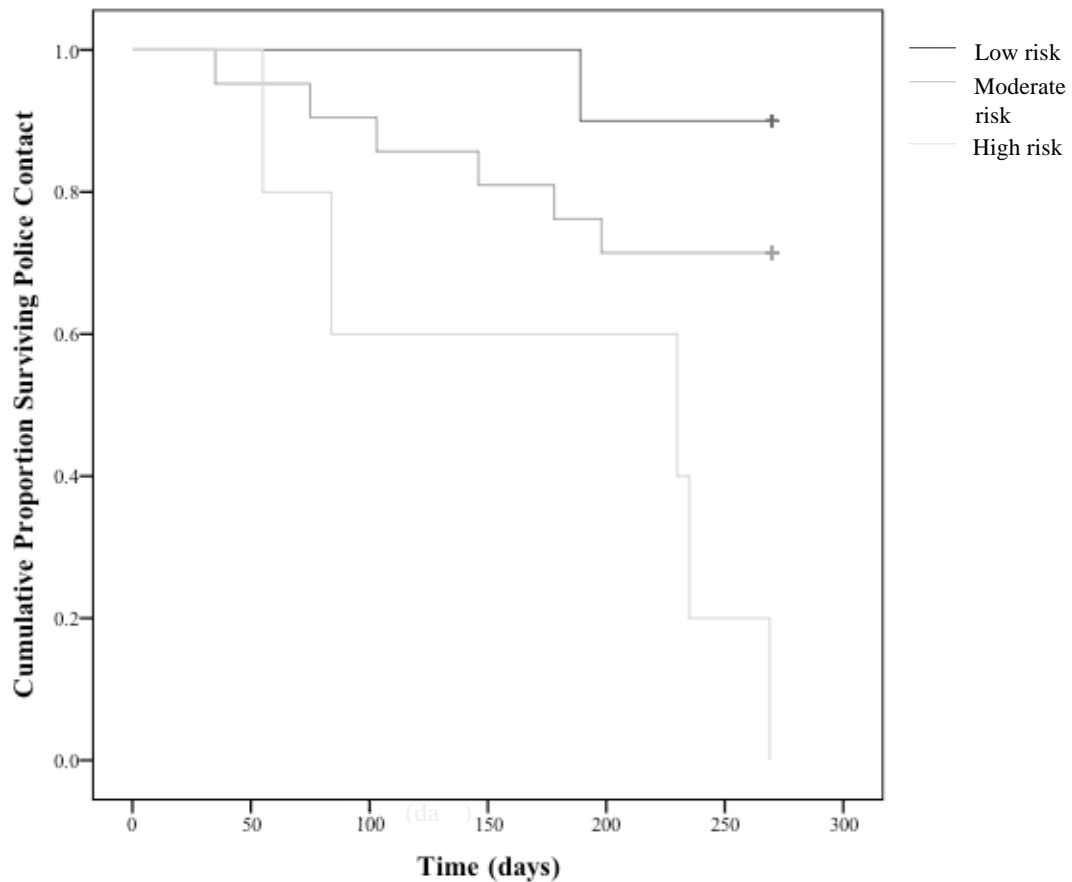


Figure 10. Group comparison of cumulative proportion with no police contact at 270 days, based on summary risk rating.

Incremental validity

Two exact logistic regressions were performed on police contact at 270 days as the outcome. The first run had one predictor, static risk factor scores, while the second run had two predictors, static risk factor and dynamic risk factor scores. The difference between the two models was then evaluated using the goodness-of-fit χ^2 process.

In the first run, a test of the full model against a constant-only model was statistically significant, conditional exact test score = 6.87, $p < .05$. Scores on the static risk factors distinguished between police contact and no police contact. In the second run, a test of the full model for the effect of the dynamic risk factor scores, after adjusting for the static risk factor scores, was not different from a constant-only model, conditional exact test score = .05, $p = .88$. Scores on the dynamic risk factors did not distinguish between police contact and no police contact.

Unconditional logistic regression predicting police contact at 270 days from the static risk factor scores gave a loglikelihood statistic of -19.45. Adding the dynamic risk

factor scores into the model changed the overall fit to -19.48. By the usual approximation and under the assumption that the dynamic risk factors do not add to the model, the difference in the model fit was .06, $p = .80$. The dynamic risk factor scores did not have a significant effect on the overall fit of the model. While this result is an approximation, it underlies the results attained above. The p value in the exact logistic regression is not significant and the dynamic risk factor scores did not add to the model by adjusting the static risk factor scores variable.

CHAPTER SEVEN

DISCUSSION

The aim of the current study was to evaluate the reliability and validity of the SARA with a New Zealand sample. Currently no intimate partner violence risk assessments have had the psychometric properties evaluated with this unique population. Specifically, the interrater reliability, and convergent, discriminant, and predictive validities of the SARA, and incremental validity of the SARA dynamic factors were investigated. Further to this, the current study investigated whether the interrater reliability coefficients improved from that attained with interview notes alone, with audio recordings of the interview or multiple sources of information. The present chapter begins with a discussion of the results of the study in the context of the literature reviewed. This is followed by a review of the implications of the findings for intimate partner violence risk assessment and interrater reliability studies. The chapter concludes with a discussion of the limitations of the study and recommendations for future directions of intimate partner violence risk assessment research in New Zealand.

Summary of the Findings

Part One: Interrater reliability

Consensus estimates of the interrater reliability of the SARA total, general violence and intimate partner violence subscales, and summary risk rating were carried out. The interviewer's (in-person) ratings were compared to the observers' (independent raters) ratings, each of who scored the SARA using three sources of information (interview notes alone, audio recording alone, and combined audio recording and interview notes). The hypothesis that the SARA would have adequate interrater agreement was partially supported. For the total and subscale scores the interrater reliability was in the slight to fair ranges in the interview notes alone method, the almost perfect range in the audio recording alone method, and the slight to substantial ranges in the combined interview notes and audio recording method. The interrater reliability for the summary risk rating was slight to fair across the three methods. In comparison, previous research found that the interrater reliability of the SARA scores when the observers had access to file information was between substantial and almost perfect (Grann & Wedin, 2002; Kropp & Hart, 2000), while the interrater reliability of the total

and subscale scores when the observers had access to an audio recording of the interview was almost perfect (all above .90; Mowat-Leger, 2001).

The low agreement in the interview notes alone method likely reflects artefacts that resulted from the quality of the notes provided to the observers. First, these notes were comprised during the interview and the level of detail was at the discretion of the interviewer. Previous studies, which found higher interrater reliability coefficients than the current one, provided the observers with correctional files. These files were more detailed and included summaries of several prior interviews, psychological and psychiatric reports, and the results of previous risk assessments for general violence. Second, the interviewer may have relied on specific idiosyncratic and minor details to score the items and determine the summary risk rating, which likely were not recorded in the notes and, therefore, not available for the observers to utilise. Therefore, the level of detail in written notes provided to the observers appears to have had an impact on low agreement.

In the audio recording alone method, which evidenced the higher interrater reliability coefficients than interview notes alone method, it is likely that less artefacts influenced the coefficients. Prosodic features of communication unique to spoken words were captured in the audio recording alone method. It has been shown that voice tone contributes more of the total meaning of a communication than spoken words (Mehrabian & Ferris, 1967). Other components of prosody that may have impacted include the emotional state of the speaker, form of the statement, emphasis, stress, and the presence of sarcasm. Further to this, in comparison to prosodic cues, visual cues have been shown to further improve the contribution to the total meaning (Argyle, Salter, Nicholson, Williams, & Burgess, 1970; Dohen & Loevenbruck, 2009; Mehrabian & Ferris, 1967). Visual cues contain both prosodic and other non-verbal behaviours, including displayed facial affect and posture. Therefore, it is likely that the interrater reliability of the SARA scores would have been further improved if the observers had access to a video recording of the interview.

The inadequate interrater agreement in the SARA summary risk ratings warrants special consideration. Across the methodologies the agreement was in the slight to fair ranges, and lower than that reported by Kropp and Hart (.63 for file review methodology; 2000). That is, the summary risk ratings made by the observers' were not consistent with those of the interviewer. In addition, the mean discrepancy between the ratings of the observers' and the interviewer were significantly different; two of the

observers systematically underrated the summary risk rating, compared to the third observer.

Given that summary risk ratings rely on clinical judgement, the experience of the observers' warrants investigation. Agreement between the interviewer and the observers ranged from 38 to 88 percent. Interestingly, one of the observers with lower agreement (38%) was the PhD student who had administered the SARA in correctional facilities, prior to the commencement of this study. The remaining observers (agreement of 38% and 88%) had similar experience; they were Doctorate of Clinical Psychology students who had previously assessed risk in mental health settings. This suggests that more experience in administering the SARA may not increase agreement with the interviewer. Alternatively, it could suggest that the interviewers' scores were inaccurate. However, when the interviews were externally scored by one of the supervisors, who had clinical and research experience in risk assessment with offender populations (including the SARA), agreement in the summary risk ratings was highest for the interviewer (75%), while agreement between the external observer and the independent observers ranged from 38 to 63 percent. In addition, in terms of accuracy, only one participant who was scored by this group went on to have police contact for an intimate partner violence offence during the follow up period. Two of the observers' (PhD and Doctorate of Clinical Psychology students) scored this person as low risk, while the interviewer, the third observer (Doctorate of Clinical Psychology student), and the external observer scored the participant's risk as moderate.

An alternative explanation for the low interrater reliability in the summary risk rating is that the lack of agreement may have resulted from rater association. Rater association concerns whether the observers have a shared understanding of the trait being assessed (Fleiss, 1973). In the current study the observers may not have had a shared understanding of risk for intimate partner violence, or may have varied in the weight given to the individual risk factors that were considered in assigning the risk rating. Given that the results of the investigation of rater bias showed that the mean difference in the SARA total score and summary risk rating discrepancy scores for the three observers' was significant, the observers may not have shared a common understanding of the construct with the interviewer. That is, the observers understanding of risk for intimate partner violence differed. This lack of agreement may have resulted from the training provided, with risk for intimate partner violence, as a construct, not portrayed to the attendees. In addition, the idiosyncratic factors unique to each observer

may have influenced their understanding of risk. In spite of these findings, and those of the Kropp and Hart (2000) study, a definitive conclusion on the interrater reliability of the SARA summary risk ratings cannot be made; the number of studies conducted to date is too small. Therefore, further studies are required to build on the current literature in order to draw conclusions about the interrater reliability of the summary risk ratings and the factors that result in imperfect agreement.

In order to investigate a possible explanation for the differences in agreement of the previously conducted studies (Grann & Wedin, 2002; Kropp & Hart, 2000; Mowat-Leger, 2001), the impact on agreement of the type of information provided to the observers was explored. It was hypothesised that there would be an incremental increase in interrater reliability depending on the method employed. Specifically, that the combined audio recording and interview method would evidence higher reliability than the audio recording method alone which, in turn, would evidence higher reliability than the interview notes alone method. The hypothesis was partially supported. Agreement in the total and subscales was highest in the audio recording alone method. For the summary risk ratings agreement was in the slight to fair ranges in the three methods. Statistically, the interrater reliability of the general violence subscale increased significantly from the interview notes alone method to the audio recording alone method.

The lack of findings may represent a Type II error; there was a failure to reject the null hypothesis when it was false. That is, there may have been significant differences between the methodologies, which this study did not find. The sample size was small, with eight ratings in each of the three methods. Therefore, the chance of detecting even a large population effect size was less than 80 percent. In addition, the lack of incremental increase between the audio recording alone and combined interview notes and audio recording methods may have resulted from two sources of information being available. The observers may have had difficulty synthesising the two sources of information. Instead, they may have attended to one of the sources of information more than the other. Given the results it is likely that this was the interview notes. In sum, however, the current study found adequate interrater reliability for the total and subscale scores, while the possible explanations for the interrater reliability of the summary risk rating requires further investigation.

Part Two: Convergent and discriminant validity

Convergent validity. To assess the construct validity of the SARA the association between the measure and other parts of the SARA, risk for intimate partner violence, risk for general violence, and violent and non-violent conflict resolution behaviour was explored. The relationships between the SARA scores showed that the SARA total score and summary risk rating were strongly related. Generally speaking, the more risk factors present, the higher the risk assigned to the participant. In terms of the summary risk rating, the relationship with the intimate partner violence subscale was greater than the relationship with the general violence subscale. That the summary risk rating was more closely aligned with risk for intimate partner violence than risk for general violence may indicate that the assessors gave more weight to the risk for intimate partner violence construct when the overall risk decision was made.

The hypothesis that the relationship between SARA scores and risk for intimate partner violence would be strong and positive was supported. Compared to previous literature, the relationship between the total score and the Domestic Violence Screening Instrument was stronger than that reported by Williams and Houghton (2004), and Wong and Hisashima (2008), while the relationship between the summary risk rating and the Domestic Violence Screening Instrument was similar to that reported by Williams and Houghton (2004). The hypothesis that the relationship between the SARA general violence subscale and risk for general violence would be strong and positive was supported. Given previous validation studies (Grann & Wedin, 2002; Kropp & Hart, 2000; Mowat-Leger, 2001) the strength of this relationship was as expected. In line with previous literature (Mowat-Leger, 2001), it was hypothesised that the relationship between the SARA intimate partner violence subscale and risk for general violence would be moderate and positive, reflecting both the shared construct of risk assessment, but also the uniqueness of the risk for intimate partner violence construct. This hypothesis was not supported—the intimate partner violence subscale was strongly related to risk for general violence.

The general trend in the relationship between the SARA subscale scores and risk for general violence that has been demonstrated by several authors was confirmed in the current study (Grann & Wedin, 2002; Kropp & Hart, 2000; Mowat-Leger, 2001). That is, the relationship between the general violence subscale and risk for general violence was stronger than the relationship between the intimate partner violence subscale and risk for general violence. While this finding would suggest that the general violence

subscale is assessing risk for general violence, both of the SARA subscales had stronger relationships with risk for intimate partner violence, compared to risk for general violence. Therefore, it appears that both of the SARA scores are measuring risk for intimate partner violence and, to a lesser extent, risk for general violence. Given the high inter-correlation between the subscales this was not unexpected, they appear to assess the same construct. This is contrary to the SARA authors' assertion that one subscale assesses risk for general violence, while the other assesses risk for intimate partner violence (Kropp et al., 1999). The similarity in the associations between the SARA and both risk for general violence and risk for intimate partner violence may also indicate that the sample consisted of men who engaged in both general and intimate partner violence (that is dysphoric/borderline and generally violent/antisocial types), rather than men who engaged in intimate partner violence alone (family only and low level antisocial types, see p .11 for an outline of the typologies; Clements et al., 2002; Holtzworth-Munroe et al., 2000).

The association between the SARA scores and behaviours associated with intimate partner violence was also explored. Given that the literature has repeatedly shown past behaviour is a good predictor of future behaviour (Campbell et al., 2003, Dutton & Kropp, 2000; Hilton & Harris, 2005; Saunders & Browne, 2000), it was hypothesised that the SARA scores would be moderately and positively associated with the physical assault, psychological aggression, sexual coercion, and injury subscales of the Revised Conflict Tactics Scale. This hypothesis was partially supported; the relationship between the SARA intimate partner violence subscale and Revised Conflict Tactics Scale injury subscale was of moderate strength and significant. The remaining relationships were not significant.

In relation to the previous literature, the relationship between the SARA total score and the physical aggression and psychological abuse subscales of the Revised Conflict Tactics Scale were smaller than those reported by Au et al. (2008) and Mowat-Leger (2001). However, those studies differed from the current study in that Au et al. (2008) used a brief version of the SARA designed for use by police officers, and Mowat-Leger used an alternative measure of past violence behaviour (the Abusive Behaviours Inventory). The relationship between the SARA total and the sexual coercion subscale was similar to that found by Au et al. (2008), who attributed the lack of correlation to the under-reporting of sexual assault. Therefore, with the exception of the intimate partner violence subscale assessing injuries caused in previous intimate

partner violence, the SARA total, general violence and intimate partner violence subscales, and summary risk rating appeared to show discriminant validity with self-reported past intimate partner violence. The construct of the SARA is likely to be broader than that of past behaviour, and more specifically related to risk.

Discriminant validity. It was hypothesised that the SARA scores would have small and negative relationships with non-violent behaviours used in conflict resolution. As with past research (Au et al., 2008), the discriminant validity of SARA total, general violence and intimate partner violence subscales, and summary risk rating was supported. The correlations with the negotiation subscale of the Revised Conflict Tactics Scale were near zero. However, unlike the study by Au et al. (2008), the relationships in the current study were positive. Nevertheless, these findings suggest that the SARA was not assessing the non-violent behaviours used in this conflict resolution construct.

Part Three: Predictive and incremental validity

Predictive validity. The ability of the SARA total score and summary risk rating to predict future intimate partner violence was assessed in relation to two prospective criteria; police contact for intimate partner violence and time to police contact for intimate partner violence. Initial analyses showed that the SARA total and subscale scores were strongly associated with police contact in the follow up period. These relationships were stronger than the moderate correlations reported by Gibas et al. (2008), who used correctional records as their outcome criterion. The SARA summary risk rating was also moderately associated with time to police contact in the follow up period. In addition, initial analyses of mean differences in the scores for each of the outcome groups showed that the participants who had police contact had significantly higher total scores and summary risk ratings than the participants who had no police contact.

The hypothesis that participants with higher scores on the SARA would have police contact during the 270 day follow-up period, was supported for the SARA summary risk rating and partially supported for the SARA total score. The participants assigned high risk status, according to the summary risk rating, were found to be more likely to have come into contact with the police for further intimate partner violence than those assigned low or moderate risk status. Furthermore, ROC curve analyses indicated that the summary risk rating had a large predictive accuracy. The predictive

accuracy of the summary risk ratings was larger than the moderate predictive accuracy reported by others who also employed community samples (Kropp & Hart, 2000; Reeves et al., 2008). For the total score, the participants in the high risk group (SARA total scores 13 and above) were no more likely to have come into contact with the police for further intimate partner violence than those in the low risk group (SARA total scores 12 and below). However, the total score had a large predictive accuracy for further police contact, which was larger than the small to moderate predictive accuracies reported elsewhere (Heckert & Gondolf, 2004; Kropp & Hart, 2000). Both of these latter studies (Heckert & Gondolf, 2004; Kropp & Hart, 2000) utilised community samples, however the SARA was scored from file information. The difference in the findings of these two analyses may reflect the base rates of the police contact; unlike tests of the significance of associations, ROC curves are not dependent on the base rate of the recidivistic behaviour.

Despite the small sample, SARA scores accurately predicted police contact for intimate partner violence. That is, the trade off between sensitivity and specificity was acceptable, and any errors present did not detrimentally impact on the confirmation of high risk cases as recidivistic and low and moderate risk cases as non-recidivistic. In the current study all of the participants (100%) assigned high SARA summary risk ratings had further police contact in the follow up period, while 10 percent of the participants assigned low risk ratings and 29 percent of the participants assigned moderate risk ratings had further police contact. Therefore, if appropriate risk management strategies had been implemented following assessment, the high accuracy of the SARA summary risk ratings suggest that the likelihood of either potential victim's experiencing undue harm or perpetrators experiencing an unjust loss of liberty through misclassification would have been reduced. In addition, the false positive rate in the current study was zero, indicating that none of the participants who would not go on to re-offend would experience a loss of liberty through unnecessary risk management strategies. However, because the base rate of further police contact (33% in this study) was below the optimal research level of 50 percent, the positive predictive power was reduced (Streiner, 2003). That is, because there were few true positives in the sample, the proportion of false positives (participants misclassified as low or moderate risk, rather than high) increased.

In addition, it is worth noting that the small sample size of the current study suggests low numbers of false negatives ($n = 7$), compared to true positives ($n = 24$)

when intervention is aimed at those deemed high risk. However, this may under represent the true scale of intimate partner violence⁴. Given that the estimated number of incidents of police contact for intimate partner violence in New Zealand each year is 56,380 (New Zealand Family Violence Clearinghouse, 2007), this false negative rate equates to approximately 10,963 of people in the population classified as low or moderate risk having further police contact. That is, there would be 10,963 incidents where potential victim's experienced undue harm.

Risk assessments also require the imminence of intimate partner violence to be assessed. To date one study has been conducted investigating the accuracy of the SARA to predict imminence, which did not report results for the summary risk rating. Given that the authors argued that the overall risk rating judgement, rather than the total score, reflects the level of risk, this highlights a void in the existing literature (Kropp et al., 1999). In light of this, it was hypothesised that participants with higher scores on the SARA would have shorter time to police contact during the 270 day follow-up period. For the SARA summary risk ratings the hypothesis was supported. The participants assigned high summary risk ratings had a shorter time to police contact than those assigned both low and moderate summary risk ratings, while no difference was found in the time to police contact for the participants assigned low and moderate summary risk ratings. For the SARA total score, the hypothesis was not supported. The participants in the high risk group did not have a shorter time to police contact, compared to the participants in the low risk group. However, this finding was not unexpected for two reasons. First, the correlation between the SARA total score and time to police contact for an intimate partner violence offence in the follow up period was not significant. Second, the non-significant findings in the total score ROC and Kaplan-Meier survival analyses suggest that the cut off score may not have accurately differentiated the low, moderate, and high risk groups.

Overall, the results of the predictive validity studies for the SARA summary risk rating were promising. The SARA summary risk ratings of the interviewer discriminated between the participants who did and did not have police contact, and accurately predicted both police contact for intimate partner violence, and time to police contact for intimate partner violence. However, they must also be interpreted with

⁴ Also, this finding may reflect a sampling artefact because the sample was not random or normal. Within a small sample and with a short follow-up time the unexpected could be expected.

caution. The low interrater reliability coefficients suggested that multiple observers could not agree on the assignment of risk ratings for the participants. As a result, it could be argued that the current study determined the predictive validity of the interviewers' risk rating assignment, rather than the predictive validity of the SARA summary risk ratings. The agreement between observers' needs to be improved before there can be certainty in the scores used to determine the predictive validity.

Incremental validity. The incremental validity of the SARA was explored to determine if the dynamic risk factors enhanced the predictive validity of the static risk factors. It was hypothesised that dynamic risk factors would not provide better police contact classification than the static factors. As expected, the results suggested that the dynamic risk factors do not contribute uniquely to the predictive accuracy of the static risk factors. Therefore, this study suggests that the dynamic risk factors did not evidence incremental validity. However, a longer follow-up period may have allowed for the dynamic risk factors to come into play and for their incremental validity to be fully assessed.

While traditionally this would indicate that the dynamic risk factors should be removed from the SARA, their inclusion in the measure may allow for risk prevention to be facilitated through risk management strategies, rather than solely risk prediction. Belfrage et al. (2011) found that the number of risk management strategies utilised mediated the relationship between SARA total score and intimate partner violence recidivism, indicating that targeting identified dynamic risk factors may facilitate the prevention of further violence. However, the authors did not investigate the effect of specific strategies. Further studies are required to determine if the SARA dynamic risk factors can be systematically altered through management strategies, and whether this results in violence being prevented. If it is shown that the dynamic risk factors facilitate the prevention of violence then their utility will be confirmed. However, if it is shown that they do not facilitate the prevention of violence then the results of this study bring into question their inclusion in the measure.

Implications

Intimate partner violence risk assessment

The current findings have implications for risk assessment practice in New Zealand. While the New Zealand police currently use a measure to ascertain the risk of homicide and serious harm, this measure has not been shown to be reliable and valid. In

addition, the proposed underlying construct is risk for lethality, as opposed to the more broadly defined construct risk for intimate partner violence that the SARA scores assess. It has been argued that any preventative approaches should attempt to reduce or eliminate intimate partner violence generally, rather than focusing on intimate partner homicide solely (Martin & Pritchard, 2010). Therefore, a void exists in current risk assessment practice in New Zealand, which means that the highly prevalent intimate partner violence (Fanslow & Robinson, 2004; Morris et al., 2003) is not being predicted and a reduction in recidivism facilitated by psychometrically valid measures.

It is proposed that the SARA is one such measure that could fill this void. The results of this study indicated that the SARA accurately predicts police contact for intimate partner violence offences. The SARA maximises both the sensitivity and specificity, meaning that the rate of false positives and false negatives is acceptable. Therefore, the SARA balances the safety of the community with the rights of the person being assessed (Douglas & Kropp, 2002). Second, higher SARA summary risk ratings are associated with shorter time to police contact for intimate partner violence offences. While the current study did not show that the dynamic risk factors contributed to risk for intimate partner violence over and above the static risk factors, preliminary research conducted by others (Belfrage et al., 2011) shows that targeting the dynamic risk factors through management strategies and interventions may reduce an individual's risk. Compared to actuarial measures, the SARA potentially provides more utility. No one is at risk for intimate partner violence at all times and in all settings—risk is not a static entity (Gendreau & Goggin, 1996). The SARA allows for the change in risk to be assessed, and may allow for risk to be systematically reduced.

The majority of validation studies undertaken with the SARA have recruited participants from Northern America, while the remaining two recruited participants from the Northern Hemisphere, specifically Sweden (Grann & Wedin, 2002) and Spain (Andrés-Pueyo et al., 2008). Studies have shown the predictive validity of risk assessment measures shrink when used with populations that differ from those used in validation studies. In New Zealand the ethnic make up of intimate partner violent offenders differs from those in the Northern Hemisphere. Of the participants followed up in the current study 17 percent were New Zealand European, and 33 percent New Zealand Māori. In comparison, approximately 80 percent of the original SARA validation study sample comprised of Canadian Europeans (Kroop & Hart, 2000). The results of the current study indicate that the SARA total and summary risk rating scores

were similar, albeit at the higher end, to the existing validation studies (see pg. 50 for the list of studies and a summary of the results), suggesting that shrinkage did not impact on the predictive validity. That is, the SARA evidenced predictive validity within this unique population. In addition, the results of the current study suggest that the SARA could be utilised with other non-European populations. That shrinkage did not impact on the current findings could have resulted from the SARA is a third generation risk assessment measure. The risk ratings are not derived from fixed algorithms developed with international samples. In addition, the studies that did evidence shrinkage utilised sexual offender populations (Allan, Dawson, & Allan, 2006; Helmus, Babchishin, & Blais, 2012; Langstrom, 2004; Ward & Dockerill, 1999). It may be that differences exist in the predictability of sexual and intimate partner violence risk, due in part to differences in the base rate of further behaviour.

The current convergent validity results suggest that risk for intimate partner violence is a separate construct to risk for general violence, although there does appear to be some shared variance in the constructs. This indicates that measures unique to risk for intimate partner violence are required for use with this special population. The differences between the forms of violence may result from the typologies of intimate partner violence offenders. That is, while typologies exist where violence is directed towards both intimate partners and non-intimate partners, there exists a type where the perpetrators engage in violence directed solely towards an intimate partner. The accuracy of the SARA with each type of intimate partner violence perpetrator may help to further understand the utility of unique risk assessments within the intimate partner violence population. Perhaps both general violence and intimate partner violence risk assessments would be useful with the types that engage in both general violence and intimate partner violence (i.e. the dysphoric/borderline and generally violent/antisocial types), while intimate partner violence assessments would be useful with the types that engage solely in intimate partner violence (i.e. the family only and low level antisocial types; Clements et al., 2002; Holtzworth-Munroe et al., 2000).

However, the promising validity findings should be viewed in light of the interrater reliability findings. While the agreement in the total and subscale scores was acceptable, the low agreement of the summary risk ratings brings into question the reliability of the risk assigned. The observers agreed on which risk factors were present, which suggested that there was consensus in the specific factors that should be targeted in order to reduce risk. However, given that all of the participants assigned high risk

state had police contact for intimate partner violence in the 270 day follow up period, misclassifying these people as moderate risk, and implementing lesser case management strategies accordingly, could have implications for the safety of the partner. That is, it could put other people at undue risk. In addition, misclassifying people who would not go on to re-offend as high risk could lead to a loss of liberties as risk management strategies are implemented. Further research is required to determine methods that increase agreement in the risk ratings. These may include standardising training programmes, which have been shown to improve interrater reliability, or utilising video recording in future validation studies.

Interrater reliability studies

The results suggest that the methodology employed in studies investigating interrater reliability may impact on the results attained. In studies where the interviewer and observers are provided with different sources of information to make their ratings, the interrater reliability coefficients attained may be lower than true coefficients because additional measurement error is introduced. This appears especially true when the observers are provided with interview or file notes and the interviewer interviewed the participant. The prosodic elements of communication are not being captured in written communication. When the two methods are more closely matched (i.e. audio recording and interviewing) the coefficients are higher. The coefficients also improve in studies that use only one method, for example when file notes are provided to both the interviewer and observers (Grann & Wedin, 2002). The results also suggest that the addition of multiple sources of information does not improve the interrater reliability coefficients above that of single sources of information. This finding is important because of the cost to researchers' of complying multiple sources of information. In sum, future studies should employ one source of information that is at least an audio recording of the interview.

It should be noted that none of these methodologies reflect real world clinical practice. Risk assessments are usually made following an interview with the perpetrator. Therefore, the results of the current study are also likely underestimates of the true interrater reliability of the SARA. It is recommended that future studies attempt to use real world methodologies to determine the consensus amongst raters. That is, both the interviewer and observers' interview the perpetrator in order to assign a risk rating.

Limitations

There are a number of limitations of the current findings, which relate to the sample size, the reliance on self-report participant interviews, the generalisability of the findings, and the impact of the treatment programme on the risk state of the participants. Each of these will be discussed below.

While every effort was made to recruit sufficient participant numbers to undertake the analyses, the small sample size limited the current study. The non-significant findings of the Kaplan-Meier survival analysis, which employed the SARA total score, and the analysis of variance of the divergence in the mean differences for the observers', which employed the subscale scores, should be interpreted with caution. Because of the small sample size it is unclear whether non-significant findings were accurate, or whether they represented a Type II error. That is, there may be genuine differences in the underlying intimate partner violence population, which, with the current sample, this study failed to find. Therefore, replication of the study with an increased sample size of greater demographic variation is strongly recommended for future research.

The second limitation was the reliance on self-report participant interview. Self-report may have affected the scoring of the SARA two ways. First, the participants may have intentionally responded in a socially desirable manner and minimised their violent behaviour and attitudes towards intimate partner violence. In future studies it is recommended that socially desirable responding be assessed so that it can be inferred whether the data is tainted by such tendencies. Second, the participants may have underestimated past behaviour. It has been shown that recall tends to be affected unless behaviour is rare or of significant importance (Schwarz, 1999). Therefore, the use of interview may have resulted in an underestimation of risk. While the authors of the SARA recommend that corroborative information be sought during the assessment process (Kropp et al., 2000) to minimise the impact of relying on self-report, the current study did not seek collateral information because it was not possible to talk to the partners and files (e.g. correction or judicial) were specifically not utilised. Instead, the ecological validity of the study was prioritised; it is likely that files and partner report will not be available to all professionals administering the SARA.

The third limitation was the generalisability of the results. Given that the sample was drawn from a community-based stopping violence programme the results may not generalise to other intimate partner violent populations, such as groups residing in New

Zealand prisons. Community samples, in general, are a lower risk group because the most serious offenders often receive prison sentences. A further generalisability issue relates to the predictive validity criterion. Given that the base rate of self-reported violence is higher than the base rate of police contact for violence (Babcock et al., 2004), the validity of the SARA to predict police contact for intimate partner violence cannot be automatically generalised to self-reported intimate partner violence. Despite this, the criterion selected was also a strength of the study. Given that self-report is influenced by high levels of participants leaving the study during the follow up period, the use of this criterion may have further impacted on the sample size and resulted in lower chance of finding genuine differences.

The final limitation related to the impact of the participants' attendance at the stopping violence programme on the base rate of future intimate partner violence. That is, by attending the programme the participants' risk state may have reduced which, in turn, may have confounded the association between the SARA and the predictive validity outcome criterion. However, performing research with an intimate partner violence sample that was withheld from the stopping violence programme was not possible. It was considered unethical to withhold intervention when the researcher knew that the participants were at risk of engaging in behaviour that harmed another person. Given that the results were favourable (the SARA scores demonstrated predictive accuracy), the stopping violence programme may not have impacted on the results in this manner. This may have been because the sample appeared to consist of participants who engaged in both general and intimate partner violence (that is, generally violent/antisocial and dysphoric/borderline types), which is not the type of intimate partner violence perpetrators targeted by stopping violence programmes (family only and low level antisocial types; Clements et al., 2002).

Future Directions

In addition to the recommendations made throughout this discussion, several other suggestions are made. First, conducting another study using a correctional sample would enable the sensitivity and specificity of the SARA with a higher risk New Zealand population to be assessed. It could be argued that accurate risk assessment and risk management practice is more important in this population, because the victims are more likely to experience severe intimate partner violence. This process would further enhance the clinical utility of the measure, and provide guidelines for its use with New

Zealand intimate partner violence perpetrators, as a whole. Second, the SARA should be validated with female perpetrators of intimate partner violence. Several studies have shown that men and women initiate violent behaviour at equal rates (Langhinrichsen et al., 1995; Moffitt & Caspi, 1999; Morse, 1995). This may also result in a better understanding of the similarities and differences in risk factors for each gender. Third, as an extension of the validation of the SARA with a New Zealand population, the utility of each of the items in predicting and facilitating the prevention of violence should be examined. Differences in several characteristics (for example, ethnicity and culture) likely create a uniqueness, which may result in important difference in how the scale would work in this population. Finally, methods of increasing the interrater reliability of the SARA should be investigated. This could include an exploration of the factors (for example the content of the training, practical experience with intimate partner violence populations, practical experience with risk assessment) that could enhance the understanding of the construct of risk for intimate partner violence. This, in turn, may result in higher agreement between observers.

Conclusion

With the high prevalence of intimate partner violence and the resultant negative consequences experienced by both partners and the children, addressing the risk posed by perpetrators is important. To date no intimate partner violence risk assessment measure has been standardised with a New Zealand population. Despite numerous international studies pertaining to the reliability and validity of several measures, including the SARA, the unique characteristics of New Zealand's population likely differentiates it from the populations of other countries and, as a result, the findings of such studies may not generalise here. That is, it is impossible to say whether the measures adequately balance the *mêlée* between potential victims experiencing undue harm and misclassified perpetrators experiencing an unjust loss of liberty. Based on the current findings, the SARA demonstrated potential to fill this void. It was shown that the construct is specific to intimate partner violence, and that the summary risk ratings predict police contact for intimate partner violence and time to police contact for intimate partner violence. However, these results should be viewed in light of the interrater reliability findings. The observers' could not agree on the summary risk ratings assigned to the participants. In terms of the methodology employed in interrater reliability studies the findings suggested that providing the observers with an audio

recording of the interview produces the most reliable results, compared to interview notes alone or both interview notes and audio recordings.

REFERENCES

- Adams, D. (1988). Treatment models for men who batter: A profeminist approach. In K. Yllö & M. Bograd (Eds.), *Feminist perspective on wife abuse* (pp. 176-199). Newbury Park, CA: Sage.
- Aldarondo, A., & Sugarman, D.B. (1996). Risk marker analysis of the cessation and persistence of wife assault. *Journal of Consulting and Clinical Psychology, 64*(5), 1010-1019.
- Allan, A., Dawson, D., & Allan, M.A. (2006). Prediction of the risk of male sexual reoffending in Australia. *Australian Psychologist, 41*(1), 60-68.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anastasi, A. (1997). Validity: Basic concepts. In A. Anastasi & S. Urbina (Eds.), *Psychological testing* (7th ed.) (pp. 113-139). Upper Saddle River, NJ: Prentice Hall.
- Andrés-Pueyo, A., López, S., and Álvarez, E. (2008). Assessment of the risk of intimate partner violence and the SARA. *Papeles del Psicólogo, 29*(1), 107-122.
- Andrews, D.A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Newark, NJ: Lexis Nexis.
- Andrews, D.A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: Lexis Nexis.
- Argyle, M., Slater, V., Nicholson, H., Williams, M., & Burgess, P. (1970). The communication of inferior and superior attitudes by verbal and non-verbal signs. *British Journal of Social and Clinical Psychology, 9*, 222-231.
- Ateah, C.A., Secco, M.L., & Woodgate, R.L. (2003). The risks and alternative to physical punishment use with children. *Journal of Paediatric Health Care, 17*(3), 126-132.
- Au, A., Cheung, G., Kropp, P.R., Yuk-chung, C., Lam, G.L.T., & Sung, P. (2009). A preliminary validation of the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER) in Hong Kong. *Journal of Family Violence, 23*(8), 727-735.
- Babcock, J.C., Green, C.E., & Robie, C. (2004). Does batterers' treatment work? A metaanalytic review of domestic violence treatment. *Clinical Psychology Review*

23(8), 1023-1053.

- Baird, C. (2009). *A Question of evidence: A critique of risk assessment models used in the justice system: Special report*. National Council on Crime and Delinquency. Retrieved from http://cjjr.georgetown.edu/pdfs/ebp/baird2009_QuestionOfEvidence.pdf.
- Balzer, R., Haimona, D., Henare, M., & Matchitt, V. (1997). *Māori Family Violence in Aotearoa: Hamilton abuse intervention pilot project*. Wellington, New Zealand: Te Puni Kokiri.
- Bandura, A. (1971). *Social learning theory*. New York, NY: General Learning Press.
- Bandura, A. (1973). *Aggression: A social learning analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Bartels, L. (2010). *Indigenous Women's Offending Patterns: A literature review*. Australian Institute of Criminology, Canberra, Australia.
- Bartko, J.J., & Carpenter, W.T. (1976). On the methods and theory of reliability. *Journal of Nervous and Mental Disease*, 163(5), 307-317.
- Bascand, G. (2012). *Household Labour Force Survey: June 2012 quarter*. Wellington, New Zealand: Statistics New Zealand.
- Beasley, R., & Stoltenberg, C.D. (1992). Personality characteristics of male spouse abusers. *Professional Psychology: Research and practice*, 23(4), 310-317.
- Beech, A.R., Fisher, D.D., & Thornton, D. (2003). Risk assessment of sex offenders. *Professional Psychology: Research and practice*, 34(4), 339-352.
- Beck, J. (1995) *Cognitive therapy: Basics and beyond*. New York, NY: Guilford Press.
- Belfrage, H., Strand, S., Storey, J.E., Gibas, A.L., Kropp, P.R., & Hart, S.D. (2011). Assessment and management of risk for intimate partner violence by police officers using the Spousal Assault Risk Assessment guide. *Law and Human Behavior*, Advanced online publication. PMID: 21431454.
- Bell, K.M., & Naugle, A. E. (2008). Intimate partner violence theoretical considerations: Moving towards a contextual framework. *Clinical Psychology Review*, 28(7), 1096 – 1107.
- Bennett Cattaneo, L., & Goodman, L.A. (2005). Risk factors for reabuse in intimate partner violence: A cross-disciplinary critical review. *Trauma, Violence, and Abuse*, 6(2), 141-175.
- Boer, D.P., Hart, S.D., Kropp, P.R., & Webster, C.D. (1997). *Manual for the Sexual Violence Risk-20: Professional guidelines for assessing risk of sexual violence*.

- British Columbia, Canada: Simon Fraser University.
- Bonta, J. (1989). Native inmates: Institutional responses, risk, and needs. *Canadian Journal of Criminology*, 31, 49-62.
- Bonta, J., LaPrairie, C., & Wallace-Capretta, S. (1997). Risk prediction and re-offending: Aboriginal and non-Aboriginal offenders. *Canadian Journal of Criminology*, 39(2), 127-144.
- Borsboom, D., Mellenbergh, G.J., & van Heerden J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071.
- Boyle, D.J., O'Leary, K.D., Rosenbaum, A., & Hassett-Walker, C. (2008). Differentiating between generally and partner-only violent subgroups: Lifetime antisocial behavior, family of origin violence, and impulsivity. *Journal of Family Violence*, 23(1), 47-55.
- Breiding, M.J., Black, M.C., & Ryan, G.W. (2008). Prevalence and risk factors of intimate partner violence in eighteen U.S. states/territories, 2005. *American Journal of Preventive Medicine*, 34(2), 112-118.
- Briggs, S.R., & Cheek, J.M. (1986). The role of factor analysis in the development and evaluation of personality tests. *Journal of Personality*, 54(1), 106-148.
- Broadhurst, R.G., & Maller, R.A. (1990). The recidivism of prisoners released for the first time: Reconsidering the effectiveness question. *Australian and New Zealand Journal of Criminology*, 23, 264-281.
- Brown, S.L. (2002). *Dynamic prediction of criminal recidivism: A three wave prospective study (1995 - 2002)* (Unpublished doctoral dissertation). Queen's University, Ontario, Canada.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Campbell, J. (1986). Nursing assessment for risk of homicide with battered women. *Advances in Nursing Science*, 8(4), 36-51.
- Campbell, J.C., Webster, D., Koziol-McLain, J., Block, C., Campbell, D., Curry, M.A., ... Laughon, K. (2003). Risk factors for femicide in abusive relationships: Results from a multi-site case control study. *American Journal of Public Health*, 93(7), 1089-1097.
- Carswell, S. (2006). *Family violence and the pro-arrest policy: A literature review*. Wellington, New Zealand: Ministry of Justice.

- Cattaneo, L.B., & Goodman, L.A. (2005). Risk factors for reabuse in intimate partner violence: A cross-disciplinary critical review. *Trauma, Violence and Abuse*, 6(2), 141–175.
- Cattell, R.B. (1946). *Description and measurement of personality*. New York, NY: World Book Company.
- Clements, K., Holtzworth-Munroe, A., Gondolf, E.W., & Meehan, J.C. (2002). *Testing the Holtzworth-Munroe et al. (2000) batterer typology among court-referred martially violent men*. Paper presented at the Association for the Advancement of Behavior Therapy Annual Meeting, Reno, NV.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Cone, J.D., & Foster, S.L. (2006). *Dissertations and theses from start to finish: Psychology and related fields (2nd ed.)*. Washington, DC: American Psychological Association.
- Correctional Service Canada. (2006). *Annual Report of the Office of the Correctional Investigator 2005-2006*. Canada: Minister of Public Works and Government Services.
- Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cross, J., & Newbold, G. (2010). Presumptive arrest in partner assault: Use of discretion and problems of compliance in the New Zealand Police. *Australian and New Zealand Journal of Criminology*, 43(1), 51-75.
- Curtis, G.C. (1963). Violence breeds violence-perhaps? *American Journal of Psychiatry*, 120, 386-387.
- Delsol, C., & Margolin, G. (2004). The role of family of origin violence in men's marital violence perpetration. *Clinical Psychology Review*, 24(1), 99-122.
- Dempster, R.J., & Hart, S.D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and nonrecidivists. *Sexual Abuse: A journal of research and treatment*, 14(2), 121-138.
- Dobash, R.E., & Dobash, R.P. (1977). Wives: The “appropriate” victims of martial violence. *Victimology*, 2, 426-442.

- Dohen, M., & Loevenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech, 52*(2-3), 177-206.
- Domestic Abuse Intervention Programmes. (2011). *What is the Duluth model?* Retrieved from <http://www.theduluthmodel.org/about/index.html>.
- Donovan, C., Hester, M., Holmers, J., & McCarry, M. (2006). *Comparing domestic abuse in same sex and heterosexual relationships*. Bristol, England: University of Bristol.
- Doone, P. (2000). *Report on Combating and Preventing Māori Crime*. Wellington, New Zealand: Department of the Prime Minister and Cabinet.
- Douglas, K.S., & Kropp, P.R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behaviour, 29*(5), 617-658.
- Douglas, K.S., & Skeem, J.L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11*(3), 347-383.
- Douglas, K.S., Otto, R.K., & Borum, R. (2003). Research methods in clinical forensic psychology. In J.A. Schinka, W.F. Velicer & I.B. Weiner (Eds.), *Handbook of psychology, volume two: Research methods in psychology* (pp. 189-211). New York, NY: John Wiley and Sons.
- Douglas, K.S., Skeem, J.L., & Nicholson, E. (2011). Research methods in violence risk assessment. In B. Rosenfeld & S.D. Penrod (Eds.), *Research methods in forensic psychology* (pp. 325-346). Hoboken, NJ: John Wiley and Sons.
- Douglas, K.S., Yeomans, M., & Boer, D.P. (2005). Comparative validity analysis of multiple measures of violence risk in a general population sample of criminal offenders. *Criminal Justice and Behavior, 32*(5), 479-510.
- Dutton, D.G. (1995). *The domestic assault of women: Psychological and criminal justice perspectives*. Vancouver, Canada: UBC Press.
- Dutton, D.G. (2003). Theoretical approaches to the treatment of intimate partner violence perpetrators: Treatment of assaultiveness. In D.G. Dutton & D.J. Sonkin (Eds.), *Intimate violence: Contemporary treatment innovations* (pp. 7-28). New York, NY: The Haworth Press.
- Dutton, D.G. (2006). *Rethinking domestic violence*. Vancouver, Canada: University of British Columbia.
- Dutton, D.G. (2007). *The abusive personality (2nd ed.)*. New York, NY: Guilford Press.

- Dutton, D.G., & Corvo, K.C. (2007). The Duluth model: A data-impervious paradigm and a flawed strategy. *Aggression and Violent Behaviour, 12*(6), 658-667.
- Dutton, D.G., & Hart, S.D. (1992a). Risk markers for family violence in a federally incarcerated population. *International Journal of Law and Psychiatry, 15*(1), 101-112.
- Dutton, D.G., & Hart, S.D. (1992b). Evidence of long-term, specific effects of childhood abuse and neglect on criminal behaviour in men. *International Journal of Offender Therapy and Comparative Criminology, 36*(2), 129-137.
- Dutton, D.G., & Kropp, P.R. (2000). A review of domestic violence risk instruments. *Trauma, Violence, and Abuse, 1*(2), 171-181.
- Dutton, D.G., Starzomski, A.J., & Ryan, L. (1996). Antecedents of borderline personality organization in wife assaulters. *Journal of Family Violence, 11*(2), 113-132.
- Dvoskin, J.A., & Heilbrun, K. (2001). Risk assessment and release decision-making: Toward resolving the great debate. *Journal of the American Academy of Psychiatry and the Law, 29*(1), 6-10.
- Fanslow J.L. (2005). *Beyond zero tolerance: Key issues and future directions for family violence work in New Zealand*. Wellington, New Zealand: Families Commission.
- Fanslow, J.L., & Robinson, E. (2004). Violence against women in New Zealand: Prevalence and health consequences. *The New Zealand Medical Journal, 117*(1206), 1173-1185.
- Fanslow, J.L., & Robinson, E. (2010). Help-seeking behaviors and reasons for help-seeking reported by a representative sample of women victims of intimate partner violence in New Zealand. *Journal of Interpersonal Violence, 25*(5), 929-951.
- 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.
- Federal Bureau of Investigation. (2002). *Uniform Crime Reports for the United States*. Washington, DC: US Government Printing Office.
- Fergusson, D.M., Horwood, L.J., & Ridder, E. (2005). Partner violence and mental health outcomes in a New Zealand birth cohort. *Journal of Marriage and Family, 67*(5), 1103-1119.

- Field, A.P. (2009). *Discovering statistics using SPSS: And sex, drugs, and rock 'n' roll* (3rd ed.). London, England: Sage.
- Field, C.A., Caetano, R., & Nelson, S. (2004). Alcohol and violence related cognitive risk factors associated with the perpetration of intimate partner violence. *Journal of Family violence, 19*(4), 249-253.
- Fleiss, J.L. (1973). *Statistical methods for rates and proportions*. New York, NY: John Wiley and Sons.
- Ford, G. (1985). *Domestic disputes: A police perspective: An analysis of domestic dispute complaints attended by the police in Hamilton*. Hamilton, New Zealand: University of Waikato.
- Foster, S.L., & Cone, J.D. (1995). Validity issues in clinical assessment. *Psychological Assessment, 7*(3), 248-260.
- Ganley, A.L. (1989). Integrating feminist and social learning analyses of aggression. In P.L. Caesar & L.K. Hamberger (Eds.), *Treating men who batter* (pp. 196-235). New York, NY: Springer Publishing Company.
- Garcia-Moreno, C., Jansen, H.A., Ellsberg, M., Heise, L., & Watts, C.H. (2006). Prevalence of intimate partner violence: Findings from the World Health Organisation multi-country study on women's health and domestic violence. *The Lancet, 368*(9549), 1260-1269.
- Garson, G.D. (2009). Chi-square significance tests. *Statnotes: Topics in multivariate analysis*. Retrieved from <http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm>.
- Gendreau, P., & Goggin, C. (1996). Principles of effective assessment for community corrections. *Federal Probation, 60*(3), 64-71.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34*(4), 575-607.
- Gibas, A., Kropp, P.R., Hart, S.D., & Stewart, L. (2008). *Validity of the SARA in a Canadian sample of incarcerated males*. Paper presented at the 8th Annual Conference of the International Association of Forensic Mental Health Services, Vienna, Austria.
- Glass, N., Laughon, K., Campbell, J., Block, C.R., Hanson, G., Sharps, P.W., & Taliaferro, P.W. (2008). Non-fatal strangulation is an important risk factor for homicide of women. *Journal of Emergency Medicine, 35*(3), 329-335.
- Godbout, N., Dutton, D.G., Lussier, Y., & Sabourin, S. (2009). Early exposure to

- violence, domestic violence, attachment representations, and marital adjustment. *Personal Relationships*, 16(3), 365-384.
- Gondolf, E.W. (1998). *Assessing woman battering in mental health services*. Thousand Oaks, CA: Sage.
- Gondolf, E.W., & White, R.J. (2001). Batterer program participants who repeatedly reassault: Psychopathic tendencies and other disorders. *Journal of Interpersonal Violence*, 16(4), 361-380.
- Grann, M., & Wedin, M. (2002). Risk factors for recidivism among spousal assault and spousal homicide offenders. *Psychology, Crime and Law*, 8(1), 5-23.
- Grant, S.C. (2009). *Family violence risk assessment: An early study of police officers' experience at the frontline* (Unpublished masters dissertation). Victoria University, Wellington, New Zealand.
- Groth-Marnat, G. (2009). *Handbook of psychological assessment (5th ed.)*. Hoboken, NJ: John Wiley and Sons.
- Grove, W.M., & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293-323.
- Haggard-Grann, U. (2007). Assessing violence risk: A review and clinical recommendations. *Journal of Counselling and Development*, 85(3), 294-301.
- Hamberger, L.K., & Hastings, J.E. (1991). Personality correlates of men who batter and nonviolent men: Some continuities and discontinuities. *Journal of Family Violence*, 6(2), 131-147.
- Hamberger, L.K., Lohr, J.M., Bonge, D., & Tloin, D.F. (1996). A large sample empirical typology of male spouse abusers and its relationship to dimensions of abuse. *Violence and Victims*, 11(4), 277-292.
- Hanson, R., Helmus, L., & Bourgon, G. (2007). *The validity of risk assessments for intimate partner violence: A meta-analysis*. Ottawa, Canada: Public Safety Canada.
- Hanson, R.K., & Wallace-Capretta, S. (2000). *Predicting recidivism among male batterers: User Report 2000-06*. Ottawa, Canada: Department of the Solicitor General of Canada.
- Harris, G.T., & Rice, M.E. (2003). Actuarial assessment of risk among sex offenders. *Annals of the New York Academy of Sciences*, 989, 198-210.

- Harris, G.T., Rice, M.E., & Cormier, C.A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior, 26*(4), 377-394.
- Harris, G.T., Rice, M.E., Quinsey, V.L., Lalumiere, M.L., Boer, D., & Lang, C. (2003). A multi-site comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*(3), 413-425.
- Harris, G.T., Rice, M.E., & Quinsey, V.L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behaviour, 20*(4), 315-335.
- Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3*(1), 121-138.
- Hart, S.D. (2001). Assessing and managing violence risk. In K.S. Douglas, C.D. Webster, S.D. Hart, D. Eaves & J.R.P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 13-25). British Columbia, Canada: Simon Fraser University.
- Hart, S.D. (2008). Preventing violence: The role of risk assessment and management. In A.C. Baldry & F.W. Winkel (Eds.), *Intimate partner violence prevention and intervention: The risk assessment and management approach* (pp. 7-18). Hauppauge, NY: Nova Science Publishers.
- Hart, S.D., Kropp, P.R., & Hare, R.D. (1988). Performance of male psychopaths following conditional release from prison. *Journal of Consulting and Clinical Psychology, 56*(2), 227-232.
- Hart, S.D., Kropp, P.R., Laws, D.R., Klaver, J., Logan, C., & Watt, K.A. (2003). *The Risk for Sexual Violence Protocol (RSVP): Structured professional guidelines for assessing risk of sexual violence*. British Columbia, Canada: Simon Fraser University.
- Hart, S.D., Webster, C.D., & Douglas, K.S. (2001). Risk management using the HCR-20: A general overview of historical factors. In K.S. Douglas, C.D. Webster, S.D. Hart, D. Eaves & J.R.P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 27-40). British Columbia, Canada: Simon Fraser University.
- Hartman, S., Janes, S., & Troy, A. (2009). Violence and nursing's response. In K.S. Lundy & S. Janes (Eds.), *Community health nursing: Caring for the public's*

- health* (2nd ed.) (pp. 506-541). Sudbury, MA: Jones and Bartlett Publishers.
- Hastings, J.E., & Hamberger, L.K. (1994). Psychosocial modifiers of psychopathology for domestically violent and nonviolent men. *Psychological Reports*, *74*(1), 112-114.
- Haynes, S.N., & Lench, H.C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment*, *15*(4), 456-466.
- Heckert, D.A., & Gondolf, E.W. (2004). Battered women's perceptions of risk versus risk factors and instruments in predicting repeat reassault. *Journal of Interpersonal Violence*, *19*(7), 778-800.
- Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior*, *21*(4), 347-359.
- Helmus, L., Babchishin, K.M., & Blais, J. (2012). Predictive accuracy of dynamic risk factors for aboriginal and non-aboriginal sex offenders: An exploratory comparison using STABLE-2007. *International Journal of Offender Therapy and Comparative Criminology*, *56*(6), 856-876.
- Helmus, L., & Bourgon, G. (2011). Taking stock of 15 years of research on the Spousal Assault Risk Assessment Guide (SARA): A critical review. *International Journal of Forensic Mental Health*, *10*(1), 64-75.
- Hemphill, J.F., Hare, R.D., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal and Criminological Psychology*, *3*(1), 139-170.
- Hetherington, S. (2009). *Evaluation of a community based programme for male perpetrators of intimate partner violence* (Unpublished masters dissertation). Massey University, Auckland, New Zealand.
- Hilton, N.Z., & Harris, G.T. (2005). Predicting wife assault: A critical review and implications for policy and practice. *Trauma, Violence, and Abuse*, *6*(1), 3-23.
- Hilton, N.Z., Harris, G.T., Rice, M.E., Houghton, R.E., & Eke, A.W. (2007). An indepth actuarial assessment for wife assault recidivism: The Domestic Violence Risk Appraisal Guide. *Law and Human Behavior*, *32*(5), 150-163.
- Hilton, N.Z., Harris, G.T., Rice, M.E., Lang, C., & Cormier, C.A. (2004). A brief actuarial assessment for the prediction of wife assault recidivism: The ODARA. *Psychological Assessment*, *16*(3), 267-275.
- Hoge, R.D., & Andrews, D.A. (1996). An investigation of risk and protective factors in a sample of youthful offenders. *Journal of Child Psychology and Psychiatry*,

37(4), 419-424.

- Holtzworth-Munroe, A., & Anglin, K. (1991). The competency of responses given by maritally violent versus nonviolent men to problematic marital situations. *Violence and Victims, 6*(4), 257-269.
- Holtzworth-Munroe, A., & Stuart, G.L. (1994). Typologies of male batterers: Three subtypes and the differences among them. *Psychological Bulletin, 116*(3), 476-497.
- Holtzworth-Munroe, A., Meehan, J.C., Herron, K., Rehman, U., & Stuart, G.L. (2000). Testing the Holtzworth-Munroe (1994) batterer typology. *Journal of Consulting and Clinical Psychology, 68*(6), 1000-1019.
- Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment, 15*(4), 443-445.
- Hunsley, J., & Meyer, G.J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15*(4), 446-455.
- Jones, A.S., & Gondolf, E.W. (2001). Time-varying risk factors for reassault among batterer program participants. *Journal of Family Violence, 16*(4), 345-359.
- Kantor, G.K., Jasinski, J.L., & Aldarondo, E. (1994). Sociocultural status and incidence of marital violence in Hispanic families. *Special issue: Violence against women of colour. Violence and Victims, 9*(3), 207-222.
- Kaplan, R.M., & Saccuzzo, D.P. (2005). *Psychological testing: Principles, applications, and issues (6th ed.)*. Belmont, CA: Wadsworth/Thomson.
- Kaslow, N.J., Thompson, M.P., Meadows, L.A., Jacobs, D., Gibb, B., Bornstein, H., ... Phillips, K. (1998). Factors that mediate and moderate the link between partner abuse and suicidal behaviour in African American women. *Journal of Consulting and Clinical Psychology, 66*(3), 553-540.
- Kelley, T.L. (1927). *Interpretation of educational measurements*. Yonkers, NY: World Book Company.
- Kelly, J.B., & Johnson, M.P. (2008). Differentiation among types of intimate partner violence: Research update and implications for interventions. *Family Court Review, 46*(3), 476-499.
- Koziol-McLain, J., Rameka, M., Giddings, L., Fyfe, E., & Gardiner, J. (2007). Partner violence prevalence among women attending a Māori health provider clinic. *Australian and New Zealand Journal of Public Health, 31*(2), 143-148.

- Kraemer, H.C., Kazdin, A.E., Offord, D.R., Kessler, R.C., Jensen, P.S., & Kupfer, D.J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry*, 54(4), 337-343.
- Kroner, D.G., & Mills, J.F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior*, 28(4), 471-479.
- Kropp, P.R. (2003). *Validity of law enforcement officers evaluations of risk*. Paper presented at the 3rd Annual Conference of the International Association of Forensic Mental Health Services, Miami, Florida.
- Kropp, P.R. (2004). Some questions regarding spousal assault risk assessment. *Violence Against Women*, 10(6), 676-697.
- Kropp, P.R., & Gibas, A. (2010). The Spousal Assault Risk Assessment Guide (SARA). In R.K. Otto & K.D. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 227-250). New York, NY: Taylor and Francis Group.
- Kropp, P.R., & Hart, S.D. (2000). The Spousal Assault Risk Assessment (SARA) guide: Reliability and validity in adult male offenders. *Law and Human Behaviour*, 24(1), 101-118.
- Kropp, P.R., & Hart, S.D. (2004). *B-SAFER (Brief Spousal Assault Form for the Evaluation of Risk): A tool for criminal justice professionals*. Ottawa, Canada: Department of Justice.
- Kropp, P.R., Hart, S.D., Webster, C.D., & Eaves, D. (1994). *Manual for the Spousal Assault Risk Assessment Guide*. Vancouver, Canada: British Columbia Institute on Family Violence.
- Kropp, P.R., Hart, S.D., Webster, C.D., & Eaves, D. (1995). *Manual for the Spousal Assault Risk Assessment Guide (2nd ed.)*. Vancouver, Canada: British Columbia Institute on Family Violence.
- Kropp, P.R., Hart, S.D., Webster, C.D., & Eaves, D. (1999). *Manual for the Spousal Assault Risk Assessment Guide (3rd ed.)*. Toronto, Canada: Multi-Health Systems.
- Krug, E.G., Dahlberg, L.L., Mercy, J.J., Zwi, A.B., & Lorano R. (Eds.). (2002). *World report on violence and health*, World Health Organisation, Geneva.
- Kulkin, H.S., Williams, J., Borne, H.F., de la Bretonne, D., & Laurendine, J. (2007). A review of research on violence in same-gender couples: A resource for clinicians. *Journal of Homosexuality*, 53(4), 71-87.

- La Taillade, J.J., & Jacobson, N.S. (1997). Domestic violence: Antisocial behavior in the family. In D.M. Stoff, J. Breiling & J.D. Maser (Eds.), *Handbook of antisocial behavior* (pp. 534-550). New York, NY: John Wiley and Sons.
- Langhinrichsen-Rohling, J., Neidig, P., & Thorn, G. (1995). Violent marriages: Gender differences in current marital violence and childhood abuse. *Journal of Family Violence, 10*(2), 159-176.
- Langstrom, N. (2004). Accuracy of actuarial procedures for assessment of sexual offender recidivism risk may vary across ethnicity. *Sexual Abuse: A journal of research and treatment, 16*(2), 107-120.
- LeBreton, J.M., & Senter, J.L. (2008). Answers to twenty questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815-852.
- Leonard, K.E., & Senchak, M. (1996). Prospective prediction of husband marital aggression within newlywed couples. *Journal of Abnormal Psychology, 105*(3), 369-380.
- Lewis, S.F., & Fremouw, W. (2001). Dating violence: A critical review of the literature. *Clinical Psychology Review, 21*(1), 105-127.
- Link, B., & Steueve, A. (1994). Psychotic symptoms and the violent/illegal behavior of mental patients compared to community control. In J. Monahan & H. Steadman (Eds.), *Violence and mental disorders: Developments of risk assessment* (pp. 137-159). Chicago, IL: University of Chicago Press.
- Litwack, T.R., & Schlesinger, L.B. (1999). Dangerousness risk assessments: Research, legal and clinical considerations. In A.K. Hess & I.B. Weiner (Eds.), *Handbook of forensic psychology (2nd ed.)* (pp. 171-217). New York, NY: John Wiley and Sons.
- Mahoney, P., Williams, L.M., & West, C.M. (2001). Violence against women by intimate relationship partners. In C.M. Renzetti, J.L. Edleson & R. Kennedy Bergen (Eds.), *Sourcebook on violence against women* (pp. 143-178). Thousand Oaks, CA: Sage.
- Marie, D., Fergusson, D.M., & Boden, J.M. (2008). Ethnic identity and intimate partner violence in a New Zealand birth cohort. *Social Policy Journal of New Zealand, 33*, 126-145.
- Martin, E.K., Taft, C.T., & Resick, P.A. (2007). A review of marital rape. *Aggression and Violent Behaviour, 12*(3), 329-347.

- Martin, J., & Pritchard, R. (2010). *Learning from tragedy: Homicide within families in New Zealand 2002-2006*. Wellington, New Zealand: Ministry of Social Development.
- McCarroll, J.E., Thayer, L.E., Newby, J.H., Ursano, R.J., Fullerton, C.S., & Norwood, A.E. (2000). Are respondents who omit Conflict Tactics Scale items more violent than those who omit none? A methodological note. *Journal of Interpersonal Violence, 15*(8), 872-881.
- McCue, M.L. (2008). *Domestic violence: A reference handbook (2nd ed.)*. Santa Barbara, CA: ABC-CLIO Incorporated.
- McFarlane, J., Campbell, J.C., & Watson, K. (2002). Intimate partner stalking and femicide: Urgent implications for women's safety. *Behavioral Sciences and the Law, 20*(1-2), 51-68.
- McKinney, C.M., Caetano, R., Ramisetty-Mikler, S., & Nelson, S. (2009). Childhood family violence and perpetration and victimisation of intimate partner violence: Findings from a national population-based study of couples. *Annals of Epidemiology, 19*(1), 25-32.
- McMaster, K., & Gregory, D. (2003). Current approaches to working with family violence. In K. McMaster & A. Wells (Eds.), *Innovative approaches to stopping family violence* (pp. 16-35). Wellington, New Zealand: Steele Roberts.
- Maynard, K., Coebergh, B., Anstiss, B., Bakker, L., & Huriwai, T. (1999). Kit e arotu – Toward a new assessment: The identification of cultural factors which predispose Māori to crime. *Social Policy Journal of New Zealand, 13*, 43-58.
- Mederos, F. (1999). Batterer intervention programs: Past and future prospects. In E. Pence & M. Shepard (Eds.), *Coordinating community responses: Lessons from Duluth and beyond* (pp. 127-150). Newbury Park, CA: Sage.
- Mehrabian, A., & Ferris, S.R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology, 31*(3), 248-252.
- Mehta, C.R., & Patel, N.R. (1995). Exact logistic regression: Theory and examples. *Statistics in Medicine, 14*(19), 2143-2160.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*(11), 1012-1027.
- Messick, S. (1993). Validity. In R.L. Linn (Ed.), *Educational measurement (2nd ed.)* (pp.13-104). Phoenix, AZ: American Council on Education and Oryx Press.

- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45(1-3), 35-44.
- Mihalic, S.W., & Elliot, D. (1997). A social learning theory model of marital violence. *Journal of Family Violence*, 12(1), 21-47.
- Millon, T. (1983). *Manual for the MCMI-II*. Minneapolis, MN: National Computer Systems.
- Millon, T. (1987). *Manual for the MCMI-II (2nd ed.)*. Minneapolis, MN: National Computer Systems.
- Mills, J.F., Kroner, D.G., & Morgan, R.D. (2011). *Clinicians guide to violence risk assessment*. New York, NY: Guilford Press.
- Moffitt, T.E., & Caspi, A. (1999). Findings about partner violence in the Dunedin Multidisciplinary Health and Development Study. *National institute of justice research in brief*. Washington, DC: US Department of Justice. Retrieved from <http://www.ncjrs.org>.
- Mojtabai, R. (2006). Psychotic-like experiences and interpersonal violence in the general population. *Social Psychiatry and Psychiatric Epidemiology*, 41(3), 183-190.
- Monahan, J. (1981). *The clinical prediction of violence*. Beverley Hills, CA: Sage.
- Monahan, J. (1992). Mental disorder and violent behavior: Perceptions and evidence. *American Psychologist*, 47(4), 511-521.
- Monahan, J., & Steadman, H. (Eds.) (1994). *Violence and mental disorder: Developments in risk assessment*. Chicago, IL: University of Chicago Press.
- Monahan, J., Steadman, H.J., Silver, E., Appelbaum, P.S., Robbins, P.C., Mulvey, E.P., ... Banks, S. (2001). *Rethinking risk assessment: The MacArthur Study of mental disorder and violence*. New York, NY: Oxford University Press.
- Montgomery, A.A., Graham, A., Evans, P.H., & Fahey, T. (2002). Inter-rater agreement in the scoring of abstracts submitted to a primary care research conference. *BMC Health Services Research*, 2(8), 1-4.
- Mooney, N.P. (2010). *Predicting offending within the New Zealand youth justice system: Evaluating measures of risk, need, and psychopathy* (Unpublished doctorate of clinical psychology dissertation). Massey University, Wellington, New Zealand.
- Morris, A., Reilly, J., Berry, S., & Ransom, R. (2003). *New Zealand national survey of crime victims 2001*. Wellington, New Zealand: Ministry of Justice.

- Morse, B.J. (1995). Beyond the conflict tactics Scale: Assessing gender differences in partner violence. *Violence and Victims, 10*(4), 251-272.
- Mossman, D. (1994). Assessing predictions of violence: being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*(4), 783-792.
- Mowat-Leger, V. (2001). *Risk factors for violence: A comparison of domestic batterers and other violent and non-violent offenders* (Unpublished doctoral dissertation). Carleton University, Ottawa, Canada.
- Mugford, J., & Mugford, S. (1992). Policing domestic violence. In P. Moir & H. Eijkman (Eds.), *Policing Australia: Old issues, new perspectives* (pp. 321-351). Melbourne, Australia: MacMillan.
- Mulvey, E.P., Odgers, C., Skeem, J.L., Gardner, W., Schubert, C., & Lidz, C. (2006). Substance use and community violence: A test of the relation at the daily level. *Journal of Consulting and Clinical Psychology, 74*(4), 743-754.
- Murphy, C.M., & Baxter, V.A. (1997). Motivating batterers to change in treatment context. *Journal of Interpersonal Violence, 12*(4), 607-619.
- Murphy, C.M., & Eckhardt, C. (2005). *Treating the abusive partner: An individualized cognitive-behavioral approach*. New York, NY: Guilford Press.
- Murphy, C.M., & O'Farrell, T.J. (1994). Factors associated with marital aggression in male alcoholics. *Journal of Family Psychology, 8*(3), 321-335.
- Murphy, C.M., Meyer, S.L., & O'Leary, K.D. (1993). Family of origin violence and MCMI-II psychopathology among partner assaultive men. *Violence Victims, 8*(2), 165-176.
- Nelson, B.S., & Wampler, K.S. (2000). Systemic effects of trauma in clinical couples: An exploratory study of secondary trauma resulting from childhood abuse. *Journal of Marital and Family Therapy, 26*(2), 171-184.
- New Zealand Department of Corrections. (2007). Over-representation of Māori in the Criminal Justice System: An exploratory report. Wellington, New Zealand: Department of Corrections.
- New Zealand Family Violence Clearinghouse. (2007). *Overview of family violence fact sheet*. Retrieved from www.nfvc.org.nz/PublicationsArea.aspx?area=Fact+Sheets.
- Newbold, G., & Cross, J. (2008). Domestic violence and pro-arrest policy. *Social Policy Journal of New Zealand, 33*, 1-14.
- Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York, NY:

McGraw-Hill.

- O'Leary, K.D., Malone, J., & Tyree, A. (1994). Physical aggression in early marriage: Prerelationship and relationship effects. *Journal of Consulting and Clinical Psychology, 62*(3), 594-602.
- Pallant, J. (2007). *SPSS survival manual (3rd ed.)*. Crows Nest, Australia: Allen and Unwin.
- Peat, J., & Baton, B. (2005). *Medical statistics: A guide to data analysis and critical appraisal*. Oxford, England: Blackwell Publishing Ltd.
- Pence, E., & Paymar, M. (1993). *Education groups for men who batter: The Duluth model*. New York, NY: Springer.
- Philipse, M.W.G., Koeter, M.W.J., van der Staak, C.P.F., & van den Brink, W. (2005). Reliability and discriminant validity of dynamic reoffending risk indicators in forensic clinical practice. *Criminal, Justice and Behavior, 32*(6), 643-664.
- Pitts, M.K., Smith, A.M.A., Mitchell, A., & Patel, S. (2006). *Private lives: A report on the health and wellbeing of GLBTI Australians*. La Trobe University, Melbourne, Australia: The Australian Research Centre in Sex, Health, & Society.
- Quinsey, V.L., Harris, G.T., Rice, M.E., & Cormier, C.A. (2006). *Violent offenders: Appraising and managing risk (2nd ed.)*. Washington, DC: American Psychological Association.
- Ratner, P.A. (1993). The incidence of wife abuse and mental health status in abused wives in Edmonton, Alberta. *Canadian Journal of Public Health, 84*(4), 246-249.
- Reeves, K., Kropp, P.R., & Cairns, K. (2008). *An independent validation study of the SARA*. Paper presented at the 8th Annual Conference of the International Association of Forensic Mental Health Services, Vienna, Austria.
- Rice, M.E., & Harris, G.T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*(5), 737-748.
- Rice, M.E., & Harris, G.T. (1997). Cross validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. *Law and Human Behavior, 21*(2), 231-241.
- Rice, M.E., & Harris, G.T. (2005). Comparing effect sizes in follow-up studies: ROC, Cohen's *d* and *r*. *Law and Human Behavior, 29*(5), 615-620.
- Rice, M.E., Harris, G.T., & Cormier, C.A. (1992). Evaluation of a maximum security

- therapeutic community for psychopaths and other mentally disordered offenders. *Law and Human Behavior*, 16(4), 399-412.
- Rice, M.E., Harris, G.T., & Quinsey, V.L. (1990). A follow-up of rapists assessed in a maximum security psychiatric facility. *Journal of Interpersonal Violence*, 5(4), 435-448.
- Richardson, J., Coid, J., Petruckevitch, A., Chung, W.S., Moorey, S., & Feder, G. (2002). Identifying domestic violence: Cross sectional study in primary care. *British Medical Journal*, 324(7332), 274-277.
- Riggs, D., Caulfield, M., & Street, A. (2000). Risk for domestic violence: Factors associated with perpetration and victimization. *Journal of Clinical Psychology*, 56(10), 1289-1316.
- Robertson, N. (1999). *Reforming institutional responses to violence against women* (Unpublished doctoral dissertation). University of Waikato, Hamilton, New Zealand.
- Robertson, N., Busch, R., Ave, K., & Balzer, R. (1991). *The Hamilton abuse intervention pilot project: The first three months*. Wellington, New Zealand: Family Violence Prevention Coordinating Committee.
- Rugge, T. (2006). *Risk Assessment of Male Aboriginal Offenders: A 2006 perspective*. Ontario, Canada: Public Safety Canada.
- Sartori, R., & Pasini, M. (2007). Quality and quantity in test validity: How can we be sure the psychological tests measure what they have to? *Quality and Quantity*, 41(3), 359-374.
- Saunders, D.G., & Browne, A. (2000). Intimate partner homicide. In R. T. Ammerman & M. Hersen (Eds.), *Case studies in family violence* (2nd ed.) (pp. 415-449). New York, NY: Plenum.
- Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences*, 3(1), 153-160.
- Schumacher, J.A., Feldbau-Kohn, S., Slep, A.M.S., & Heyman, R.E. (2001). Risk factors for male-to-female partner physical abuse. *Aggression and Violent Behavior*, 6(2-3), 281-352.
- Schwartz, R.C., Reynolds, C.A., Austin, J.F., & Petersen, S. (2003). Homicidality in schizophrenia: A replication study. *The American Journal of Orthopsychiatry*, 71(1), 74-77.
- Scott, H.A. (1995). *The determinants of sexual aggression: An examination of risk*

- factors in self-reported and convicted sexual aggressors* (Unpublished doctoral dissertation). Carleton University, Ottawa, Canada.
- Sherman, L.W., & Berk, R.A. (1984). *The Minneapolis domestic violence experiment*. Washington, DC: Police Foundation Reports.
- Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.
- Smith, C., & Thornberry, T.P. (1995). The relationship between childhood maltreatment and adolescent involvement in delinquency. *Criminology*, 33(4), 451-481.
- Stark, E. (2007). *Coercive control: The entrapment of women in personal life*. New York, NY: Oxford University Press.
- Statistics Canada. (2011). *Family violence in Canada: A statistical profile*. Ontario, Canada: Statistics Canada.
- Stemler, S.E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research and Evaluation*, 9(4). Retrieved from <http://PAREonline.net/getvn.asp?v=9&n=4>.
- Stewart, A., Dennison, S., & Waterson, E. (2002). *Pathways from child maltreatment to juvenile offending*. Canberra, Australia: Criminology Research Council. Retrieved from www.aic.gov.au/crc/reports/stewart.html.
- Stith, S.M., Smith, D.B., Penn, C.E., Ward, D.B., & Tritt, D. (2004). Intimate partner physical abuse perpetration and victimization risk factors: A meta-analysis review. *Aggression and Violent Behavior*, 10(1), 65-98.
- Straus, M.A. (2004). Prevalence of violence against partners by male and female university students worldwide. *Violence Against Women*, 10(7), 790-811.
- Straus, M.A. (2004). *Scoring the CTS2 and CTSPC*. Retrieved from <http://gauss.unh.edu/~mas2/CTS28a3.pdf>.
- Straus, M.A., Hamby, S.L., Boney-McCoy, S., & Sugarman, D.B. (1996). The revised Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues*, 17(3), 283-316.
- Straus, M.E., & Smith, G.T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, 5(1), 1-25.
- Streiner, D.L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, 80(3), 217-222.
- Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Boston,

MA: Allyn and Bacon.

- Taft, C.T., Murphy, C.M., King, D.W., Musser, P.H., & DeDeyn, J.M. (2003). Process and treatment adherence factors in group cognitive-behavioral therapy for partner violent men. *Journal of Consulting and Clinical Psychology, 71*(4), 812-820.
- Tanilon, J., Segers, M., Vedder, P., & Tillema, H. (2009). Development and validation of an admission test designed to assess samples of performance and academic tasks. *Studies in Educational Evaluation, 35*(4), 168-173.
- Thompson, R., Bonomi, A., Anderson, M., Reid, R., Dimer, J., Carrell, D., ... Rivara, F. (2006). Intimate partner violence prevalence, types, and chronicity in adult women. *American Journal of Preventive Medicine, 30*(6), 447-457.
- Tinsley, H., & Weiss, D. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology, 22*(4), 358-376.
- Tjaden, P., & Thoennes, N. (1998). *Extent, nature, and consequences of intimate partner violence*. Washington, DC: United States Department of Justice.
- Tolan, P., D. Gorman-Smith and D. Henry (2006). "Family violence". *Annual Review of Psychology, 57*:557-583.
- Tolman, R.M., & Edleson, J.L. (1995). Intervention for men who batter: A review of research. In S.M. Stith & M.A. Straus (Eds.), *Understanding partner violence: Prevalence, causes, consequences, and solutions* (pp. 262-272). Minneapolis, MN: National Council on Family Relations.
- Uebersax, J. (2010). *Statistical methods for rater and diagnostic agreement*. Retrieved from <http://www.john-uebersax.com/stat/agree.htm>.
- Vega, E M., & O'Leary. K.D. (2007). Test-retest reliability of the Revised Conflict Tactics Scales (CTS2). *Journal of Family Violence, 22*(8), 703-708.
- Walby, S. (2004). *The cost of domestic violence*. London, England: Women and Equality Unit. Retrieved from <http://www.devon.gov.uk/de/text>.
- Walter, S.D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine, 17*(1), 101-110.
- Waltz, J., Babcock, J., Jacobson, N., & Gottman, J. (2000). Testing a typology of batterers. *Journal of Consulting and Clinical Psychology, 68*(4), 658-669.
- Ward, A., & Dockerill, J. (1999). The predictive accuracy of the violent offender treatment program risk assessment scale. *Criminal Justice and Behavior, 26*(1), 125-140.

- Webster, C.D., Dickens, B., & Addario, S. (1985). *Constructing dangerousness: Scientific, legal, and policy implications*. Toronto, Canada: University of Toronto.
- Webster, C.D., Douglas, K.S., Belfrage, H., & Link, B.G. (2000). Capturing Change: An approach to managing violence and improving mental health. In S. Hodgins (Ed.) *Violence among the mentally ill: Effective treatments and management strategies* (pp. 119-144). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Webster, C.D., Douglas, K.S., Eaves, D., & Hart, S.D. (1997). *HCR-20: Assessing risk for violence (Version 2)*. British Columbia, Canada: Simon Fraser University.
- Webster, C.D., Martin, M.-L., Brink, J., Nicholls, T.L., & Middleton, C. (2004). *Manual for the Short-Term Assessment of Risk and Treatability (START)*. Ontario, Canada: Forensic Psychiatric Services Commission.
- Weisz, A N., Tolman, R.M., & Saunders, D.G. (2000). Assessing the risk of severe domestic violence: The importance of survivors' predictions. *Journal of Interpersonal Violence, 15*(1), 75-90.
- Whitfield, C.L., Anda, R.F., Dube, S.R., & Felitti, V.J. (2003). Violent childhood experiences and the risk of intimate partner violence in adults: Assessment in a large health maintenance organization. *Journal of Interpersonal Violence, 18*(2), 166-185.
- Widom, C.S. (1989). The cycle of violence. *Science, 244*(4901), 160-166.
- Wiggins, G. (1993). Assessment: Authenticity, context, and validity. *Phi Delta Kappa, 75*(3), 200-208.
- Williams, K.R., & Houghton, A.B. (2004). Assessing the risk of domestic violence reoffending: A validation study. *Law and Human Behavior, 28*(4), 437-455.
- Wilson, J. (1998). *Guidelines for clinical risk assessment management in mental health services*. Wellington, New Zealand: Ministry of Health.
- Wilson, M., Daly, M., & Wright, C. (1993). Uxoricide in Canada: Demographic risk patterns. *Canadian Journal of Criminology, 35*, 263-291.
- Wong, T., & Hisashima, J. (2008). *Domestic violence exploratory study on the DVSI and SARA, state of Hawaii, 2003–2007 (ICIS Technical Report No. 1)*. Hawaii, HI: Hawaii State Department of Health. Retrieved from <http://hawaii.gov/icis/documents/SARA-DVSI%20Exploratory%20Study%20%28Oct%202008%29.pdf>

- Wright, K. N., Clear, T. R., & Dickson, P. (1984). Universal applicability of probation risk- assessment instruments. *Criminology*, 22 (1), 113-134.
- Zorrilla, B., Pires, M., Lasheras, L., Morant, C., Seoane, L., Sanchez, L.M., ... Durbán, M. (2010). Intimate partner violence: Last year prevalence and association with socio-economic factors among women in Madrid, Spain. *European Journal of Public Health*, 20(2), 169-175.

APPENDIX A: DULUTH MODEL WHEELS



Figure A1. Power and control wheel (Domestic Abuse Intervention Project).

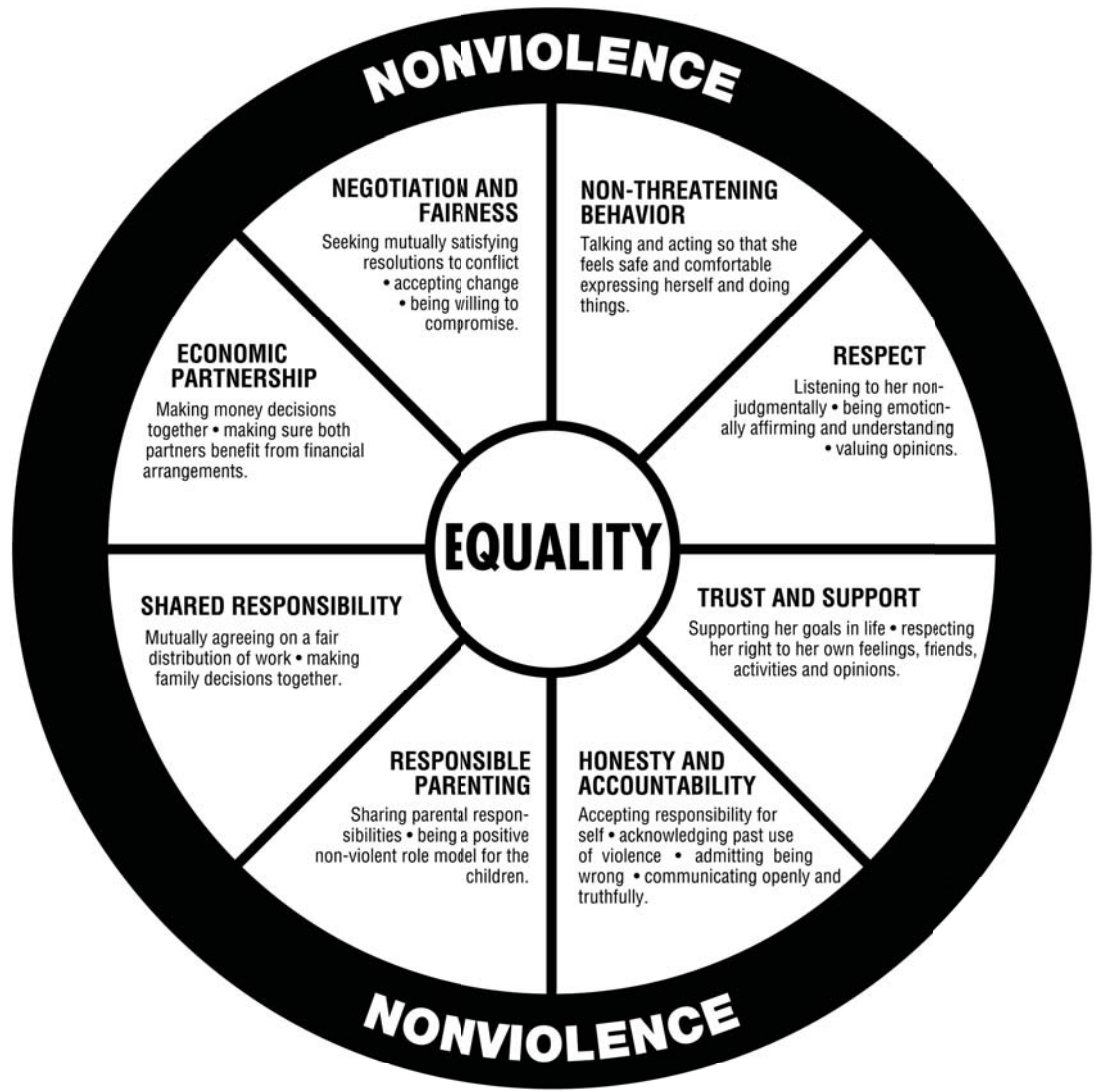


Figure A2. Equality wheel (Domestic Abuse Intervention Project).

APPENDIX B: PARTICIPANT INFORMATION AND CONSENT FORMS

The Evaluation of the Spousal Assault Risk Assessment Guide



Participation Information Sheet

You are invited to take part in a research study that will evaluate the Spousal Assault Risk Assessment Guide for a New Zealand population. Please read this information sheet carefully before deciding whether or not to participate.

Researcher information

This research is being conducted by myself, Uvonne Bartkiw, who is a Doctorate of Clinical Psychology student at Massey University. I can be contacted via email (06050352@massey.ac.nz) or phone (021 238 6943). The supervisor for this study is Mei Wah Williams who can be contacted via email (m.w.williams@massey.ac.nz) or phone (09 414 0800, extn: 41222).

Who will the project involve?

I wish to recruit 120 men for this study, 100 men with a history of intimate partner violence and 20 men with a history of general violence, but no documented history of intimate partner violence. All of the participants will be recruited from stopping violence programs such as the one you are currently involved in. Any person attending the programme who is over the age of 18 and English speaking is welcome to participate.

What will I have to do?

If you wish to take part in this study you will be asked to participate in one initial session involving an interview and the completion of five questionnaires, focused on your past. This will take between one and two hours. This interview will take place at the stopping violence program you are attending. Some of the interviews will be recorded on audiotape so that the interview can be checked to make sure it has been carried out correctly.

Additionally, I will ask your permission to gather information from the New Zealand Police, if you have had contact with them. The information collected will include:

1. Your offence history.
2. A Police summary of events from your charges.

You can participate in this research with or without giving me permission to collect this information. I will seek no other information.

Each time you meet with me for a session your petrol will be reimbursed with a \$20.00 Motor Transport Association petrol voucher.

What are the benefits of taking part in this research?

In New Zealand there is no standardised risk assessment tool that can be used. It is important to have a standardised tool so that men with a history of intimate partner

violence receive more appropriate court decisions and are directed to more useful stopping violence programmes. Your participation in this study would help future men who have been involved in intimate partner violence. This is the first time this research has been conducted in New Zealand.

Is there any risk or discomfort to participating?

The study does not foresee any direct harm to participants. If you experience any part of the sessions as distressing you will be encouraged to talk to your programme facilitator. The questions asked should not be any more distressing than those you will be asked during your sessions with the stopping violence program you are attending.

Participation

Your participation is entirely voluntary (your choice). Participating in, or withdrawing from, this study will not affect your legal status (if relevant) or your involvement in the programme you are attending. Participation in this research will not count towards the required sessions of the programme you are attending, and is separate from your treatment programme.

General Information

Results of this study may be published or presented at conferences or seminars. No individual will be identified. At the end of this study the list of participants and their study identification number will be disposed of. Any raw data from the study will be retained in secure storage for 10 years, after which it will be destroyed.

A summary of the study findings will be available to all participants who request it.

Everyone who participates has the option of having the consent forms and measures read to them, rather than reading it themselves.

Participant's Rights

You are under no obligation to accept this invitation. If you decide to participate, you have the right to:

- decline to answer any particular question
- withdraw from the study at any time
- ask any questions about the study at any time during participation
- provide information on the understanding that your name will not be used unless you give permission to the researcher;
- be given access to a summary of the project findings when it is concluded
- ask for the audiotape recorder to be turned off at any time during the interview, if your interview is recorded.

Confidentiality

No material, which could personally identify you, will be used in the study. The data will be used only for the purposes of this study. Only myself, the researcher will have access to personal information and this will be kept secure and strictly confidential. All questionnaires, interview material and police information will have an identification number on them, instead of your name. This information will be kept separate from the consent form and stored in a locked cabinet.

However, if you tell me that you intend to harm yourself or another person I will discuss this with you and your programme director. I may also contact my supervisors. If I believe that these steps have not reduced the risk of harm I may need to contact other people to ensure the safety of yourself and other people.

Committee Approval Statement

This project has been reviewed and approved by the Massey University Human Ethics Committee: Northern, Application 09/005. If you have any concerns about the conduct of this research, please contact Dr Denise Wilson, Chair, Massey University Human Ethics Committee: Northern, telephone (09) 414 0800 x9070, email humanethicsnorth@massey.ac.nz.

Thank you for considering participating in this study

**The Evaluation of the Spousal
Assault Risk Assessment Guide**



Consent Form

1. I have read and understand the information sheet for taking part in the study to evaluate the Spousal Assault Risk Assessment Guide.
2. I have had the opportunity to discuss the research and am satisfied with the answers I have been given.
3. I understand that I can ask more questions at any time.
4. I understand that taking part in this research is voluntary (my choice) and that I can withdraw from the study at any time.
5. I understand that participating in or withdrawing from this research will not affect my legal status (if relevant) or my involvement in the programme I am attending.
6. I understand that taking part in this study is confidential and that no material that could identify me will be used in the research. All identifying information will be kept in a locked cabinet.
7. I agree to have this interview audiotaped.

Yes No

8. I understand that if I take part in this study I can request a written summary of the results and findings by ticking the box below. In doing so I understand that I will have to provide my postal address or email address where the results and findings will be sent.

Address:.....
.....
.....

9. Due to the limits of confidentiality, I also understand the if I tell the researcher that I intend to harm myself or another person that the researcher will discuss this with me and my programme facilitator and/or director. If necessary other people may have to be informed for my or others safety.
10. I have discussed with the researcher that the questions are not intended to find out about future harm, but past behaviour.
11. I know that the person I contact if I have any questions about this study is the researcher Uvonne Callan-Bartkiw (Email: 06050352@uni.massey.ac.nz), or the research supervisor Mei Williams (Phone: 09 414-0800, ext 41222 or Email: M.W.Williams@massey.ac.nz).

I.....(full name- printed)
give consent to take part in this study.

Signature:.....Date:.....

Consent Form to Collect Information from the New Zealand Police

I..... (full name- printed) have a police record and give permission for Uvonne Callan-Bartkiw (the researcher) to collect information about myself from the New Zealand Police.

The information collected will include:

1. My offence history, including the date of my first arrest and the type of charges I have faced.
2. A Police summary of events from my charges.

I agree that I do not have to sign this form. If I do not sign this form I can still take part in the research.

I understand that the information will only be released to the researcher and confidentiality of this information will remain.

I also understand that a copy of this consent form will be kept in my police file.

Signature:.....Date:.....

Witnessed by (full name- printed):.....

Signature:.....Date:.....

APPENDIX C: SEMI-STRUCTURED INTERVIEW GUIDE

Today I want to talk about what happened that got you to this programme. I'd like you to tell me how it came about and what happened during the incident. I may also ask you some questions about the incident. After that, I'll ask you about some more general questions. This will include what it was like when you were growing up, your experience at school and work, your health, and a little bit about your past relationships.

Firstly, what is your age? Ethnicity? Are you court or self referred? -

Offence

Tell me a little bit about what happened during the incident that brought you here.

If not covered during the description, follow up with these questions:

How old were you when this happened?

Who was the victim? _____

Had you been drinking or using drugs before this incident? _____

Was it planned or more spur of the moment? _____

Were there any injuries, and if so, what were they? _____

Have you ever used a weapon in your offences? Who? Threatened? _____

Have you ever threatened someone with death? Who? _____

What did you actually do (kind of violence)? _____

Have you ever been violent with any family members? How long ago? What happened? How often did it happen? Did you ever threaten to? _____

Have you ever been in any physical fights? How long ago? What happened? How often did it happen? Did you ever threaten to? _____

Do you think being arrested for this incident was fair? _____

What kind of effect do you think the current charge will have on your life? _____

What other offences have you committed? _____

How old were you when you were first arrested? _____

What is the most serious offence you have committed? What happened? _____

Robbery _____ Arson _____ Threatening with a weapon _____
Threatening without a weapon _____ Theft _____ Mischief to public or
private property _____ Break and enter and commit burglary _____ Break
and enter with the plan to commit burglary _____
Fraud _____ Possession of a weapon _____ Prostitution _____
Trafficking drugs _____ Dangerous driving or DIC _____ Resisting arrest
_____ Disturbing the peace _____ Wearing a disguise with the intent to
commit an offence _____ Indecent exposure _____
Who was to blame for these offences? _____

Why do you commit crime? _____

Why did you start committing crime? _____

Do you regret committing these crimes? Why/ why not? _____

How old were you when you started getting into trouble? _____

How old were you when you were first arrested? _____

Were you ever caught by the police under the age of 12? _____

When you were young did you used to steal from your parents, vandalise other property, tell a lot of lies, or hurt animals for fun? How often? How old were you? Did you get caught? What happened? _____

Have you been bailed, had a protection order or on parole? Have you breeched this order? Were you arrested for it? Is that breech the reason you are here today?

School

How many different primary and intermediate schools did you attend? Why did you change? _____

How many different high schools did you attend? Why did you change? _____

What was your attendance like at school? How old were you when you started skipping? _____

Did you like school? How would your teachers describe you? _____

How was your behaviour at school? _____

Fights? How often? How old were you? How often did you start them? Ever hurt anyone badly? Ever use a weapon? _____

Classroom problems? Did you get into trouble a lot in class? Why? How old were you when it started? _____

Ever suspended/ expelled? At what age? What did you do after that? _____

How old were you when you left school? Why did you leave? _____

Work

Are you working at the moment? How long have you been there? Is it full time hours?

How many full time jobs have you had in the past year? _____

Have you ever been fired? Why? _____

Health

Have you used alcohol or drugs in the last year? Which ones? How often? Have you OD? _____

How old were you when you started drinking? (IF BEFORE 18, assess amount) _____

Ever been to a psychologist or psychiatrist? For what? How long ago? Any diagnosis? Are you on any meds or have you been? How has this affected your life? _____

Have you ever been told that you have a personality disorder? Which one? _____

Have you ever thought or tried to commit suicide? When? How often? Did you have a plan? What happened? _____

In the past year, but not currently, have you thought about killing anyone else? How long ago? Who? What changed your mind? _____

Family

Tell me about your childhood

If not covered in description of childhood, follow-up with these questions

Who raised you? Did either of your parents pass away before you were 16? _____

Did you have strict rules? How often did you break them? What happened when you got in trouble? How were you disciplined? _____

Did your parents drink? (GUAGE LEVEL) _____

Were you ever physically or sexually abused? How old? By who? _____

Did you get along with your parents? _____

Did they get along ok with each other, or did they fight? _____

Did they have physical fights that you know about? _____

How did this affect you? _____

Do you have any brothers or sisters? Did you get along with them? _____

Did anyone in your family drink or take drugs? Who? How much? How often? _____

Was there violence in your family? Between who? _____

How old were you when you left? Why did you leave? _____

Relationships

Tell me about your current or most recent partner. _____

How long have you/ had you been together? Did you love her, or was it more of a physical relationship? _____

Did you argue a lot? What happens when you disagree? _____

Were there any physical fights? Any other violence? Has this been escalating over the past year? _____

Was it like this with your previous partners? _____

Has there been any sexual aggression in your past relationships? _____

When is it ok to have a physical fight? _____

(IF IT ENDED) Why did it end? How did you feel when you broke up? _____

Are there any problems with jealousy in your relationship? _____

General Questions

Do you think people are easy to con or manipulate? Do you ever do it? _____

Do you sometimes put on a show of feelings even though you are not feeling it, just because that's what you think others expect? _____

APPENDIX D: PARTICIPANT ASSIGNMENT FOR EACH RATER

Table D1

Participant assignment for each rater, by method.

Participant	Rater 1	Rater 2	Rater 3
1	C	A	N
2	A	N	C
3	A	N	C
4	N	C	A
5	N	C	A
6	C	A	N
7	A	N	C
8	N	C	A

Note: N=Interview notes alone method

A=Audio recording alone method

C=Combined interview notes and audio recording method

APPENDIX E: NORMAL PROBABILITY PLOTS IN THE TOTAL SAMPLE

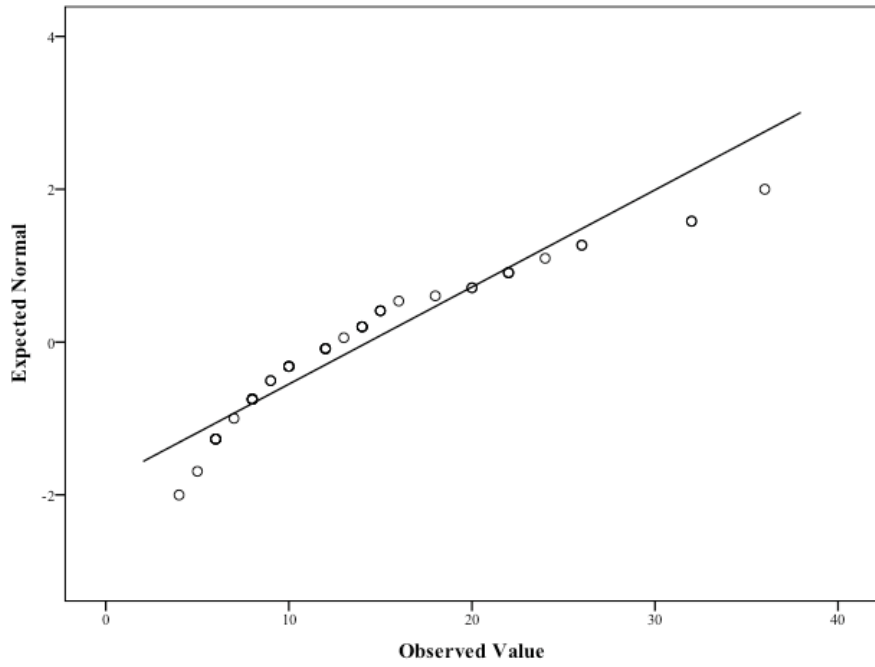


Figure E1. Normal Q-Q plot for the SARA total score.

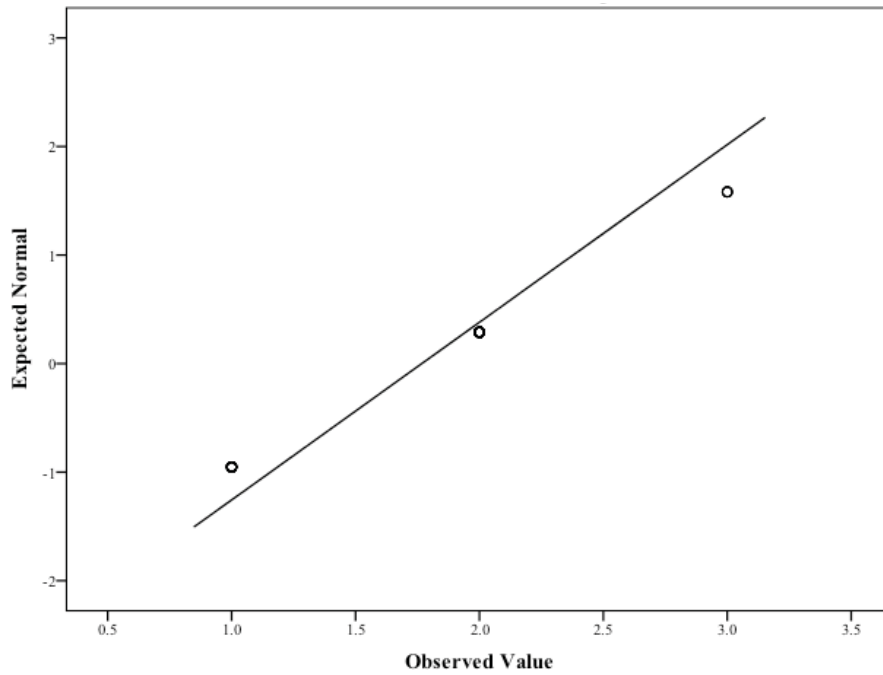


Figure E2. Normal Q-Q plot for the SARA risk rating.

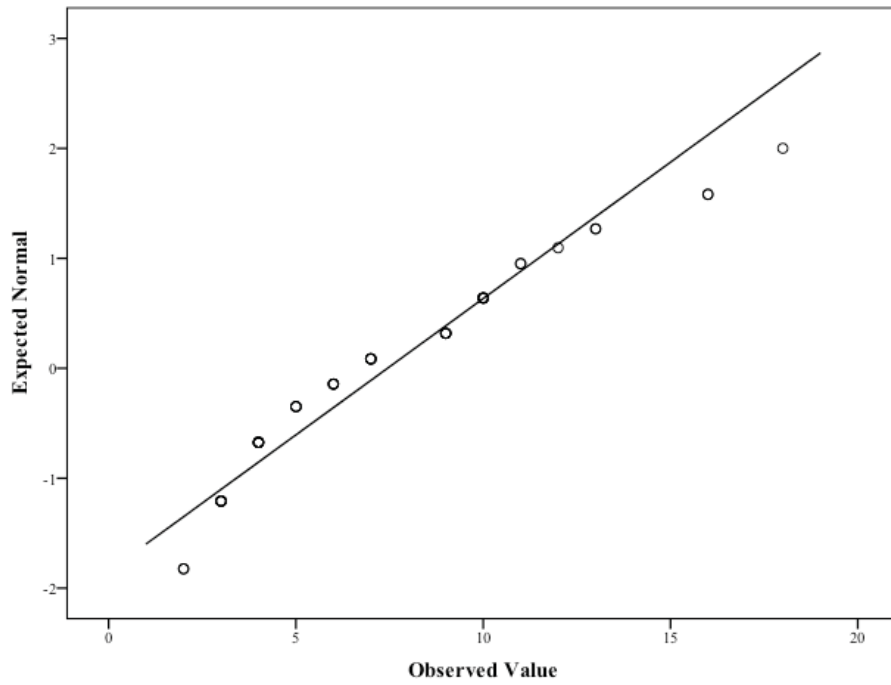


Figure E3. Normal Q-Q plot for the SARA general violence subscale.

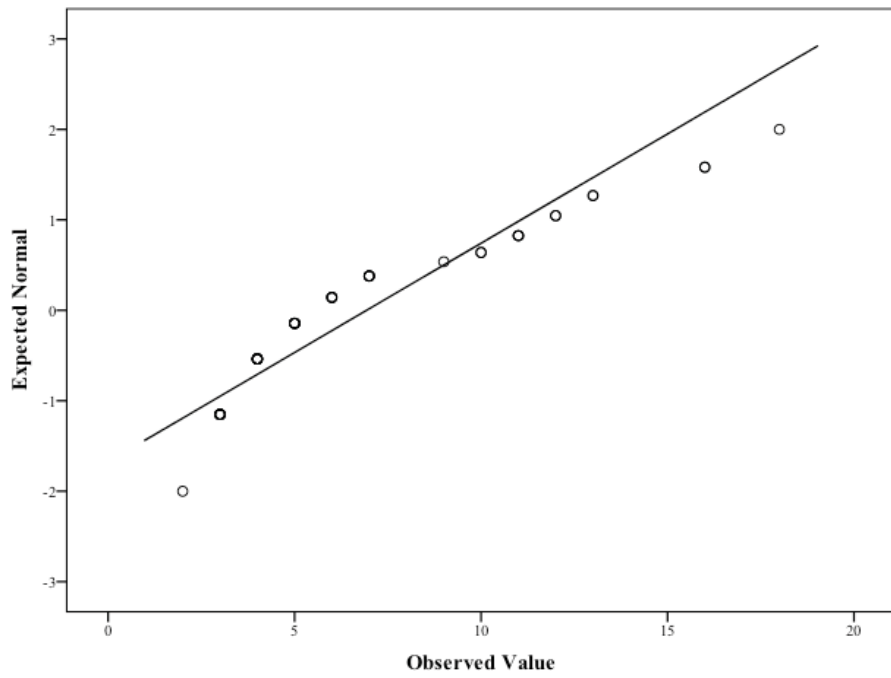


Figure E4. Normal Q-Q plot for the SARA intimate partner violence subscale.

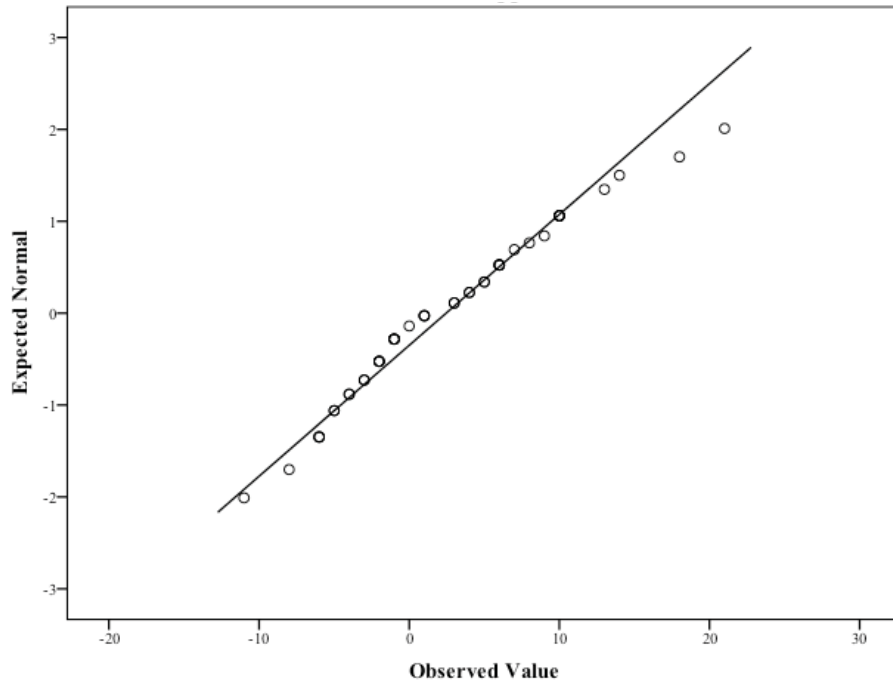


Figure E5. Normal Q-Q plot for the Violence Risk Appraisal Guide.

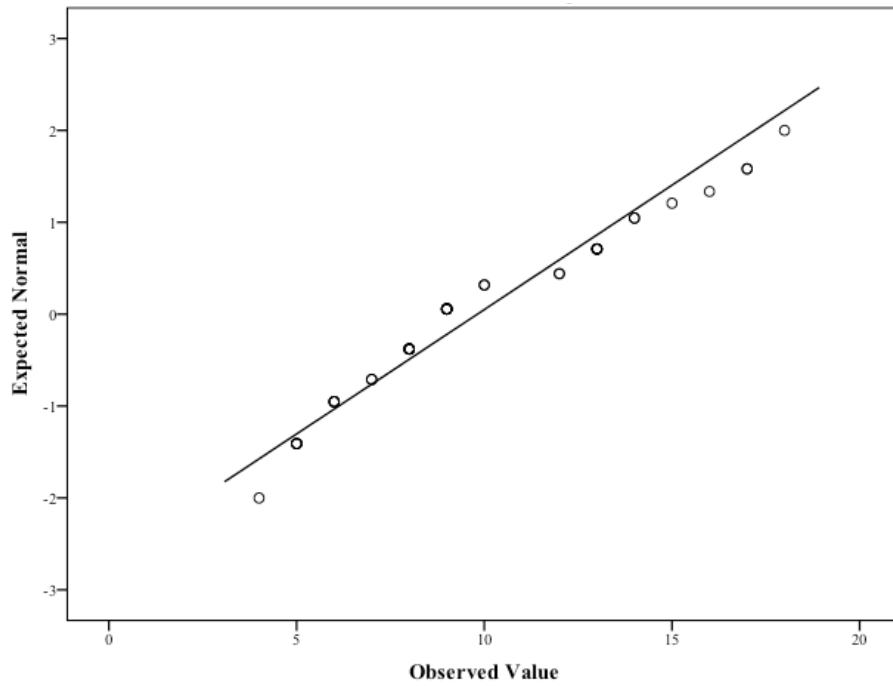


Figure E6. Normal Q-Q plot for the Domestic Violence Screening Instrument.

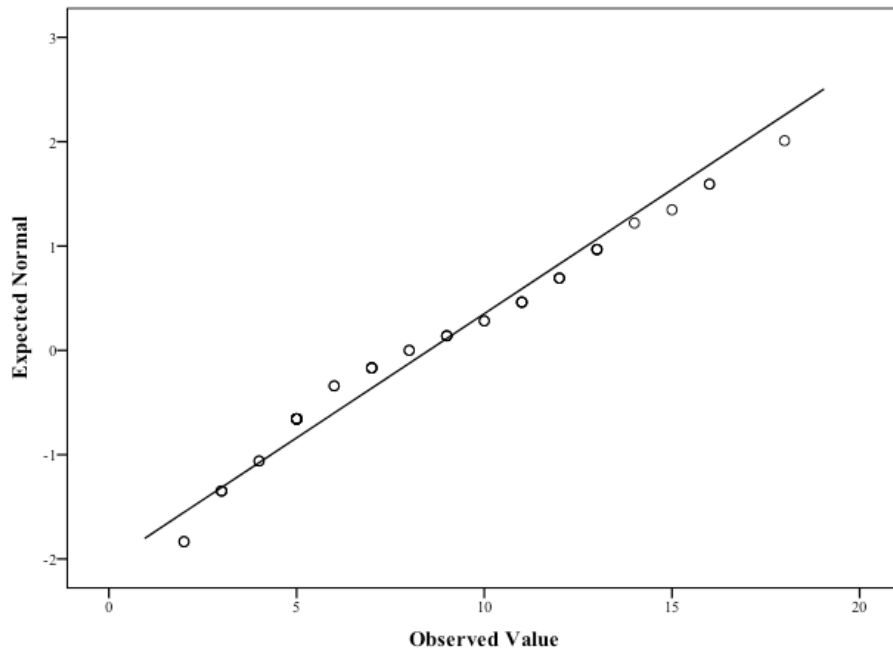


Figure E7. Normal Q-Q plot for the Violent Offender Treatment Programme-Risk Assessment Scale.

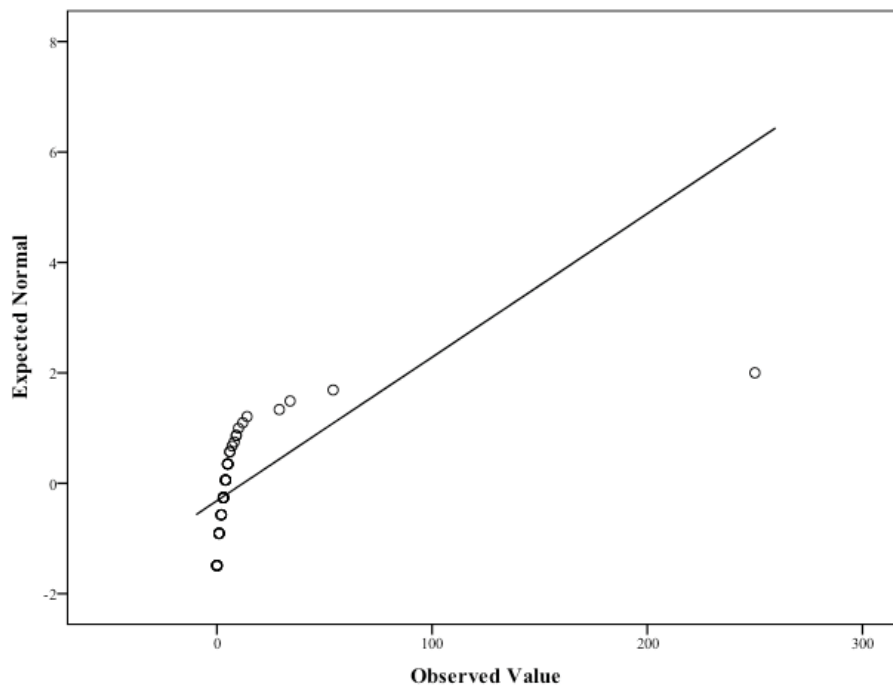


Figure E8. Normal Q-Q plot for the Revised Conflict Tactics Scale physical assault subscale.

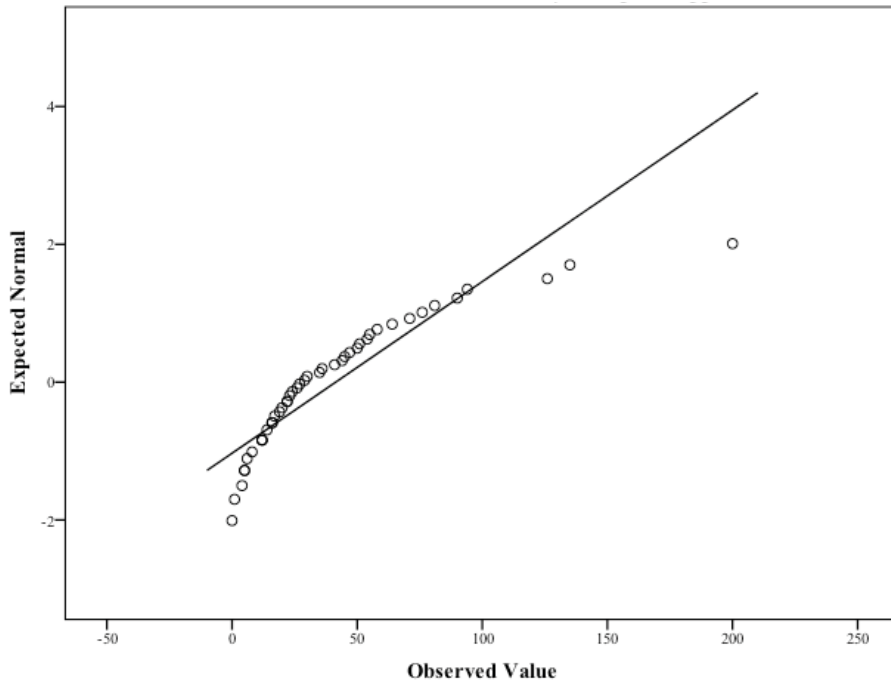


Figure E9. Normal Q-Q plot for the Revised Conflict Tactics Scale psychological aggression subscale.

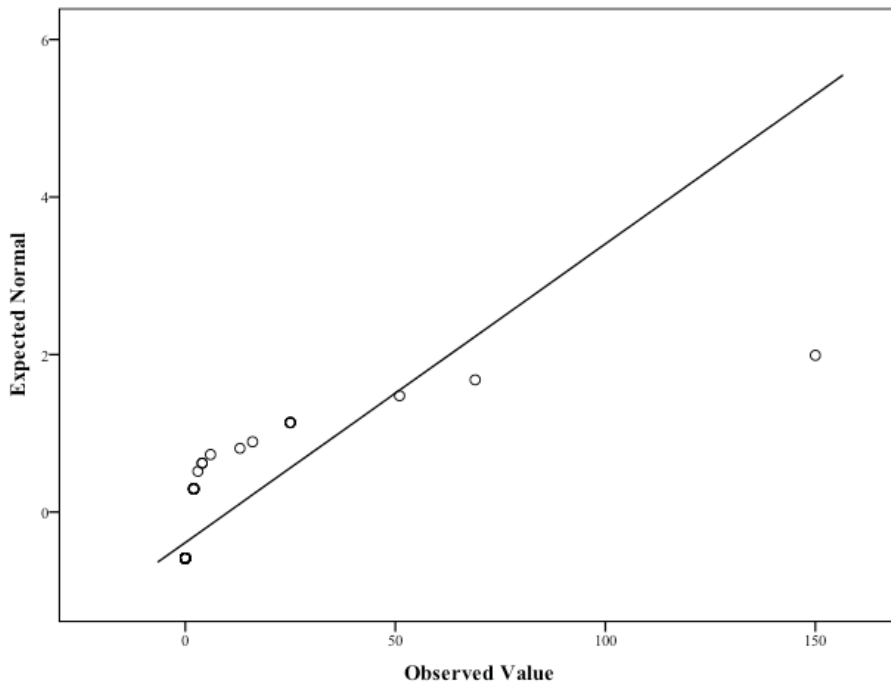


Figure E10. Normal Q-Q plot for the Revised Conflict Tactics Scale sexual coercion subscale.

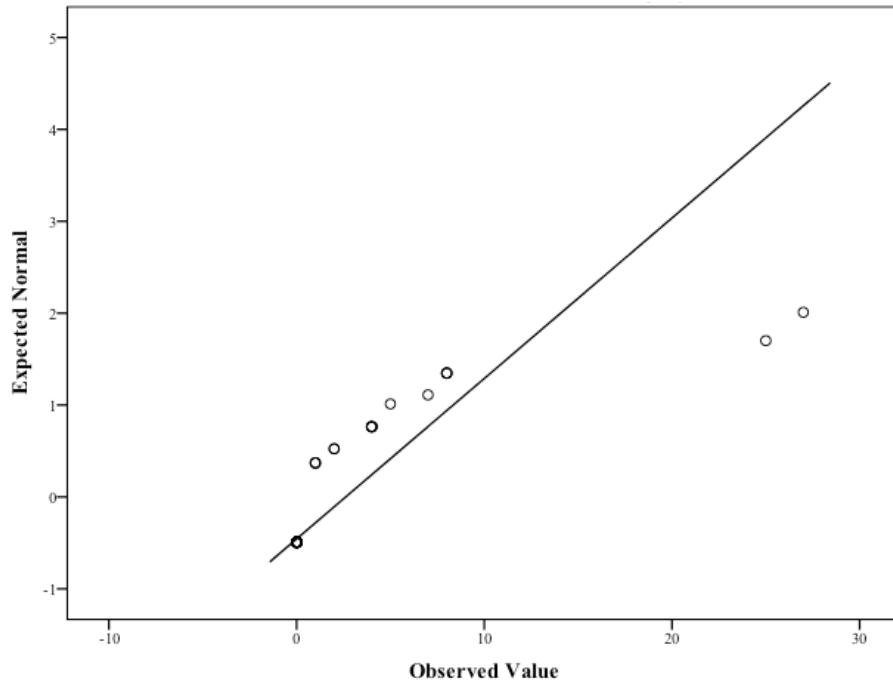


Figure E11. Normal Q-Q plot for the Revised Conflict Tactics Scale injury subscale.

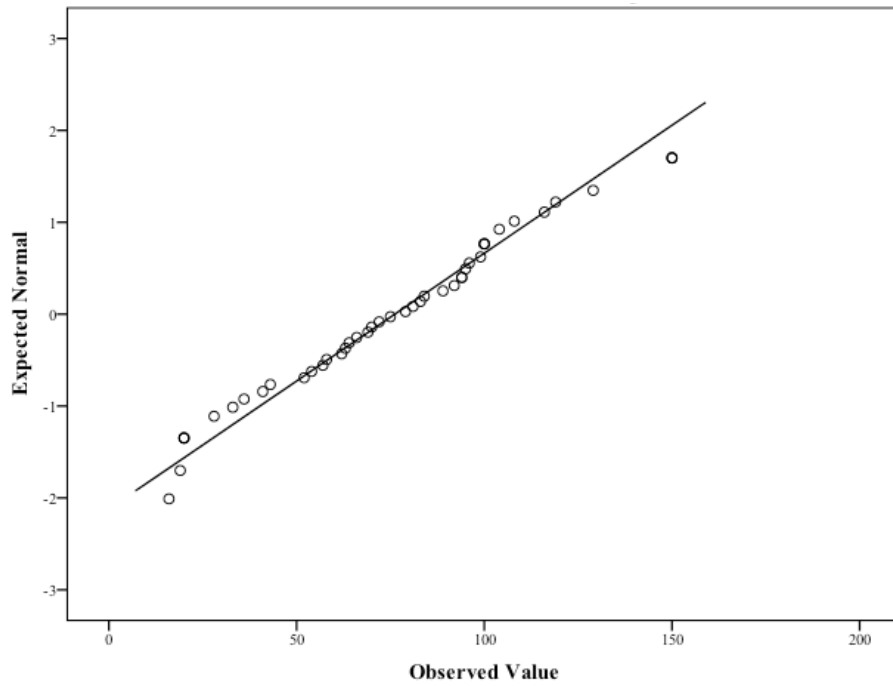


Figure E12. Normal Q-Q plot for the Revised Conflict Tactics Scale negotiation subscale.

APPENDIX F: NORMAL PROBABILITY PLOTS IN THE FOLLOW UP SUB-SAMPLE

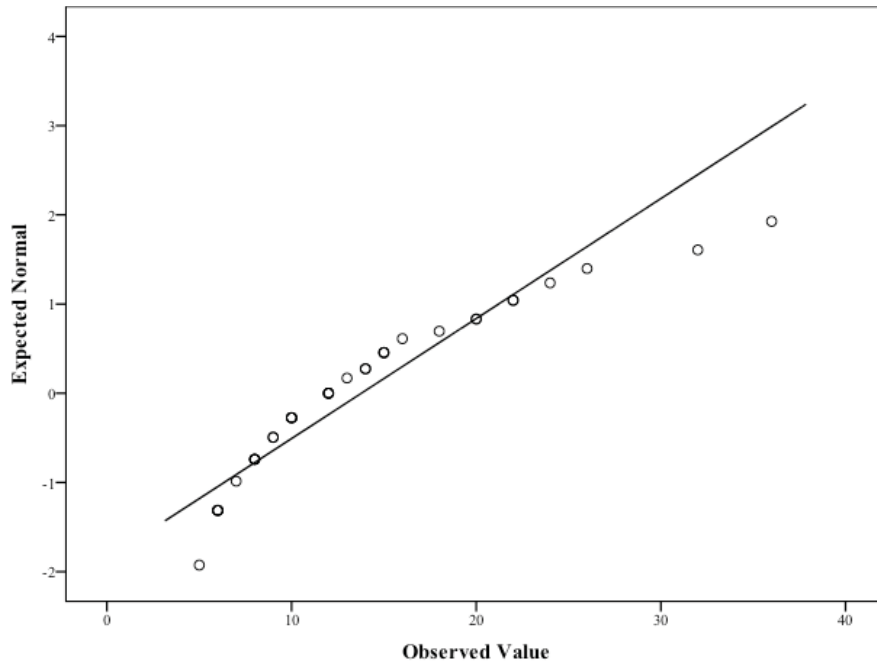


Figure F1. Normal Q-Q plot for the SARA total score.

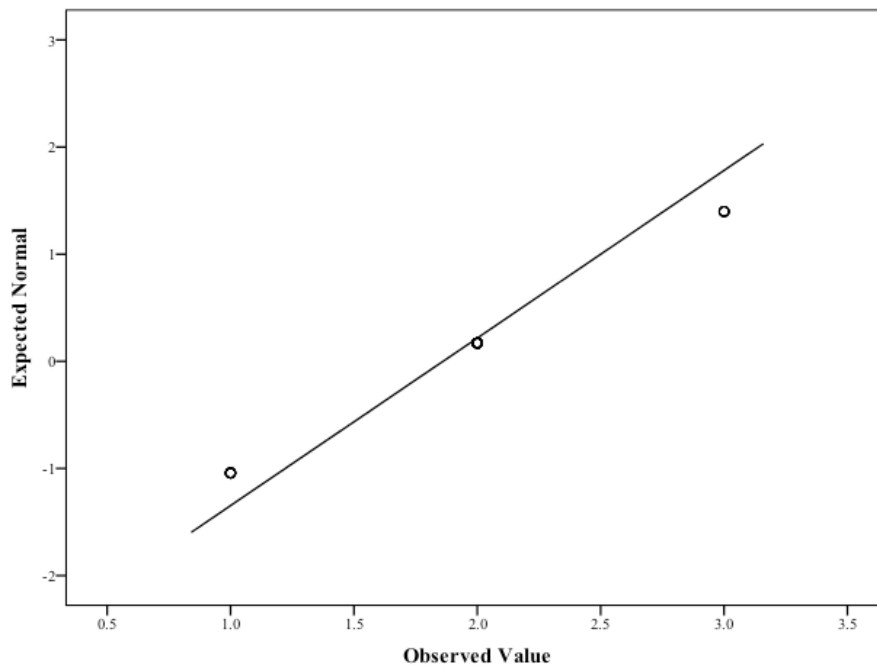


Figure F2. Normal Q-Q plot for the SARA summary risk rating.

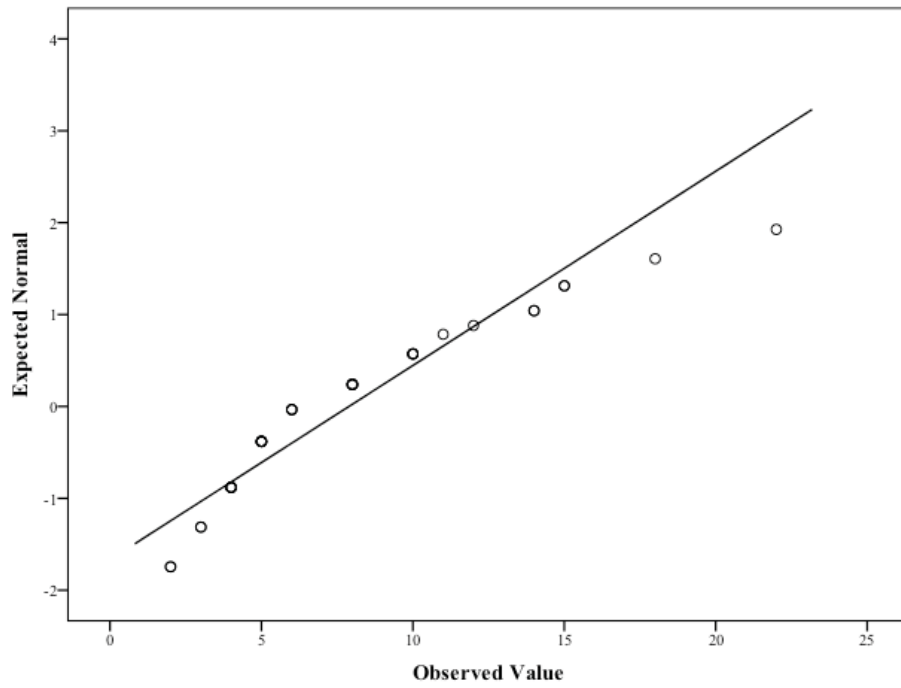


Figure F3. Normal Q-Q plot for the SARA static risk factors.

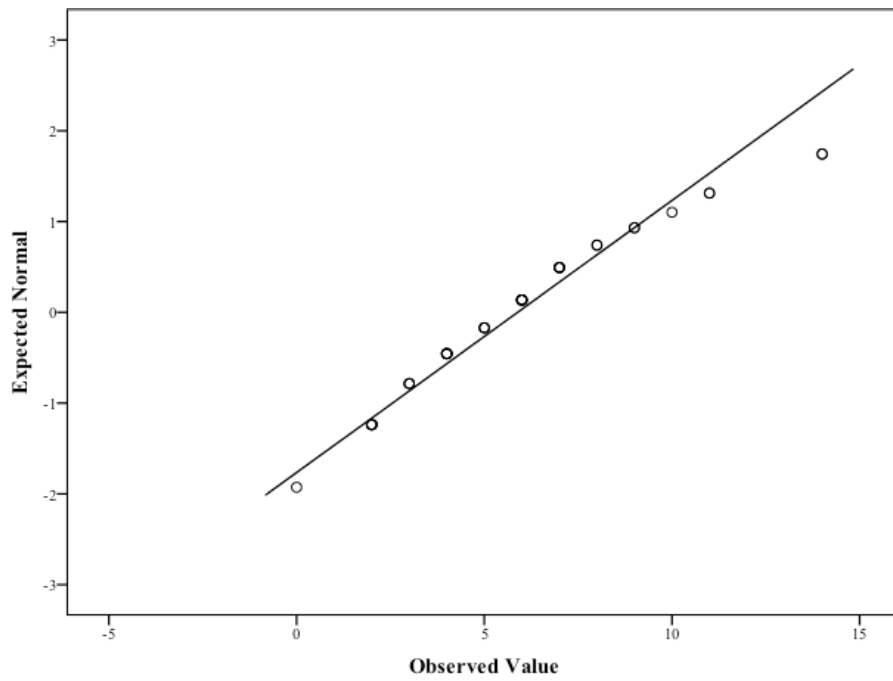


Figure F4. Normal Q-Q plot for the SARA dynamic risk factors.

**APPENDIX G: DISTRIBUTION OF THE SARA SCORES IN THE FOLLOW-
UP SUB-SAMPLE**

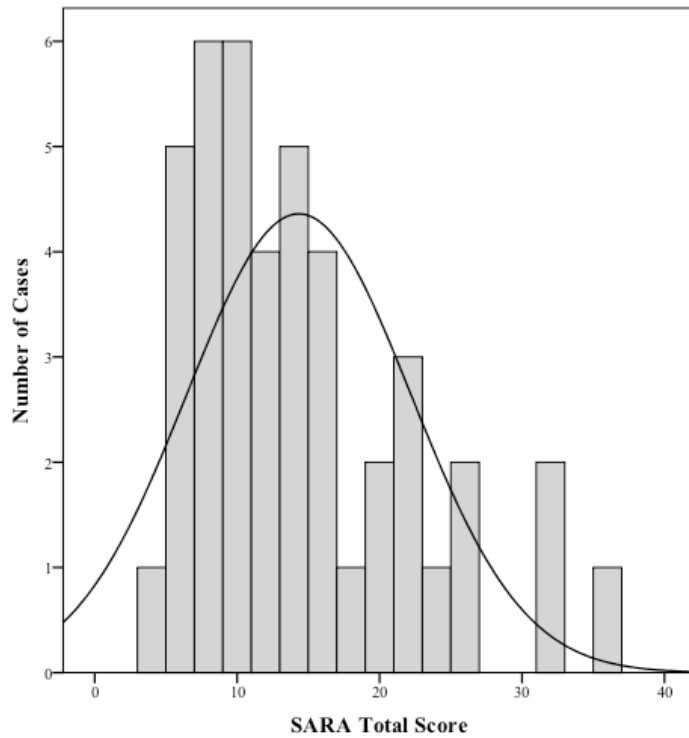


Figure G1. Distribution on the SARA total score.

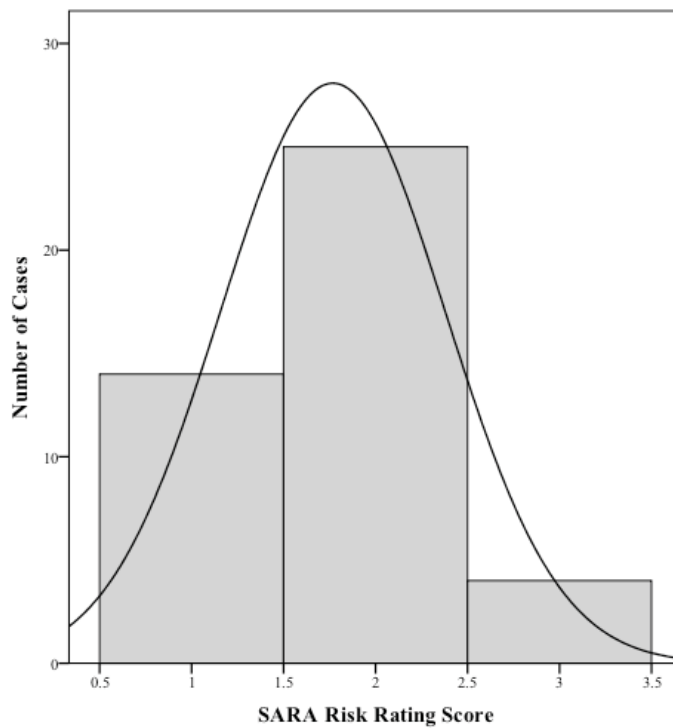


Figure G2. Distribution on the SARA summary risk rating.

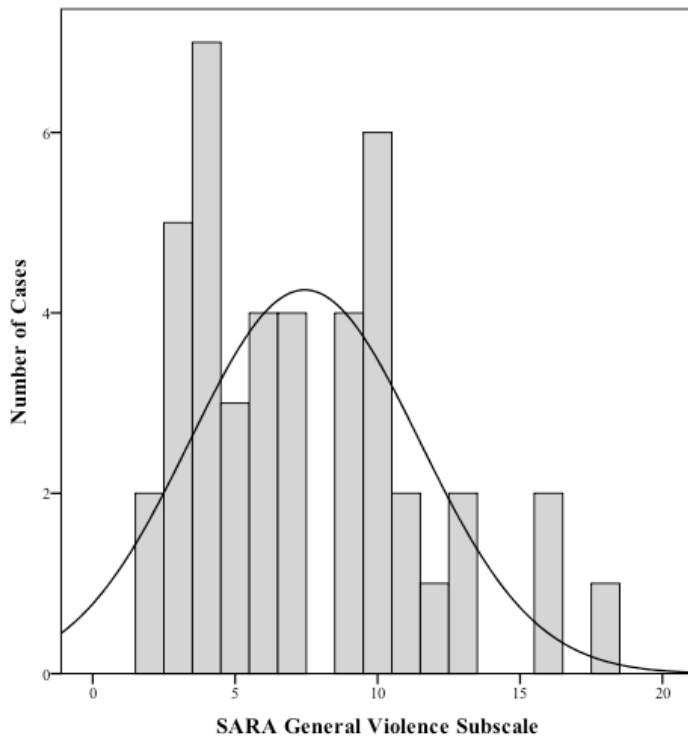


Figure G3. Distribution on the SARA general violence subscale.

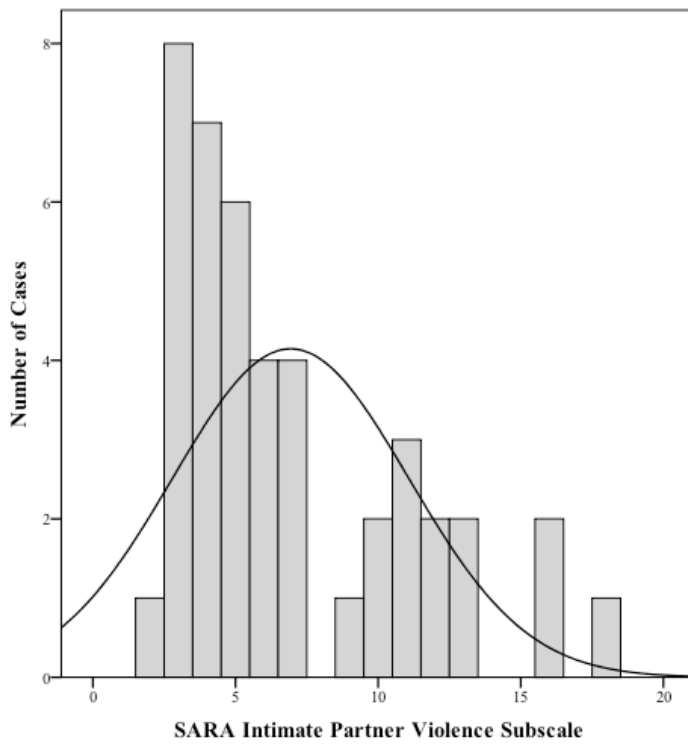


Figure G4. Distribution on the SARA intimate partner violence subscale.