Risk Assessment for Family Violence Aggressors in Aotearoa New Zealand

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Psychology

at

The University of Waikato

by

APRIEL JOLLIFFE SIMPSON

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2023

Abstract

In recent decades Family Violence (FV) researchers have developed risk assessment instruments to enable practitioners to identify the people who are most likely to commit FV in the future, and the families who are most likely to experience ongoing harm. Indeed, risk assessments have become a standard procedure for FV practitioners who are frequently required to make decisions with potential ramifications for families' safety and wellbeing. Yet, despite considerable developments in risk assessment research and practice, risk assessment instruments have limitations that curb their value to FV practice. In addition, such instruments should be empirically validated, but there is limited evidence for the validity of risk assessment instruments used by agencies that respond to FV in Aotearoa New Zealand. Therefore, in this thesis, we examined the risk assessment procedures that New Zealand Police and the Integrated Safety Response (ISR) used for FV, with the purpose of contributing to the evidence base about risk assessment for FV in New Zealand and identifying strengths, weaknesses, and opportunities for improvement.

This thesis contains three manuscripts either published in or submitted to peer-reviewed academic journals; as well as supplementary chapters that support, extend, and integrate the research findings. In the first manuscript, we examined the predictive validity of risk categories from the Dynamic Risk Assessment for family violence (DYRA) and Static Assessment of Family Violence Recidivism (SAFVR)—the two risk assessment instruments that New Zealand Police use when responding to FV cases—for a sample of 2,115 cases with FV episodes reported to police in Waikato and parts of Canterbury between 1 November – 9 December 2018 (1,817 of which had complete risk categories). We examined the instruments' ability to predict recurrence (i.e., a further FV-related call for police service) at three intervals (3 days, 12 weeks, and 24 weeks), and found both instruments had a poor ability to discriminate between aggressors who had a recurrence and those who did not, with

the SAFVR outperforming the DYRA, and the DYRA performing especially poorly for non-intimate partner cases, and aggressors who were Māori or women.

In the second manuscript, we explored the risk assessments that Integrated Safety Response (ISR) triage team members conducted for a sub-sample of 842 cases with FV episodes reported to police between 1 – 14 November 2018. We examined the relationship between the factors triage teams recorded during their risk assessments and the risk categories the teams assigned, then tested the association of both the factors and categories with recurrence and physical recurrence within 24 weeks. The predictive validity of ISR triage teams' risk categories was comparable to the results produced for the DYRA and SAFVR in manuscript one. The ISR risk categories again performed poorly for aggressors who were women, Māori, and for non-intimate partner cases; and overall, fewer than half of the factors triage teams recorded were associated with recurrence or physical recurrence. Then, in an additional chapter, we explored the triage teams' factors' utility for case management and found the factors could be used to create three sub-groups of cases with potential implications for their treatment and management.

Finally, in the third manuscript, we modeled the behavioral patterns of 2,115 aggressors with a FV episode reported to police in Waikato and parts of Canterbury between 1 November – 9 December 2018, by collecting reports for further FV episodes involving those aggressors during the two years after the index episode. We used this information to describe three latent states behind the information reported to police and identified four common pathways through the latent states over time. We concluded this manuscript by discussing changes that could contribute to improvements in how risk for FV aggressors is conceptualized, assessed, and communicated in practice.

In each chapter, we discussed the theoretical and practical implications of the findings, before integrating those findings and implications in a general discussion chapter.

Taken together, the results presented herein indicate that the risk assessment instruments used in New Zealand can predict risk for FV recurrence better than chance, but that those assessments should be further developed to improve their value to FV practice and ensure they adequately capture risk for different types of FV cases and groups of people. This thesis advanced our understanding of the limitations of—and opportunities in—risk assessment research and contributed to the evidence base for risk assessment for FV in New Zealand.

Acknowledgements

Table 1

People Who I Could Not Have Completed this Thesis Without

	B	*p*
New Zealand Police, Evidence Based Policing Centre, Waikato District Police & Integrated Safety Response	Provided access to data and supported the dissemination of research findings	< .001
University of Waikato Doctoral Scholarship & William Georgetti Scholarship	Funded scholarships that made it possible to enroll in doctoral study full-time	< .001
Devon Polaschek	Encouraged me to access opportunities that I never thought were possible, and helped me achieve my goals with invaluable guidance, direction, and support	< .001
Chaitanya Joshi	Attentively supervised the research process and patiently taught me techniques that greatly extended my knowledge and abilities	< .001
Bill Burnett	Brought light to my life and understood while finishing this thesis consumed my days	< .001
Jordan, Sam, Lisa, Emily, Kerry & all other past and present Lab members	Accompanied me on the journey of graduate research, sharing feedback, successes, and challenges along the way	< .001
Mum, Nana, Poppa, Dad, Jeff, family & Burnett family	Encouraged me to read, to be creative, to try, and to be myself	< .001
Nikalah, Jodie, Alyce, Jemma, TEAGA & friends	Supported me to persevere, feigned interest where necessary, and provided good laughs	< .001
Fluffy	Helped me through the most challenging year of my life so far	< .001

Note. Model $R^2 = 1$; $\chi^2(9) = \infty$, $p < .001$. All contributions were significant at $\alpha = .05$.

Table of Contents

List of Publications

Peer-reviewed Journal Articles

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2023). Modeling behavioral patterns of family violence aggressors in Aotearoa New Zealand [Manuscript submitted for publication].

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2023). Unpacking multiagency structured professional judgement risk assessment for family violence. *Journal of Interpersonal Violence.* Advance online publication. https://doi.org/10.1177/08862605221147069

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2021). Predictive validity of the DYRA and SAFVR: New Zealand Police's family violence risk assessment instruments. *Criminal Justice and Behavior, 48*(10), 1487-1508. https://doi.org/10.1177/0093854821997525

Non Peer-reviewed Articles

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2021). Predictive ability of New Zealand Police's family violence risk assessment instruments: Practical implications for frontline policing. *Police Science, 5*(2), 24-25.

Jolliffe Simpson, A. D., Moore, L., Joshi, C. & Polaschek, D. L. L. (2019). Perpetrators of family harm already serving correctional sentences: Risk profiles and episode characteristics. *Psychology Aotearoa.*

Conference Presentations

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2021, October). Modelling the behavioral patterns of family violence aggressors. Paper presented at Australia & New Zealand Evidence Based Policing Conference, online.

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2021, September). The structure of risk factors for intimate partner violence. Paper presented at New Zealand Psychological Society Annual Conference, Auckland.

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2021, September). Modelling the behavioral patterns of family violence aggressors. Paper presented at Department of Corrections Psychological Services National Training Event, Rotorua.

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2021, August). Understanding how practitioners assess risk for intimate partner violence. Paper presented at Waikato Waitaha Forensic Psychology Conference, online.

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2021, July). Understanding how practitioners assess risk for intimate partner violence. Paper presented at School of Computing and Mathematical Sciences Seminar, University of Waikato, Hamilton.

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2020, May). Is there an association between controlling behavior and intimate partner violence? University of Waikato/Evidence Based Policing Centre Conference, online.

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2020, October). Looking for the structure in structured professional judgement: Risk factors for repeat family violence recorded by a multi-agency family violence response. Paper presented at Department of Corrections Psychological Services National Training Event, online.

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2020, September). Predictive validity of the DYRA and SAFVR: New Zealand Police's family violence risk assessment measures. Paper presented at Evidence Based Policing Centre Conference, online.

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2020, July). Predictive validity of the DYRA and SAFVR: New Zealand Police's family violence risk assessment

measures. Paper presented at Waikato Waitaha Forensic Psychology Conference, online.

Jolliffe Simpson, A. D., Moore, L., Joshi, C. & Polaschek, D. L. L. (2019, October). Perpetrators of family harm already serving correctional sentences: Risk profiles and episode characteristics. Paper presented at Department of Corrections Psychological Services National Training Event, Rotorua.

Jolliffe Simpson, A. D., Moore, L., Joshi, C. & Polaschek, D. L. L. (2019, August). Perpetrators of family harm already serving correctional sentences: Risk profiles and episode characteristics. Paper presented at New Zealand Psychological Society Annual Conference, Rotorua.

Jolliffe Simpson, A. D., Moore, L., Joshi, C. & Polaschek, D. L. L. (2019, May). Perpetrators of family harm already serving correctional sentences: Risk profiles and episode characteristics. Poster presented at North American Correctional and Criminal Justice Psychology Conference, Halifax, Canada.

Chapter One: Introduction

Each time a Family Violence (FV) related death is reported, Aotearoa New Zealand engages in collective soul-searching. We ask ourselves: What happened? Was the family receiving any support? And—most importantly—could we have foreseen this outcome and intervened before it was too late? Between 2009 and 2018 there were 125 deaths attributed to Intimate Partner Violence (IPV; New Zealand Family Violence Death Review Committee, 2021a), and a further 70 deaths due to child maltreatment (New Zealand Family Violence Death Review Committee, 2021b). In hindsight, reports reveal many of those cases had previous contact with agencies responsible for intervening and preventing FV, causing those agencies to look inward and search for ways to improve their decision making and prevent further harm and loss of life.

To that end, completing risk assessments has become a standard procedure for FV practitioners who are frequently required to make decisions that have potential ramifications for the safety and wellbeing of families or whānau[1]. During these assessments, information that could indicate an aggressor is likely to commit FV again in the future, or suggest a family is at greater risk of experiencing further harm, is used to decide whether and how practitioners should intervene (Henning et al., 2021; Kropp, 2004). Risk assessment is, then, a crucial aspect of responding to FV cases; it is important we get it right, but little research exists for the risk assessment instruments used by agencies that respond to FV in New Zealand. Accordingly, in this thesis, we will examine the risk assessment procedures that New Zealand Police and the Integrated Safety Response (ISR) use for FV, with the purpose

[1] Whānau are a social unit similar to family or extended family in Te Reo Māori, the Māori language.

of contributing to the evidence base about risk assessment for FV in New Zealand and identifying strengths, weaknesses, and opportunities for improvement in those procedures.

Definition of FV

FV is a pervasive social problem that has been known by many other names, including wife abuse, battering, and domestic violence (Kropp, 2004). In recent decades the collective understanding of FV has broadened from exclusively focusing on men's violence against their wives and girlfriends to include other types of FV such as child maltreatment, elder abuse, sibling violence, and violence perpetrated by children against their parents. This change recognizes that people commit harm in a variety of family relationships and when one form of harm occurs within a family or whānau, it is often accompanied by another (Chan et al., 2021; Dixon et al., 2007; Ryan & Roman, 2021).

The definition of FV now also recognizes non-physical forms of harm, such as psychological abuse and controlling behavior (Fanslow, Malihi, et al., 2021; Robinson et al., 2018; Ryan & Roman, 2021). In New Zealand, FV is defined in the Family Violence Act (2018, § 9) as physical, sexual, or psychological violence inflicted against a person 'whom [the aggressor] is, or has been, in a family relationship with' (Family Violence Act 2018, § 9). In New Zealand, understandings of FV are further broadened by the Māori concept of whakapapa that refers to genealogical links traversing generations and recognizes transgressions against whakapapa and tikanga (customs and traditional values) as further forms of harm (Dobbs & Eruera, 2014).

Prevalence of FV

It is difficult to measure the prevalence of FV because of its varying definitions and its hidden nature. In New Zealand, responding to FV accounts for more than two-fifths of frontline police officers' time (New Zealand Family Violence Clearinghouse, 2017). Police conducted 165,039 FV investigations in 2020 (New Zealand Police, 2022); but two-thirds of

FV investigations did not involve criminal offenses (i.e., represented "argument-only" calls for service), which highlights how prevalence estimates may vary widely according to the definition of FV applied (New Zealand Police, 2022). Moreover, an estimated 80% of FV episodes are not reported to police, indicating the true problem is much larger and victimization surveys may provide a more complete picture than official statistics (e.g., New Zealand Crime and Victims Survey, 2018).

Research on FV prevalence usually focuses on measuring one sub-type (e.g., IPV or child maltreatment). For example, Devries and colleagues (2013) estimated that in 2010, 30% of women aged over 15 across the globe experienced physical or sexual IPV. Consistent with this finding, a recent survey found around 30% of New Zealand women experienced physical IPV in their lifetime, 13.1% experienced sexual IPV, and a third experienced at least two acts of psychological IPV (Fanslow, Hashemi, et al., 2021; Fanslow, Malihi, et al., 2021). Research using community samples shows men experience similar rates of IPV to women, but women experience more severe and fatal IPV (Holtzworth-Munroe, 2005). Studies of self-reported IPV aggression produce similar estimates to victimization surveys, but with slightly higher rates reported by women than men (Desmarais et al., 2012).

On average, physical child maltreatment was estimated to affect 22.6% of children across the globe, and sexual maltreatment was estimated to affect 12.7% (Stoltenborgh et al., 2015) with the rate of physical or sexual maltreatment reaching one in four among children with disabilities (Jones et al., 2012). Furthermore, it was estimated that 36.3% of children experience emotional maltreatment, 16.3% experience physical neglect and 18.4% experience emotional neglect (Stoltenborgh et al., 2015). Like other sub-types of FV, child maltreatment rates differ widely according to the type of abuse measured, country, data source, and gender (Moody et al., 2018; Stoltenborgh et al., 2015).

Research has also investigated the prevalence of other types of FV. For example, Tucker and colleagues (2013) found about 36.7% of children experienced sibling violence in the preceding year, with higher rates observed among brothers. Estimates of the prevalence of elder abuse also vary widely, ranging between 3.2-27.5% (Cooper et al., 2008). In addition, child to parent physical FV is estimated to occur in 10% of families (Arias-Rivera et al., 2020), and aggression from adolescents towards their parents represents 1-7% of FV episodes reported to police in Australia (Moulds et al., 2019). While these sub-types of FV are often examined in isolation, a global meta-analysis found rates of experiencing more than one type of FV (i.e., poly-victimization) ranged between 9-36%, indicating different sub-types often co-occur and research may benefit from examining FV at the level of the family unit instead (Chan et al., 2021; Dixon et al., 2007).

Recent global events have highlighted how contextual factors also complicate the process of estimating FV prevalence. FV reportedly increased during the COVID-19 pandemic as family members experienced isolation, stress, and spent more time together (Boxall et al., 2020). Indeed, one in ten people in a representative survey of New Zealanders reported experiencing some form of FV during the nation's first lockdown in 2020 (Every-Palmer et al., 2020). The prevalence rates listed in this section are difficult to unify but together indicate that FV is a common and wide-ranging social problem, warranting further attention from researchers and practitioners alike.

Responding to FV

The conceptualization of FV as a social problem and its treatment by agencies has transformed rapidly over the last 50 years. It was not until 1982 that violence within families was specifically recognized in New Zealand legislation, with the Domestic Protection Act (1982) introducing non-violence, non-molestation, occupancy, and tenancy orders to protect victims of domestic violence (victims who, at this time, were understood to be women

assaulted by their husbands). Since then, responses to FV have varied in their approaches and effectiveness; for example, the 1980s saw the introduction of pro-arrest policies, where Police were encouraged to arrest the predominant aggressor[2] at FV call outs (Hamel, 2011), which lead to unintended consequences including the arrest of victims using self-defense.

Historically, FV responses were fragmented because agencies had different goals and mandates (Dobbs & Eruera, 2014; Little & Kantor, 2002). For example, the main role of Police was to detect and prosecute criminal offenses, whereas Women's Refuge provided haven for women leaving violent relationships and their children, and Tuu Oho Mai Services (formerly the Hamilton Abuse Intervention Project) provided non-violence programs to people named as aggressors on protection orders or who self-referred themselves for treatment. With these diverging missions, agencies could not provide comprehensive, 'wrap-around' services; instead, families receive different types and levels of intervention, depending on where and how they entered the system (Dobbs & Eruera, 2014).

Multi-agency responses emerged in response to this fragmentation, such as the Multi Agency Risk Assessment Conferences operating in the United Kingdom (Hague & Bridge, 2008; Robbins et al., 2014) and Multi Agency Risk Assessment Committees in Canada (Contini & Wilson, 2019). These responses included representatives from local organizations (e.g., Police, Health, refuges) responsible for responding to FV, and were created to improve collaboration by facilitating information sharing, streamlining service provision, and targeting resources at victims at the highest risk of harm (Robinson & Clancy, 2020; Robinson & Tregidga, 2007; Sloper, 2004). Research suggests multi-agency responses have improved service provision and victim safety (Hague & Bridge, 2008; Mossman et al., 2017).

[2] In unclear situations, the predominant aggressor is often selected based on the parties' physical capability to cause harm, contributing to a gender bias towards the male partner in heterosexual relationships (Hamel, 2011).

The Integrated Safety Response

The Integrated Safety Response (ISR) is a multi-agency FV response operating in Waikato and parts of Canterbury in New Zealand (Integrated Safety Response, 2019). The ISR includes representatives from organizations such as New Zealand Police, Ara Poutama Aotearoa (Department of Corrections), local District Health Boards, Oranga Tamariki (Ministry for Children), Accident Compensation Corporation, and non-governmental organizations (Mossman et al., 2017). Contrary to the approach taken by other multi-agency responses (e.g., MARACs; Robbins et al., 2014), where only suspected high-risk cases are assessed, all FV episodes reported to police in the ISR's catchment areas are referred to the ISR's triage team for risk assessment and case management.

ISR triage teams, which contain representatives from ISR's partner organizations, meet daily for safety assessment meetings (Integrated Safety Response, 2019; Mossman et al., 2017). During these meetings, the triage team reads the police report for the episode in question; shares relevant information about the family from their organization's database (often leading to a more complete picture of the family' situation); collectively chooses a risk category (low, medium, or high), records factors to support that categorization, and selects the appropiate services for the family based on those factors. Each episode is usually assessed at a safety assessment meeting within 72 hours of being reported to police (Integrated Safety Response, 2018, 2019) and risk categories are updated following new episodes or successful interventions.

Around half of cases are rated as low risk by ISR triage teams (Mossman et al., 2017), and receive low-level intervention in the form of a phone call from a non-governmental agency to offer support (e.g., Lifeline; Integrated Safety Response, 2019). The ISR takes on increased responsibility for victim safety and provides a greater number and variety of services to cases with higher risk categorizations (Integrated Safety Response, 2018; New

Zealand Police, n.d.-a)[3]. Approximately 45% of cases receive medium risk categorizations, with the remaining ~5% being labelled as high risk (Mossman et al., 2017). The predictive validity of the ISR triage teams' risk assessment has not previously been examined. A broad range of services may be offered to these higher risk cases including alcohol and drug counselling, whānau support workers, safe housing, healthcare, counselling, budgeting help, and legal assistance, among other services. In general, a greater number of services are offered to, and accepted by, victims than aggressors (Integrated Safety Response, 2019).

For cases that are assessed as high risk, the ISR undertakes further weekly intensive case management meetings attended by additional representatives, rather than the core triage team that attends safety assessment meetings (Mossman et al., 2017). At intensive case management meetings, representatives discuss and review new and existing high-risk plans. They may assess engagement with services, assign new services, follow up on information gathering, update the case's risk categorization where appropriate (Integrated Safety Response, 2019). These assessments exemplify how the decision making and actions of organizations responsible for managing FV cases are increasingly guided by empirical research (Hoyle, 2008).

The Ecological Model

Many of the risk factors included in FV risk assessment instruments are drawn from theoretical explanations of the causes for FV, and when risk factors are atheoretically selected based on their statistical relationship with recurrence, researchers use theory to explain *why* those factors are linked to recurrence. FV is rarely caused by one factor, or explained by a single theory; instead, episodes occur due to the interaction of people,

[3] According to New Zealand Police (n.d.-a) in cases with higher risk levels victims are less able to be responsible for their safety, therefore Police and other agencies must help protect those victims.

stressors, and the environment (Whiting et al., 2020). For example, people who harm their family members may be influenced by their emotions and cognition, but those individual factors intersect with relationship dynamics in the family, stressors, cultural norms, and economic conditions (Whiting et al., 2020). Therefore, rather than discussing single-factor theories for FV, in this section we will instead briefly outline various factors that contribute to FV using the ecological model as a framework (Belsky, 1980; Dutton, 2006).

Bronfenbrenner's (1981) ecological model describes how human development and behavior are affected by factors in different nested 'levels' of the environment. The ecological model organized the factors affecting human development and behavior into four nested environments: the ontogenic level, microsystem, exosystem, and macrosystem (Belsky, 1980; Bronfenbrenner, 1981). Researchers have since applied the ecological model to various phenomena, including child maltreatment (Belsky, 1980; Krishnan & Morrison, 1995), IPV (Carlson, 1984; Dutton, 1995, 2006; Smith Slep et al., 2014), elder abuse (Schiamberg & Gans, 1999, 2000), and the co-occurrence of child maltreatment and IPV (Little & Kantor, 2002; Ryan & Roman, 2021). Factors exist at different levels of the model, which are nested within one another, and factors from the different levels interact as people interact with their environment. Consequently, the ecological model puts forward that risk factors should be considered together, rather than in isolation, when assessing risk for FV cases (Belsky, 1980; Dutton, 2006).

Table 1 shows a summary of some risk factors for FV from meta-analyses on IPV and child maltreatment samples[4]. The ontogenic level of the nested ecological model—also known as the individual level—relates to individual characteristics that predispose people to FV and shape their responses to factors at the other levels. The table displays some of the risk

[4] Meta-analyses were not available for other types of FV (e.g., elder abuse or sibling violence).

factors for using FV measured at the ontogenic level, such as age, alcohol use, drug use, and education; as well as experiencing, witnessing, or using FV (Belsky, 1980; Dutton, 2006; Mallory et al., 2016; Stith et al., 2009; Stith et al., 2004).

The microsystem—or the family level—describes the interactional patterns within the family unit that precipitate and maintain FV (Belsky, 1980). Table 1 shows risk factors for using FV measured at this level include controlling behavior, employment, jealousy, relationship satisfaction, and stress (Belsky, 1980; Dutton, 2006; Mallory et al., 2016; Stith et al., 2009; Stith et al., 2004). Next, the exosystem—or community level—describes factors that influence families and stimulate conflict by creating stress (Belsky, 1980; Dutton, 2006). Table 1 shows some of the risk factors for using FV at the exosystem level, including socio-economic status and social support (Belsky, 1980; Dutton, 2006; Stith et al., 2009).

The macrosystem is the societal level, which describes characteristics of the cultural fabric of society that endorse or help trigger FV (Belsky, 1980; Weeks & Leblanc, 2011). Examples of variables that could be measured at the macrosystem level are societal attitudes towards violence (including corporal punishment), and women's socioeconomic and political power (Dutton, 2006). Table 1 does not include any risk factors from the macrosystem; factors at this level are typically excluded from risk assessments because they are far removed from the individual and are difficult to change.

Table 1

Summary of Risk Factors for FV from Meta-Analytic Research

Risk factor	Ecological level	Sub-type	*r*	Reference
Age	ONT	IPV	-.13	(Stith et al., 2004)
		IPV	-.06	(Mallory et al., 2016)
Alcohol use	ONT	IPV	.24	(Stith et al., 2004)
		IPV	.23	(Mallory et al., 2016)
		IPV	.19	(Foran & O'Leary, 2008)
Controlling behavior	MIC	IPV	.36	(Mallory et al., 2016)
Drug use	ONT	Child maltreatment	.08	(Stith et al., 2009)
		IPV	.31	(Stith et al., 2004)
		IPV	.29	(Mallory et al., 2016)
Education	ONT	IPV	-.13	(Stith et al., 2004)
		IPV	-.15	(Mallory et al., 2016)
Employment	MIC	Child maltreatment	-.15	(Stith et al., 2009)
		IPV	-.10	(Stith et al., 2004)
		IPV	-.05	(Mallory et al., 2016)
History of experiencing FV	ONT	Child maltreatment	.21	(Stith et al., 2009)
		IPV	.23	(Mallory et al., 2016)
History of witnessing FV	ONT	IPV	.29	(Mallory et al., 2016)
History of using FV	ONT	IPV	.24	(Stith et al., 2004)
Jealousy	MIC	IPV	.17	(Stith et al., 2004)
Relationship satisfaction	MIC	Child maltreatment	.16	(Stith et al., 2009)
		IPV	.27	(Stith et al., 2008)
		IPV	-.17	(Mallory et al., 2016)
Social support	EXO	Child maltreatment	-.18	(Stith et al., 2009)
Socio-economic status	EXO	Child maltreatment	.14	(Stith et al., 2009)
Stress	MIC	Child maltreatment	.19	(Stith et al., 2009)
		IPV	.26	(Stith et al., 2004)

Note. ONT = ontogenic, MIC = microsystem, EXO = exosystem.

The RNR Model

The Risk Need Responsivity (RNR) model is an influential framework of empirically derived principles that inform many aspects of correctional treatment, including risk assessment (Bonta & Andrews, 2016). The model's risk principle explains that criminal behavior can be predicted, and that treatment intensity should match the level of risk a given offender poses. Under this view, a person at high risk of reoffending should receive high intensity treatment, whereas a person at low risk of reoffending should receive low intensity treatment, or in some cases, no treatment (Bonta & Andrews, 2016; Polaschek, 2012).

The need principle outlines the most well-established risk factors for criminal behavior (i.e., the central eight; Bonta & Andrews, 2016), and explains that treatment targets should focus on the changeable characteristics currently contributing to a person's offending (e.g., substance abuse; Bonta & Andrews, 2016). The central eight risk factors include a history of antisocial behavior, antisocial personality pattern, antisocial cognitions, antisocial associates, education/employment, family/marital, leisure/recreation, and substance abuse (Bonta & Andrews, 2016); many of which clearly overlap with the risk factors identified in Table 1. Risk may be static (i.e., fixed), measured by risk factors that cannot change once they have occurred (e.g., age at first offense). On the other hand, risk can also be dynamic (i.e., changeable), measured by risk factors that fluctuate with context or vary over time (e.g., employment status; Bonta & Andrews, 2016; Douglas & Skeem, 2005). Some dynamic risk factors tend to change slowly and are considered 'stable', whereas other dynamic risk factors change rapidly and are described as 'acute'.

The RNR model's responsivity principle states that practitioners should deliver treatment services in ways that maximize people's ability to engage and benefit from them (Bonta & Andrews, 2016). Responsivity is comprised of two components. General responsivity outlines the overarching principles of effective treatment delivery; whereas

specific responsivity captures the individual factors that could impede treatment engagement (e.g., motivation to receive treatment, cognitive ability, personality, and maturity; Bonta & Andrews, 2016). Where possible, practitioners should try to accommodate specific responsivity factors to maximize behavior change (Eisenberg et al., 2022).

The RNR model's risk and need principles have greatly influenced the evolution of correctional risk assessment practices. Risk assessment is systematic speculation about future acts a person might commit, with the goal of informing risk reduction and management (Bonta & Andrews, 2016; Hilton & Ennis, 2020). It is a pragmatic task; higher-risk offenders typically commit more harm than lower risk cases and receive greater benefit from interventions, therefore limited resources should be targeted to higher risk cases to achieve maximum impact (Bonta & Andrews, 2016; Robinson & Clancy, 2020).

Predictive Accuracy

It is important to examine risk assessment instruments' predictive accuracy because they are widely used throughout the criminal justice system used to inform dispositions ranging from arrest or police safety orders[5], to imprisonment and parole that can have considerable impacts on peoples' lives (Bonta & Andrews, 2016). Moreover, practitioners responding to FV use risk assessment instruments to help them make decisions for families' safety and wellbeing (Kropp, 2004, 2008). Therefore, the risk factors and algorithms used to generate these risk categories should be transparent so they can be empirically evaluated and improved (Bonta & Andrews, 2016; Kropp, 2004).

Predictive accuracy is measured in research using a variety of statistics. One way to determine the predictive validity of a risk assessment instrument is calculating the Area

[5] A police safety order is an order issued by police to separate the parties for several days after an episode (New Zealand Police, n.d.-b).

Under the Curve (AUC). The AUC represents the probability that a randomly selected person with a recurrence would have a higher risk classification than a randomly selected person without a recurrence (Helmus & Babchishin, 2016). AUCs can range from 0-1; an AUC of .5 indicates predictive ability similar to chance, and values closer to 1 indicate better accuracy. Criteria from Rice and Harris (2005) explain how to interpret the size of AUCs; .56-.63 = small, .64-.70 = moderate, .71-1.0 = large.

In addition, predictions can be categorized as true positive (i.e., a correct prediction that there would be a recurrence), false positive (i.e., an incorrect prediction that there would be a recurrence), true negative (i.e., a correct prediction that there would be no recurrence), or false negative (i.e., an incorrect prediction that there would be no recurrence), and accuracy is a trade-off between maximizing true and minimizing false predictions. If an instrument predicts every person will have a recurrence it will capture all people who reoffend but do so at the expense of imposing dispositions on people who never would have reoffended. Therefore, it is important that we get the balance right; a balance that depends on weighing the costs of incorrectly placing dispositions on a person versus incorrectly choosing not to, and failing to prevent harm (Bonta & Andrews, 2016). Because the consequences of predicting there *will* be further harm are arguably lower than the consequences of incorrectly predicting there *will not* be further harm, many risk assessment instruments have a high rate of false positives. Resultantly, a significant number of people who would not have gone on to commit further harm receive interventions they do not need (Bland & Ariel, 2015; Bland, 2020; Heckert & Gondolf, 2005).

Unstructured Clinical Judgement

The three main approaches to risk assessment differ in their predictive accuracy and according to how much they incorporate professional experience. The earliest approach to risk assessment is unstructured clinical judgement, whereby practitioners sort and evaluate

case information, using their professional experience to determine what information is most important, before forming an intuitive impression of the likelihood of future harm (Hilton et al., 2006). This approach to risk assessment attends to psychological mechanisms of criminal behavior and is favored by practitioners who aim to be client-centered (Hilton & Eke, 2017).

However, unstructured clinical judgment has long been demonstrated as an unreliable method for risk assessment, with AUCs often indicating predictive validity akin to a coin flip (Grove & Meehl, 1996; Hilton et al., 2006). Practitioners rely on heuristics and assumptions to make these judgments (Kahneman & Klein, 2009), and are affected by the demands of high workloads and limited resources (De Bortoli et al., 2017). In addition, unstructured clinical judgements are not transparent, so it is difficult to measure where and how problems occur during the risk assessment process (Hilton et al., 2006). Yet despite the poor predictive validity of unstructured clinical judgement risk assessments, many practitioners continue to rely on their intuition, and sometimes even override the results from risk assessment instruments (Cohen et al., 2020; Hilton et al., 2006).

Actuarial Risk Assessment

On the other hand, the development of actuarial risk assessment method demonstrably improved the ability to predict recidivism (Hanson et al., 2007). Actuarial risk assessment involves selecting items to include in an instrument by atheoretically adding and removing items based on their statistical relationship with the outcome of interest (Henning et al., 2021). Under this approach, items used as risk factors do not need to have a clear causal relationship with the outcome, but instead may act as proxy variables for another unmeasured construct directly associated with FV (Hilton et al., 2006). For example, the number of offenses a person has committed in the last year may not directly cause them to offend again but may indicate the person has an antisocial lifestyle and is therefore more likely to reoffend (Bonta & Andrews, 2016; Douglas & Skeem, 2005).

Widely used actuarial risk assessment instruments in the domain of FV include the Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2010), Danger Assessment (DA; Messing et al., 2017), and the Domestic Violence Screening Instrument (DVSI; Williams & Grant, 2006). Meta analyses of the predictive validity of FV risk assessment instruments and found they had AUCs greater than chance, indicating these instruments can improve our ability to estimate whether an aggressor will harm their family members again when administered properly (Messing & Thaller, 2013; van der Put et al., 2019). As well as the risk assessment instruments available for use, many law enforcement agencies have developed in-house risk assessment instruments. These instruments are pragmatic; making use of information already routinely collected by police and avoiding the extra data collection needed to complete some instruments (Henning et al., 2021). For example, police and researchers in Australia developed the Victoria Police Screening Assessment for Family Violence Risk (VP-SAFvR; Spivak et al., 2020), and the Family Violence Risk Assessment Tool (FVRAT; Dowling & Morgan, 2019).

In 2018, New Zealand Police developed the Static Assessment of Family Violence Recidivism (SAFVR; Bissielo & Knight, 2016), and the Dynamic Risk Assessment (DYRA), to create in-house instruments to replace the ODARA, which they had used since 2012 (Mossman, 2014). The DYRA uses 16 items in the form of questions police ask victims at the scene of FV episodes (e.g., is the aggressor regularly abusing alcohol or drugs?) to inform police's three day safety plan. The predictive ability of the DRYA has not previously been examined. In contrast, the SAFVR uses eight static items (e.g., age at first FV offense, gender, number of previous offenses) to predict whether an aggressor will commit a FV-related offense within 2 years, and has an AUC of .77, indicating good predictive ability (Bissielo & Knight, 2016).

Research consistently shows actuarial risk assessment is the most accurate form of risk assessment (Hilton et al., 2006; Messing & Thaller, 2013). But while this type of risk assessment is transparent and has good predictive ability, questions remain about whether it should be used (Bland, 2020; Prins & Reich, 2018; Woldgabreal et al., 2020). Just as individual practitioners' assumptions affect unstructured clinical judgement, biases may remain embedded in those instruments, meaning they can maintain unfairness and racism (Woldgabreal et al., 2020).

Structured Professional Judgement

The Structured Professional Judgement (SPJ) approach to risk assessment emerged to maintain the improved predictive ability achieved by actuarial methods while offering practitioners the flexibility to include case-specific information (Garrington & Boer, 2020; Hart et al., 2016). In general, SPJ instruments include a list of empirically derived risk factors, but practitioners can include other information outside the instruments' list of items and decide themselves how that information should be combined when they select a categorical label to communicate the level of risk present in the case (e.g., low, moderate, or high; Hilton et al., 2006). When they are used actuarially, the predictive validity of SPJ risk assessment instruments approaches that of actuarial instruments (van der Put et al., 2019). But when the risk categories from SPJ instruments are used to predict outcomes, they tend to have poorer predictive validity, which could be due in part to categorical labels having less variance than raw scores (Helmus & Babchishin, 2016; van der Put et al., 2019).

More recent advances in guidelines for SPJ assessments mean they are increasingly used to manage rather than predict risk (De Bortoli et al., 2017). Newer instruments include case formulation and risk scenarios that prompt practitioners to identify how risk factors interact and consider potential causal mechanisms for violence (Hilton & Ennis, 2020). For example, the Historical, Clinical, Risk Management, Version 3 (HCR-20V3; Douglas et al.

2013) includes items that practitioners score based on their presence and *relevance*, determined according to how each item has contributed previous violence and could potentially contribute to future violence (Hilton & Ennis, 2020).

SPJ has been promoted as an especially suitable risk assessment method for FV due to its idiographic approach that can accommodate the large number of heterogeneous risk factors for families experiencing FV (Garrington & Boer, 2020; Smith Slep & O'Leary, 2001). SPJ risk assessment instruments developed for IPV include the Spousal Assault Risk Assessment Guide (SARA; Kropp et al., 1994) and the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER; Kropp et al., 2005); both contain specific items that practitioners score and combine, with higher scores indicting a greater likelihood of future violence (Garrington & Boer, 2020). But, in the fields of FV and IPV, the term SPJ is also often used to describe approaches that more closely resemble unstructured clinical judgement than SPJ (e.g., DASH; Richards, 2009; Turner et al., 2019). These approaches loosely guide decision-making and are not consistent with how SPJ is used for general (e.g., HCR-20V3; Douglas et al. 2013) or sexual offending (e.g., Risk for Sexual Violence Protocol Version 2; RSVP V2; Boer et al., 2017).

Limitations of Risk Assessment for FV

In this chapter we described the problem of FV, provided some background about how practitioners respond to FV, and positioned risk assessment as a crucial part of that response. But, despite considerable developments in risk assessment research and practice, risk assessment instruments have limitations that curb their value to FV practice. For example, most research about FV risk assessment instruments focuses on one type of FV (e.g., IPV) even though different types of FV often co-occur (Chan et al., 2021; Dixon et al., 2007; McEwan et al., 2018; Saxton et al., 2022). In addition, when researchers examine the predictive validity of risk assessment instruments, they often use data generated for research

rather than practice, hence results may not reflect how instruments truly perform on the front line (van der Put et al., 2019). Researchers also often examine FV instruments' predictive validity for a single dichotomous outcome (e.g., reoffending; Heckert & Gondolf, 2005); which may not capture predictive validity for forms of FV that are not criminalized (e.g., psychological harm; Ansara & Hindin, 2011; Dichter et al., 2018; Wiener, 2017), or account for variation in aggressor behavior over time (Bland & Ariel, 2015; Heckert & Gondolf, 2005; Jones et al., 2010; Piquero et al., 2006; Swartout et al., 2012). Finally, there is little evidence about the validity of risk assessment instruments used by agencies that respond to FV in New Zealand; the predictive validity of New Zealand Police's DYRA risk assessment instrument and the ISR triage teams' risk assessments remains unknown.

This Thesis

Due to limitations in existing research about FV risk assessment, there is a clear need for research that includes all types of relationships and forms of harm within the wide-ranging definition of FV, which produces results that are relevant to FV practice. Accordingly, in this thesis, we[6] examine the risk assessments that New Zealand Police and the ISR use to guide their decision making when responding to FV cases in Waikato and parts of Canterbury in New Zealand. Then, we explore the utility of the ISR's risk assessment for treatment and case management and model the behavioral patterns of FV aggressors over a two-year period. The purpose of this research is to contribute to the evidence base about risk assessment for FV aggressors in Aotearoa New Zealand; and in doing so, identify strengths, weaknesses, and opportunities for improvement. Two overarching research questions guided the research presented in this thesis:

[6] Throughout this thesis I use the terms 'we' and 'our' to recognize that although the thesis is my own work, the research was conducted in collaboration with my supervisors.

1. What are the current risk assessment processes used for FV in Aotearoa New Zealand?

2. How can we improve risk assessment for FV in Aotearoa New Zealand?

The first chapter of this thesis overviewed the issue of FV, including its definitions and prevalence; and the practice of risk assessment, including its strengths and limitations. Chapter two includes manuscript one, which was published in Criminal Justice and Behavior and is reproduced in this thesis with permission from SAGE. In this first manuscript we examined the predictive validity of New Zealand Police's actuarial risk assessment instruments for predicting recurrence (i.e., a further FV-related call for police service) within 2 days, 12 weeks, and 24 weeks. This manuscript established foundational knowledge about FV risk assessment and case management in New Zealand; Police's risk assessments are the basis for their initial response to FV episodes, which are then referred to ISR triage teams for consideration in their subsequent assessment and service allocation.

Next, chapter three includes manuscript two, which was published in the Journal of Interpersonal Violence and is reproduced in this thesis with permission from SAGE. In manuscript two we turned our attention to the next phase of the FV response and explored the risk assessments that the ISR used to triage cases and allocate intervention services. ISR triage teams used a relatively unstructured SPJ risk assessment approach. Therefore, in this manuscript we documented the factors ISR triage teams recorded during their risk assessments and examined whether those factors were associated with the risk categories triage teams allocated to cases; before testing the association of the factors and risk categories with recurrence and physical recurrence within 24 weeks.

The results from manuscript two lead us to wonder whether some of the factors ISR triage teams recorded added value to their risk assessments by highlighting barriers to responsivity (Bonta & Andrews, 2016) and informing case management in addition to (or

rather than) contributing to the assessments' ability to predict recurrence of FV. Therefore, in chapter four, we present additional analyses where we used latent class analysis of the factors ISR triage teams recorded to identify and describe the types of cases the ISR manages and explore the potential value of these descriptions for the triage teams' risk assessment and case management approaches.

In chapter five we present manuscript three, where we modeled the behavioral patterns of 2115 FV aggressors. In this study we describe the latent (i.e., unmeasurable) states behind the information reported to police, and examine aggressors' movement between those states over a two-year period (Jones et al., 2010), with the goal of generating new ideas for approaches to risk assessment. At the time of submission, this manuscript is under review at the Journal of Family Violence. Finally, chapter five synthesizes the findings from this research and details its practical and theoretical implications.

Chapter Two: Manuscript One

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2021). Predictive validity of the DYRA and SAFVR: New Zealand Police's family violence risk assessment instruments. *Criminal Justice and Behavior, 48*(10), 1487-1508. https://doi.org/10.1177/0093854821997525

Abstract

Rapid access to accurate risk assessment information is essential for effective police responses to family violence (FV) calls for service. This study describes the predictive validity of the Dynamic Risk Assessment for family violence (DYRA) and Static Assessment of Family Violence Recidivism (SAFVR), currently in use by the New Zealand Police. We used 1,817 police reports of FV episodes to predict recurrence (i.e., repeat call for police service) over three follow-up periods. Regardless of follow-up, the DYRA and SAFVR each displayed poor ability to discriminate between episodes with and without a recurrence. Both instruments substantially over-predicted recurrence, and performed relatively consistently across subsamples (e.g., intimate partners vs. other family relationship; aggressor gender, ethnicity, age). The especially poor performance of the DYRA suggests further research on dynamic risk factors and their contribution to police responses for FV is needed to make these instruments more useful for agencies working with families.

Key words: family violence; police risk assessment; actuarial risk assessment

Predictive Validity of the DYRA and SAFVR: New Zealand Police's Family Violence Risk Assessment Instruments

Family violence (FV) causes billions of dollars of societal harm per year, even in a small country like New Zealand (Kahui & Snively, 2014). But this harm is distributed very unevenly; more than half is experienced by 1% of adults (Adams, 2016). In other words, exposure to FV varies widely, not only in the severity of individual episodes, but in the likelihood of repeated episodes, and in the extent to which further exposure may be mitigated by legal and social responses. Only some FV episodes result in calls for service to the police; about 1 in 3 victims had an episode reported to police in the latest nationwide crime survey (New Zealand Crime and Victims Survey, 2018). And even though most episodes are not reported, police still spend an estimated 41% of frontline time responding to those incidents that are reported (New Zealand Family Violence Clearinghouse, 2017). The volume of calls and the demand they place on services make it an imperative for law enforcement to be able assess risk rapidly and accurately prioritize directing limited resources to where they can be most effective; where the level of risk is the highest. In this paper we report on one aspect of predictive validity for repeat FV perpetration—discriminative ability—for two risk assessment instruments currently used on the frontline by New Zealand police.

Definition of FV

The definition of FV in the New Zealand Family Violence Act 2018 is broad, including physical, sexual or psychological abuse inflicted against a person by any other person with whom that person is, or has been, in a family relationship (Family Violence Act 2018, § 9). The legislation further specifies that a family relationship is one between partners, family members, or people who ordinarily share a household. But family relationships can also include close personal relationships, based on factors that include the amount of time spent together, how shared time is spent, and the duration of the relationship. Such

relationships do not have to include shared sexual activity and may be between carers and care recipients. Applying this definition of FV leads to the inclusion of a wider range of episodes involving a wider range of people than those in many previous research populations (Williams, 2012); which were often limited to men prosecuted for intimate partner violence (IPV; defined as the physical, sexual or psychological abuse of a current or former intimate partner). In New Zealand, most FV-related calls for service do not result in charges being laid; approximately two-thirds are "argument-only" calls for service, where police do not report detecting the use of physical violence. Moreover, approximately one-third of episodes are not any type of IPV. They involve conflict between siblings, parents and children, or people in other familial relationships. These differences in context give us an opportunity to use New Zealand data to extend earlier research, by shedding light on the use of FV risk assessment instruments with the broad range of episodes police attend, some of which may be relatively minor in nature.

Research on Risk Assessment Instruments for FV

Risk assessment instruments vary along a number of dimensions, including the types of FV they are designed to predict, the types of items used, the outcome criteria they are validated against, and the practical purpose for which they were designed. Most risk assessment instruments are created for predicting and preventing repeat IPV (e.g., Spousal Assault Risk Assessment; Kropp & Hart, 2000). Meta-analyses show that using these instruments actuarially, most have small to moderate AUCs in predicting IPV recidivism (.54-.67; Messing & Thaller, 2013), although AUCs can vary widely (e.g., .46-.87; van der Put et al., 2019). Recidivism in these studies typically refers to the recurrence of criminal behaviors, and can be measured in different ways, including victim reports of repeated violence, arrests, convictions, or even death (Graham et al, 2019).

Besides IPV, the main other type of FV for which specific instruments have been constructed is child maltreatment (e.g., Child Abuse Risk Assessment Scale; Chan, 2012). A meta-analysis of 30 studies found that on average actuarial instruments created for child maltreatment had moderate AUCs (.70; van der Put et al., 2017). Other specific types of FV are sufficiently uncommon in comparison to child maltreatment and IPV that specialized risk assessment instruments been rarely developed for them.

It is not clear yet whether the predictive accuracy of IPV and child maltreatment risk assessment instruments extends to instruments created for FV in the broader sense, because research using a single instrument to predict any type of FV recurrence is still in its infancy, and only a small number of these instruments exist (e.g., VP-SAFvR; Spivak, et al., 2020). From a practical perspective, instruments tailored to specific types of FV may be valuable for those assessing and providing treatment services for FV. But, first responders are often confronted with complex and competing challenges at a scene that may itself include multiple aggressors, and multiple victims or "at risk" family members. A single instrument that can be used for rapid triage to determine immediate responses is ideal in this context.

The relative dearth of broader FV risk assessment instruments makes it difficult to situate research findings using such instruments among the siloed bodies of research on IPV and child maltreatment (McEwan et al., 2018). Variables in risk assessment instruments for FV or IPV include criminal history, alcohol or drug use, mental health, suicidal behavior, pregnancy, recent separation, and financial stress (Dowling & Morgan, 2019; Graham et al., 2019; López-Ossorio et al., 2019; Spivak, et al., 2020; Steiner et al., 2019). Some of these indicators of risk are static in nature; they cannot change once they have occurred (e.g., age at first offense), or change only slowly (e.g., age). Others are dynamic, and some of these can change relatively quickly (e.g., current levels of substance use).

Risk assessment instruments for IPV also include factors that are predictive of other types of criminal behavior (Magdol, et al., 1998; Nicholls, et al., 2013; Svalin & Levander, 2019). The overlap of risk factors for general criminal behavior and IPV perpetration suggests that instruments' predictive ability may extend across different types of outcome criteria, including types of crime, and perhaps even types of FV. Indeed, Hanson et al. (2007) found that across four studies predicting spousal re-assault and general recidivism, AUCs remained similar across FV and non-FV outcome criteria (e.g., Hilton & Eke, 2001). Moreover, results from Spivak, et al. (2020) suggest that general FV instruments may perform similarly across FV sub-types (AUC of .64 for IPV c.f., .67 for other types of FV). Taken together these findings suggest that even if there are some risk factors that are uniquely important for IPV relative to other types of FV (e.g., sexual jealousy), it should be possible to develop from common risk factors a triage instrument for use across FV.

Risk assessment instruments may be designed for use by parts of the work force: for example, for health, psychology, corrections, or social support professionals, to guide their decisions during treatment; or for law enforcement, to inform an immediate safety response (Svalin & Levander, 2019). But, until recent years most evaluations of IPV risk assessment instruments were completed using data generated for research rather than from practice (Graham et al., 2019; Messing & Thaller, 2013; van der Put, 2019), and with scores rather than the risk categories (e.g., High, Moderate, Low) that are actually used to inform decisions (Singh et al., 2013). As awareness increases about the need for evidence-based policing practices, risk assessment instruments for use by law enforcement, validated with field data, have begun to emerge. The earliest instrument of this nature was the Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2010), designed to predict future assaults in episodes of IPV involving physical violence. But, because a substantial minority of cases dealt with by police involve different types of FV, and it is not operationally feasible to use

multiple instruments, newer policing risk assessment instruments have been developed for use with FV, more broadly defined. For example, in Australia, there are at least two frontline FV risk assessment instruments: the Victoria Police Screening Assessment for Family Violence Risk (VP-SAFvR; Spivak, et al., 2020), and the Family Violence Risk Assessment Tool (FVRAT; Dowling & Morgan, 2019).

FV Risk Assessment by New Zealand Police

For New Zealand police (NZP), the purpose of FV risk assessment instruments is to triage FV episodes; frontline staff need to know whether or not FV will happen again, and how quickly, in order to determine the level of response or intervention that may be required. NZP has used a variety of risk assessment instruments to identify and prioritize IPV cases by risk level since 2004, and began using the ODARA in 2012 (Grant, 2009). A review of the ODARA's predictive ability in New Zealand (Mossman, 2014) found small to moderate AUCs of .62 - .69 (Rice & Harris, 2005), and in response, NZP embarked on a project to revise their risk assessment instruments (Bissielo & Knight, 2016). The decision was made to develop both a dynamic and a static risk instrument, and to validate these instruments for all types of FV.

The first instrument that arose from this project was the Static Assessment of Family Violence Recidivism (SAFVR), an actuarial instrument designed to predict a FV-related offense within 2 years. The SAFVR was developed from the ODARA validation study, after selecting the most promising predictive variables available in police and criminal history records for a sample of 39,317 people with a FV offense in New Zealand. After developing the SAFVR with 70% of their sample, Bissielo and Knight (2016) validated the instrument with the remaining 30%. The resulting combination of variables yielded a large AUC of .77 (Bissielo & Knight, 2016; Rice & Harris, 2005).

The second new instrument was the Dynamic Risk Assessment (DYRA), another actuarial instrument designed to determine the imminent risk of the aggressor harming their family member(s) and support police in developing a 3-day safety plan. The DYRA is comprised of "yes/no" questions, asked by police at the scene of the episode, about the aggressor and the current circumstances within the family unit. The aggressor is defined as the person causing the most harm or posing the most risk in the FV episode, but their behavior is not necessarily criminal (New Zealand Police, n.d.-a). In instances where police data-gathering at the scene does not result in the clear identification of the aggressor, police make this selection based on previous history, and absent this, they choose the person with the greatest physical capacity to cause harm. As in other jurisdictions, selecting the aggressor is a fraught issue with a gender bias towards the male partner in heterosexual relationships (Hamel, 2011). This method has the advantage of not inadvertently criminalizing self-defensive behavior, particularly by women, but the disadvantage of sometimes erroneously labelling men as aggressors when they are not.

The DYRA was designed to identify imminent risk and guide police-initiated interventions over the 3 days following a particular call for service. It complements the criminal history information in the SAFVR with the inclusion of items that are more psychologically meaningful (Douglas & Skeem, 2005; Mann et al., 2010), and intervention-relevant (e.g., current mental health, behavior toward children in the household, jealousy). Although the DYRA is described as dynamic in nature, three items are worded as static items (e.g., "Has X ever …") in that they cannot change once they have occurred, while other items are more clearly dynamic (e.g., "Is X currently ..."). However, even the static questions may reveal risk information that would otherwise go undetected and thus have the potential to improve prediction over the SAFVR alone. For example, the question "has X ever deliberately harmed any child?" may lead to the disclosure of unreported concerning

behavior indicative of an aggressor's ongoing capacity to harm children. To date, there has been no evaluation of the DYRA's ability to predict FV recurrence (D. Scott, personal communication with third author, 20 August, 2020).

The Current Study

The DYRA's predictive ability has not yet been formally investigated, so a primary aim of this study is to evaluate the discriminative ability of the DYRA for FV using field data (i.e., data from police completing the instrument on the frontline), rather than performing a traditional psychometric validation with data generated specifically for research. We evaluate validity using the outcome criterion the DYRA was designed to predict: any new call for service for FV—including "argument-only" episodes—which we collectively refer to as *recurrence*. We will perform analyses using the time period for which the DYRA was designed to predict recurrence (3 days), but will also examine performance at 12 and 24 weeks. These are intervals of practical utility for those using DYRA information for monitoring and intervening with FV, given that even when interventions are in place, when recurrence does occur, it is often in the following several months after an index episode (Gondolf, 2002). Finally, in field use, responses to the DYRA are automatically categorized for police into three discrete risk categories that serve as the basis for determining further intervention (New Zealand Police, n.d.-a). We use these categories, not individual scores, as the predictor variable here, to reflect how the DYRA is used in practice (Singh et al. 2013).

We will conduct similar analyses with the SAFVR data. In practice, the DYRA and SAFVR are used together to make judgements about risk and service response. But the SAFVR has already been shown to have good predictive ability (Bissielo & Knight, 2016), so it provides a point of comparison for the unevaluated DYRA, as well as a comparison of criminal history-based prediction vs. prediction based on factors from the event and surrounding circumstances. Lastly, we will divide our sample into sub-samples of different

types of FV and FV aggressors. We took this step because most FV risk assessment instruments are created for specific types of FV (e.g., IPV or child maltreatment), and we need to know whether these more generic instruments work equally well for different groups in order to justify their use.

Research Questions

1. How well does the DYRA discriminate between FV aggressors with and without a FV recurrence?
2. How does the discriminative ability of the DYRA compare to that of the SAFVR?
3. To what extent do the discriminative abilities of the DYRA and SAFVR differ across FV sub-samples?

Method

Data Source

This study used archival data from the 'Integrated Safety Response'; a multi-agency FV response pilot with a dedicated database (the Family Safety System; FSS). The Integrated Safety Response is similar to other international FV responses (e.g., Multi Agency Risk Assessment Conferences, in the United Kingdom; SafeLives; n.d.). The initial sample for this project contained 2,115 reports for aggressors with a FV episode reported to police between 1 November and 9 December 2018 in two regions of New Zealand: Waikato and Christchurch. Our main source of information was reports completed by police on an application on their smartphones, usually at the scene of the episode. Police recorded information relating to the episode, such as the presence of children, drugs, alcohol or firearms, and the types of harm used.

Risk Assessment Instruments

DYRA

The DYRA was designed to determine the likelihood of the aggressor harming their family member(s) and informs police's 3-day safety plan. The DYRA is administered by trained frontline police officers. Each DYRA is checked and approved by the supervising Sergeant, and police also do random quality assurance audits which result in updates to training advice for frontline users (D. Scott, personal communication with first author, August 17, 2020). The DYRA contains up to 16 "yes/no" questions asked at the scene of the episode during an interview with the adult victim, or in the case of child victims, an adult family member other than the aggressor (e.g., the mother; New Zealand Police, n.d.-a). The DYRA measures some indicators of risk at the level of the family unit; questions relate to current stressors affecting the family, and the behavior, substance use, and mental health of the aggressor. Ten questions apply to all types of FV (asked in 98.3–99.9% of episodes with a completed DYRA in our sample), two further items are applicable if episodes involve IPV (asked in 67.3–67.4% of episodes), and four items are asked if children normally reside at the address (asked in 52.6% of episodes).

DYRA items with a 'yes' response are summed to create a total score ranging between 0 and 16. A score of 0 or 1 is categorized as low-risk; 2 or 3 is moderate-risk, and 4 or greater is high-risk. There are also three 'flag' items; when endorsed, any one leads to a high-risk rating (New Zealand Police, n.d.-a). Because (a) we could not access scores for the SAFVR and therefore had to use risk categories, and (b) this was a field study of police practice, which is done based on categories for both instruments, for this study we used the DYRA risk category calculated for each aggressor at the index episode. Table 2 (final column) shows that about two fifths of aggressors in the final sample (see Data Preparation section) were rated as high-risk, approximately one quarter were rated as moderate, and the

remainder as low-risk. Of the aggressors rated as high-risk, 103 (13.6%) received high-risk ratings due to flag items alone (i.e., these cases had fewer than the necessary four items that otherwise would lead to a high-risk rating).

SAFVR

The SAFVR is an actuarial instrument that is automatically updated daily with administrative data from police and the Ministry of Justice. The SAFVR was validated to predict whether a FV-related charge will be laid against an aggressor within two years of an episode date (Bissielo & Knight, 2016). The SAFVR includes eight aggressor-based variables: gender, age at first FV offense, presence of a prior conviction, presence of a prior prison sentence of 30 days or more, number of prior offenses of any type, number of prior breaches, number of prior FV episodes, and the presence of a FV episode in the past year (New Zealand Police, n.d.-a). Variables in the SAFVR are combined by a computerized algorithm using weights derived from logistic regression analyses (Bissielo & Knight, 2016). SAFVR scores range between 0 and 100. A score of 0–14.99 is categorized as low-risk; 15–19.99 is moderate-risk, and 20 or greater is high-risk (D. Scott, personal communication with first author, August 17, 2020). Frontline police have access only to the aggressor's categorical SAFVR risk level (High, Moderate, Low, or No score) which are made available to them via a smartphone application.

For this study, we did not have access to the SAFVR scores, only the resulting risk category calculated for each aggressor on the date of the index episode. Table 2 (bottom row) shows that approximately half of aggressors in the final sample (see Data Preparation section) were categorized as high-risk, a further tenth were moderate-risk, while about a fifth were low-risk, and the remaining 281 were in the 'no score' category. We retained people with 'no score' for analysis because for the majority, not having a score indicates a lack of the

necessary criminal history for one to be calculated (D. Scott, personal communication with first author, January 28, 2020).

Procedure

New Zealand Police extracted a large dataset from the FSS and relevant police databases containing information relating to the first episode for each aggressor reported between 1 November to 31 December 2018. Data were exported to Excel, cleaned, and organized. We aimed for a sample size of approximately 2000, so we coded the first 2,115 episodes provided, which covered a five-week period up to December 9. One report was retained for each aggressor; where an aggressor had a subsequent episode during the sampling period, this was counted as recurrence.

Some descriptive information was missing in 4% of cases (n = 80). For the variables *police safety order* issued (i.e., an order issued by police to separate the parties for several days after an episode; New Zealand Police, n.d.-b), protection order breached, alcohol consumed prior, drugs consumed prior, children present (either witnessing, or in the vicinity but not witnessing, the FV episode), pregnancy, and firearms, we read the narrative description of the episode written by police to ascertain the correct value and entered it manually into the database. If the variable was not mentioned in the narrative, we entered a 0. Where more than one relationship type, or none, was coded, we chose the correct type again by reading the narrative description to ascertain the relationship between the aggressor and victim. We did not complete any missing DYRA items because the purpose of this study was to investigate the predictive ability of field ratings.

For each episode, police defined the nature of each person's involvement (e.g., victim, informant, witness, mutual participant, suspect). In most cases (86.6%), we determined who was the aggressor by selecting the individual to whom police had designated the role of 'perpetrator' or 'suspect' in the first instance. In a smaller proportion of cases (13.4%), police

had instead used the role ‘mutual participant’ to indicate that they were unable to determine a single aggressor at the scene. In these cases, we selected the person who had been identified by police as the aggressor in the DYRA.

After we had identified a aggressor in each episode, we requested recurrence information: defined as a subsequent FV call for service where the same individual had the role of perpetrator, suspect, or mutual participant. Recurrence data were examined over three follow-up periods: 3 days, 12 weeks, and 24 weeks (no recurrence = 0, recurrence = 1). The first recurrence per aggressor was used throughout and the same set of individuals was included across all intervals (i.e., a recurrence at 3 days was also counted as recurrence at the remaining follow-up periods; an aggressor whose first recurrence occurred between 12 and 24 weeks was counted as having no recurrence in 3-day and 12-week analyses and having a reccurence in 24-week analyses).

Sample Characteristics

After additional data preparation (see below) the final sample contained 1,817 episodes. Table 1 shows that most aggressors were New Zealand European or Māori. Their average age was 32.7 (*SD* = 12.4, range = 8–93), with 7.6% under 18 years old. Three quarters of aggressors were male. They had an average of 4.0 prior episodes of FV reported to police (in any role; *SD* = 5.1, range = 0–40). In 101 episodes (5.6%), aggressors were noted to have gang affiliations. Few aggressors had a diagnosed mental health condition, but this variable only contained instances where a mental health condition was diagnosed and recorded in police databases and likely underestimates the prevalence of mental illness among people in the sample. Most victims were also New Zealand European or Māori. The average age of victims was 35.0 (SD = 14.0, range = 2–84), and 5.6% of victims were under 18 years old. Three quarters of victims were female.

Table 1

Descriptive Statistics for Aggressors, Victims and Episodes in the Initial Sample (N = 2,115), and Final Sample Used in Analyses (n = 1,817)

Variable	Initial *n*(%)	Final *n*(%)	χ^2	*p*	OR [95% CI]
Aggressor gender					
Male	1599(75.6)	1368(75.3)	0.59[a]	.441[a]	1.12 [0.84-1.50][a]
Female	513(24.3)	446(24.5)			
Gender diverse	3(0.1)	3(0.2)			
Aggressor ethnicity			72.91	< .001	-
Māori	930(44.0)	735(40.5)			
European	929(43.9)	860(47.3)			
Pasifika	63(3.0)	51(2.8)			
Asian	48(2.3)	42(2.3)			
Other	145(6.9)	129(7.1)			
Aggressor under 18 years old	152(7.2)	139(7.6)	4.15	.042	1.82 [1.02-3.25]
Aggressor diagnosed mental health condition	42(2.0)	37(2.0)	0.17	.681	0.82 [0.32-2.11]
Aggressor gang affiliation	129(6.1)	101(5.6)	6.58	.010	0.57 [0.37-0.88]
Relationship to victim			18.31	.001	-
Current intimate partner	1046(49.5)	873(48.0)			
Former intimate partner	420(19.9)	376(20.7)			
Parent or child	393(18.6)	355(19.5)			
Sibling	127(6.0)	108(5.9)			
Other	129(6.1)	105(5.8)			
Victim gender					
Male	476(22.5)	404(22.2)	0.95[a]	.331[a]	1.15 [0.87-1.53][a]
Female	1636(77.4)	1412(77.7)			
Victim ethnicity			77.33	< .001	-
Māori	858(40.6)	675(37.1)			
European	968(45.8)	893(49.1)			
Pasifika	40(1.9)	29(1.6)			
Asian	52(2.5)	48(2.6)			
Other	197(9.3)	249(13.7)			
Victim under 18 years old	133(6.3)	102(5.6)	10.17	.001	0.51 [0.33-0.78]
Alcohol consumption detected at the scene	593(28.0)	497(27.4)	3.00	.083	0.79 [0.61-1.03]
Drugs consumption detected at the scene	200(9.5)	180(9.9)	3.05	.081	1.53 [0.95-2.47]
Children present in the episode	1045(49.4)	918(50.5)	6.40	.011	1.38 [1.07-1.76]
Pregnancy	48(2.3)	36(2.0)	4.83	.028	0.48 [0.25-0.94]
Firearms present at the scene	23(1.1)	22(1.2)	1.82	.177	3.64 [0.49-27.11]

Table 1 Continued

Variable	Initial *n*(%)	Final *n*(%)	χ^2	*p*	OR [95% CI]
Offense detected in the episode	280(13.2)	227(12.5)	6.24	.012	0.66 [0.48-0.92]
Police safety order issued in the episode	286(13.5)	262(14.4)	8.87	.003	1.92 [1.24-2.98]
Protection order breached in the episode	50(2.4)	42(2.3)	0.15	.694	0.86 [0.40-1.85]
Type of harm					
Physical harm	654(30.9)	548(30.2)	3.51	.061	0.78 [0.61-1.01]
Property damage	300(14.2)	270(14.9)	4.83	.028	1.56 [1.05-2.32]
Sexual violence	6(0.3)	4(0.2)	1.84	.175	0.33 [0.06-1.79]
Threats of harm	292(13.8)	267(14.7)	8.55	.003	1.88 [1.22-2.89]
Verbal abuse	1951(92.2)	1683(92.6)	2.59	.107	0.71 [0.47-1.08]

Note CI = Confidence Interval. An OR greater than one indicates the odds of the variable being present were greater among cases excluded from the final sample than cases included in the final sample.

[a] An OR greater than one for aggressor or victim gender indicates a larger proportion of men were in those roles in cases excluded from the final sample than cases included in the final sample. We excluded gender diverse aggressors from this analysis because the sample size was not large enough for results to be a fair representation.

In terms of the relationship between the aggressor and victim, half were current intimate partners, and a fifth were former intimate partners; meaning that almost two thirds of episodes were classified as IPV. Of the IPV episodes, 48 (3.8%) included same-sex couples. More than one in ten aggressors were children of victims (13.2%), and 6.3% were parents. Of the aggressors who victimized parents, 36.3% (n = 87) were under 18 years of age. Lastly, 5.9% of aggressors were siblings of the victim, and the remaining aggressors had other types of familial relationships (e.g., grandparent, mother-in-law, cousins).

Episode Characteristics

As Table 1 shows, children were present (they witnessed, or were present but did not witness, the FV episode) in half of the episodes in the sample. Pregnancy (of a woman in any episode role) and firearms were rarely noted. Prior consumption of alcohol was observed in almost a third of episodes, and drugs in a tenth; however, records did not indicate whether these variables related to the aggressor or victim.

Most episodes included verbal abuse, and a third involved physical harm. Fewer episodes involved property damage, threats of harm, and sexual violence. Based on these five recorded categories, the mean number of types of harm recorded in episodes was 1.5 (*SD* = 0.7). This number increased to 2.0 (*SD* = 0.7) when excluding the 864 reports for "argument only" episodes (*n remaining* = 953). An offense was detected by police in around one in ten episodes, and police safety orders were issued in around one in ten episodes. A small proportion of episodes represented breaches of protection orders (e.g., aggressor makes unwanted contact via social media or text message, leading to a call for service).

Plan for Analysis

All analyses were completed using IBM SPSS Statistics version 26. First, we completed descriptive statistics and examined rates of recurrence across the risk categories of the DYRA and SAFVR at increasing follow-up periods (3 days, 12 weeks, and 24 weeks). Then we entered the DYRA and SAFVR's risk categories into separate binary logistic regressions predicting FV recurrence at the same three follow-up periods to determine whether the risk categories were a meaningful predictor of recurrence. We chose 'low-risk' as the reference category so we could examine whether the moderate- and high-risk categories had different probabilities of recurrence compared to the low-risk category. After we had examined the DYRA and SAFVR separately, we added each instrument into a model already containing the other instrument, to determine whether they incrementally improved prediction of FV recurrence when controlling for the other.

Next, we calculated the area under the Receiver Operating Characteristics curve (ROC) for the prediction of recurrence, using the DYRA and SAFVR risk categories. The ROC plots sensitivity (true positive rate) against 1-specificity (false positive rate), and the Area Under the Curve (AUC) represents the probability that a randomly selected person with a recurrent episode would have a higher risk classification than a randomly selected person

without a recurrent episode, with greater values indicating better accuracy, and AUCs of .5 being similar to chance. We followed convention and used interpretive criteria for AUCs from Rice and Harris (2005); .56-.63 = small, .64-.70 = moderate, .71-1.0 = large.

To supplement our AUC analyses, we calculated a number of threshold-bound measures of discriminative ability: sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Sensitivity refers to the proportion of aggressors who were identified by the instrument to have a recurrence out of the total number of actual aggressors with a recurrence. Conversely, specificity refers to the proportion of aggressors predicted to remain recurrence-free out of the total number of aggressors who were recurrence-free. In a similar vein, PPV refers to the proportion of aggressors predicted to have a recurrence who actually did, whereas NPV is the proportion of aggressors predicted to remain recurrence-free who did. For all threshold-bound measures we used a threshold of 1.5, which means a low-risk rating was considered a negative outcome prediction (i.e., a prediction that there would be no recurrence) and a moderate or high-risk rating was a positive outcome prediction (i.e., a prediction that there would be a recurrence). We chose this threshold because it mirrors decisions made in practice: screening out low-risk aggressors, and putting resources into moderate- and high-risk aggressors. To address our third and final research question, we repeated AUC analyses with different FV sub-samples.

Results

Data Preparation

Because the focus of this study concerns the performance of the DYRA and SAFVR risk instruments, we removed reports for episodes where the DYRA was not completed (*n* = 297, 14.0%), or the SAFVR was missing (i.e., the field was blank, rather than having ‘no score’; *n* = 6, 0.3%): leaving 1,817 cases remaining. We refer to this as the “final sample”

and use it throughout the analyses. DYRAs were typically left uncompleted because victims were unwilling or unable to answer questions about the aggressor (e.g., victim too young or too intoxicated to answer DYRA items, absent from the scene, or unwilling to engage with police). We compared the characteristics of the initial and final samples and present them in Table 1. There were some statistically significant differences between the two samples, which indicate greater attrition of cases with certain characteristics, relative to the proportion of those cases in the original sample: for example, Māori and Pasifika aggressors and victims, cases with aggressors or victims under 18 years, current rather than former IPV, cases with gang connections, or where an offense was detected or a police safety order given.

Recurrence Rates for the DYRA and SAFVR Risk Categories

Table 2 shows the distribution of aggressors across the DYRA and SAFVR risk categories overall, and at each combination of the instruments' risk categories, with the corresponding rates of recurrence. There was a small positive relationship between aggressors' DYRA and SAFVR risk categories, indicating the two instruments shared a small amount of risk variance ($\rho = .06$, $p = .011$). About 1 in 7 aggressors (15.2%) were rated as high-risk on the SAFVR and low-risk on the DYRA, and a smaller proportion (7.8%) were rated as low-risk on the SAFVR and high-risk on the DYRA.

The base rate of recurrence increased from 5.0% at 3 days, to 34.8% at 12 weeks, and 44.5% at 24 weeks (see extreme right cell in bottom row of Table 2). Of the 407 aggressors (22.4%) who were rated as high-risk on both instruments, 57.7% had recurrence by 24 weeks. This rate is much higher than the rate of recurrence for aggressors in the high-risk category of the DYRA and the moderate-risk category of the SAFVR (38.5%), but is only slightly higher than the rate for aggressors in the moderate-risk category of the DYRA and the high-risk category of the SAFVR (55.6%). The lowest rate of recurrence was observed among

aggressors in the low-risk category of the DYRA and the no score category of the SAFVR (12.5%).

Overall, total recurrence rates for the DYRA categories differed only modestly, with little distinction between recurrence rates for the moderate- and high-risk categories and the base rate at any follow-up period. At the 3-day follow-up period, the DYRA's low-risk category had around half the rate of recurrence of the moderate- and high-risk groups. At 24 weeks, there was less than 9% between the rates of recurrence across the DYRA's three risk categories.

Table 2

The Number (%) of Aggressors in Rach Risk Category of the DYRA and SAFVR and the Number (%) of Aggressors with FV Recurrence at Three Follow-up Periods (n = 1,817).

	SAFVR				
DYRA	No score	Low	Moderate	High	Total
Low[a]	96 (5.3)	128 (7.0)	67 (3.7)	277 (15.2)	568 (31.3)
3 days[b]	2 (2.1)	5 (3.9)	0 (0.0)	12 (4.3)	19 (3.3)
12 weeks[b]	7 (7.3)	34 (26.6)	15 (22.4)	108 (39.0)	164 (28.9)
24 weeks[b]	12 (12.5)	42 (32.8)	21 (31.3)	148 (53.4)	223 (39.3)
Moderate[a]	87 (4.8)	99 (5.4)	57 (3.1)	250 (13.8)	493 (27.1)
3 days[b]	6 (6.9)	5 (5.1)	4 (7.0)	19 (7.6)	34 (6.9)
12 weeks[b]	19 (21.8)	26 (26.3)	19 (33.3)	112 (44.8)	176 (35.7)
24 weeks[b]	23 (26.4)	32 (32.3)	27 (47.4)	139 (55.6)	221 (44.8)
High[a]	98 (5.4)	142 (7.8)	109 (6.0)	407 (22.4)	756 (41.6)
3 days[b]	5 (5.1)	5 (3.5)	3 (2.8)	25 (6.1)	38 (5.0)
12 weeks[b]	22 (22.4)	54 (38.0)	34 (31.2)	182 (44.7)	292 (38.6)
24 weeks[b]	25 (25.5)	62 (43.7)	42 (38.5)	235 (57.7)	364 (48.1)
Total[a]	281 (15.5)	369 (20.3)	233 (12.8)	934 (51.4)	1817 (100.0)
3 days[b]	13 (4.6)	15 (4.1)	7 (3.0)	56 (6.0)	91 (5.0)
12 weeks[b]	48 (17.1)	114 (30.9)	68 (29.2)	402 (43.0)	632 (34.8)
24 weeks[b]	60 (21.4)	136 (36.9)	90 (38.6)	522 (55.9)	808 (44.5)

Note [a]The number and percentage of aggressors in the relevant categories of the DYRA and SAFVR.
[b]The number and percentage of aggressors with recurrence, at each follow-up period, out of those in the relevant categories of the DYRA and SAFVR.

On the other hand, there was some separation in the total rates of recurrence for the SAFVR categories; by 24 weeks more than half of aggressors in the high-risk category (55.9%) had a recurrence, with recurrence rates for all other categories being lower than the base rate. However, more aggressors in the low-risk category had a recurrent episode than those in the moderate-risk category at follow-up periods of 3 days and 12 weeks. Ultimately, as expected, the proportion of aggressors with a recurrence in each category increased over time.

Examining the Predictive Ability of the DYRA and SAFVR Risk Categories

Logistic regression analysis for DYRA risk categories (Model 1; Table 3) for all follow-up periods were statistically significant at the alpha level of .05, but pseudo R^2 calculations indicated the DYRA categories did not perform much better than a null model. At the 3-day and 12-week follow-ups, aggressors in the moderate-risk category had greater odds of a recurrent episode than aggressors in the low-risk category, but this difference was not statistically significant at 24 weeks. At the 12- and 24-week follow-up periods, aggressors in the high-risk category had greater odds of a recurrent episode than aggressors in the low-risk category; this difference was not statistically significant at 3 days.

Logistic regression analysis for SAFVR risk categories (Model 2; Table 3) showed that models for the 12- and 24-week follow-up periods were statistically significant, but again, pseudo R^2 calculations indicated the SAFVR did not perform much better than a null model. In the 12- and 24-week models, aggressors in the no score category had lower odds of recurrence, and aggressors in the high-risk category had greater odds of recurrence, than those for the low-risk category. The moderate-risk category could not be distinguished from the low-risk category at any follow-up period.

Table 3

Binary Logistic Regressions Using DYRA and SAFVR Risk Categories to Predict FV Recurrence at Three Follow-up Periods (n = 1,817)

	Model 1 – DYRA					Model 2 – SAFVR					Model 3 – SAFVR + DYRA				
	B	*SE*	Wald χ^2	*p*	*OR* [95% CI]	*B*	*SE*	Wald χ^2	*p*	*OR* [95% CI]	*B*	*SE*	Wald χ^2	*p*	*OR* [95% CI]
3 days															
DYRA															
Low	-	-	6.80	.033	-	-	-	-	-	-	-	-	6.66	.036	-
Moderate	0.76	0.29	6.73	.009	2.14 [1.20-3.80]	-	-	-	-	-	0.75	0.29	6.59	.010	2.13 [1.20-3.78]
High	0.43	0.29	2.20	.138	1.53 [0.87-2.68]	-	-	-	-	-	0.42	0.29	2.09	.148	1.52 [0.86-2.66]
SAFVR															
No score	-	-	-	-	-	0.14	0.39	0.12	.727	1.15 [0.54-2.45]	0.12	0.39	0.09	.762	1.13 [0.53-2.41]
Low	-	-	-	-	-	-	-	4.55	.208	-	-	-	4.40	.221	-
Moderate	-	-	-	-	-	-0.31	0.47	0.45	.501	0.73 [0.29-1.82]	-0.33	0.47	0.49	.486	0.72 [0.29-1.80]
High	-	-	-	-	-	0.41	0.30	1.89	.169	1.51 [0.84-2.70]	0.39	0.30	1.74	.187	1.48 [0.83-2.66]
Model	R^2 = .01; $\chi^2(2)$ = 7.05, p = .030					R^2 = .01; $\chi^2(3)$ = 4.91, p = .178					R^2 = .02; $\chi^2(5)$ = 11.80, p = .038 [a]				
12 weeks															
DYRA															
Low	-	-	13.77	.001	-	-	-	-	-	-	-	-	11.62	.003	-
Moderate	0.31	0.13	5.63	.018	1.37 [1.06-1.77]	-	-	-	-	-	0.32	0.14	5.64	.018	1.38 [1.06-1.79]
High	0.44	0.12	13.58	< .001	1.55 [1.23-1.96]	-	-	-	-	-	0.41	0.12	11.19	.001	1.50 [1.18-1.90]
SAFVR															
No score	-	-	-	-	-	-0.78	0.19	15.87	< .001	0.46 [0.32-0.68]	-0.78	0.20	15.88	< .001	0.46 [0.31-0.67]
Low	-	-	-	-	-	-	-	68.85	< .001	-	-	-	67.06	< .001	-
Moderate	-	-	-	-	-	-0.08	0.18	0.20	.656	0.92 [0.64-1.32]	-0.11	0.18	0.34	.558	0.90 [0.63-1.29]
High	-	-	-	-	-	0.53	0.13	16.15	< .001	1.69 [1.31-2.18]	0.51	0.13	15.03	< .001	1.66 [1.29-2.15]
Model	R^2 = .01; $\chi^2(2)$ = 14.03, p = .001					R^2 = .06; $\chi^2(3)$ = 76.67, p < .001					R^2 = .07; $\chi^2(5)$ = 88.51, p < .001 [b]				

Table 3 Continued

	Model 1 – DYRA					Model 2 – SAFVR					Model 3 – SAFVR + DYRA				
	B	*SE*	Wald χ^2	*p*	*OR* [95% CI]	*B*	*SE*	Wald χ^2	*p*	*OR* [95% CI]	*B*	*SE*	Wald χ^2	*p*	*OR* [95% CI]
24 weeks															
DYRA															
Low	-	-	10.38	.006	-	-	-	-	-	-	-	-	7.66	.022	-
Moderate	0.23	0.13	3.36	.067	1.26 [0.98-1.61]	-	-	-	-	-	0.24	0.13	3.29	.070	1.26 [0.98-1.63]
High	0.36	0.11	10.35	.001	1.44 [1.15-1.79]	-	-	-	-	-	0.32	0.12	7.46	.006	1.38 [1.09-1.73]
SAFVR															
No score	-	-	-	-	-	-0.77	0.18	17.84	< .001	0.47 [0.33-0.66]	-0.77	0.18	17.81	< .001	0.47 [0.33-0.66]
Low	-	-	-	-	-	-	-	115.0	< .001	-	-	-	112.76	< .001	-
Moderate	-	-	-	-	-	0.08	0.17	0.19	.662	1.08 [0.77-1.51]	0.06	0.17	0.10	.751	1.06 [0.75-1.48]
High	-	-	-	-	-	0.78	0.13	37.57	< .001	2.17 [1.70-2.78]	0.76	0.13	36.18	< .001	2.14 [1.67-2.75]
Model	R^2 = .01; $\chi^2(2)$ = 10.45, p = .005					R^2 = .09; $\chi^2(3)$ = 126.77, p < .001					R^2 = .10; $\chi^2(5)$ = 134.48, p < .001 [c]				

Note CI = confidence interval. Reference category = Low. All pseudo R^2 are Nagelkerke.

[a] $\chi^2(3) = 4.76$, $p = .190$ for addition of the SAFVR to the DYRA ; $\chi^2(2) = 6.89$, $p = .032$ for addition of the DYRA to the SAFVR

[b] $\chi^2(3) = 74.47$, $p < .001$ for addition of the SAFVR to the DYRA; $\chi^2(2) = 11.83$, $p = .003$ for addition of the DYRA to the SAFVR

[c] $\chi^2(3) = 124.02$, $p < .001$ for addition of the SAFVR to the DYRA; $\chi^2(2) = 7.71$, $p = .021$ for addition of the DYRA to the SAFVR

Last, we evaluated whether each instrument incremented the prediction of FV recurrence over the other (Model 3; Table 3). The results showed that adding the DYRA to the SAFVR statistically significantly improved prediction of FV recurrence at all three follow-up periods, and that adding the SAFVR to the DYRA significantly improved prediction of FV recurrence at 12- and 24- weeks, but not at 3 days.

Assessing the Discriminative Ability of the DYRA

Recall our primary research question: How well does the DYRA discriminate between FV aggressors with and without a FV recurrence? To answer this question, we calculated AUC statistics for the DYRA risk level in predicting recurrence within the same follow-up periods as for the regression analyses, and present the AUCs in the top half of Table 4. AUCs for the 3-day follow up period were not statistically significant, but promisingly, NPV of .97 indicates only 3% of people in the low-risk category had recurrence, whereas PPV of .06 indicates 6% of people in the moderate- and high-risk categories had recurrence. AUCs for follow-up periods of 12 and 24 weeks were statistically significant, but they were close to chance (.54–.55; Rice & Harris, 2005).

Much of the DYRA's poor performance can be attributed to a high rate of false positives; although sensitivity statistics indicated the correct detection of most aggressors who had a recurrence, poor PPV highlighted that only a small proportion of aggressors predicted to have a recurrence actually did, specificity indicated there were many false positives. This pattern is reflected simply in Table 2, where can see that fewer than half of aggressors in the moderate- and high-risk categories of the DYRA had a recurrence within 24 weeks.

Comparing the Discriminative Ability of the DYRA to the SAFVR

Our second research question was: how does the discriminative ability of the DYRA compare to that of the SAFVR? To address this question, we calculated AUC statistics for the

SAFVR risk categories in predicting recurrence within the same range of follow-up periods and present these results in the lower part of Table 4. AUCs for the follow-up periods of 12 and 24 weeks were statistically significant; however, they were small at 12 weeks (.61), and moderate at 24 weeks (.64; Rice & Harris, 2005). We calculated the difference in the area under the ROC curves for the DYRA and SAFVR and found the SAFVR had a better ability to discriminate between aggressors with and without recurrence than the DYRA at both the 12-and 24-week follow-up periods.

Table 4

AUCs for DYRA and SAFVR Risk Categories Predicting FV Recurrence at Three Follow-up Periods (n = 1,817)

	AUC[a] [95% CI]	*SE*	*p*	SN	SP	PPV	NPV
DYRA							
3 days[b]	.53 [.48-.59]	0.03	.296	.79	.32	.06	.97
12 weeks[c]	.55 [.52-.58]	0.01	< .001	.74	.34	.38	.71
24 weeks[d]	.54 [.51-.57]	0.01	.003	.72	.34	.47	.61
SAFVR							
3 days[b]	.55 [.48-.61]	0.03	.145	.69	.36	.05	.96
12 weeks[c]	.61 [.58-.63]	0.01	< .001	.74	.41	.40	.75
24 weeks[d]	.64 [.61-.66]	0.01	< .001	.76	.45	.52	.70

Note SN = sensitivity; SP = specificity, threshold = 1.5

[a] Interpretive criteria for AUCs from Rice & Harris (2005): .56-.63 = small, .64-.7 = moderate, .71-1 = large

[b] There was no significant difference in the DYRA and SAFVR AUCs at the 3-day follow-up period.

[c] There was a statistically significant difference in the DYRA and SAFVR AUCs at the 12-week follow-up period; z = -3.32, p = .001, AUC difference = -.06 [-.10- -.02].

[d] There was a statistically significant difference in the DYRA and SAFVR AUCs at the 24-week follow-up period; z = -5.62, p < .001, AUC difference = -.10 [-.13- -.06].

Table 5

AUC for DYRA and SAFVR Risk Categories Predicting FV Recurrence Within 24 weeks Across Different Relationship and Aggressor Characteristics

	n (% recurrence) [a]	AUC[b] [95% CI]	*SE*	*p*	SN	SP	PPV	NPV
DYRA								
Type of FV								
IPV (all)	1249 (47.0)	.55 [.52-.58]	0.02	.004	.76	.31	.50	.59
IPV + children	611 (46.6)	.55 [.50-.60]	0.02	.036	.81	.29	.40	.63
IPV - children	638 (47.3)	.54 [.50-.59]	0.02	.052	.72	.34	.49	.57
Other FV (all)	568 (38.9)	.52 [.47-.56]	0.03	.535	.63	.40	.40	.63
Other FV + children	307 (38.1)	.51 [.44-.58]	0.03	.771	.64	.37	.39	.63
Other FV - children	261 (39.8)	.53 [.45-.60]	0.04	.501	.62	.43	.42	.63
Aggressor characteristics								
Male	1368 (47.0)	.54 [.50-.57]	0.02	.026	.74	.32	.49	.58
Female	446 (36.5)	.54 [.49-.60]	0.03	.125	.67	.39	.39	.67
Below 18	139 (37.4)	.55 [.45-.64]	0.05	.375	.71	.35	.39	.67
Above 18	1678 (45.1)	.54 [.51-.57]	0.01	.004	.73	.34	.47	.60
Māori	735 (52.5)	.53 [.48-.57]	0.02	.224	.68	.37	.54	.51
European	860 (42.2)	.54 [.50-.58]	0.02	.035	.77	.29	.44	.63
SAFVR								
Type of FV								
IPV (all)	1249 (47.0)	.63 [.60-.66]	0.02	< .001	.81	.39	.54	.69
IPV + children	611 (46.6)	.61 [.56-.65]	0.02	< .001	.81	.35	.52	.68
IPV - children	638 (47.3)	.64 [.60-.69]	0.02	< .001	.80	.44	.56	.71
Other FV (all)	568 (38.9)	.64 [.59-.68]	0.02	< .001	.63	.56	.48	.70
Other FV + children	307 (38.1)	.61 [.55-.68]	0.03	.001	.51	.62	.46	.67
Other FV - children	261 (39.8)	.67 [.60-.73]	0.03	< .001	.76	.49	.50	.76
Aggressor characteristics								
Male	1368 (47.0)	.61 [.58-.64]	0.02	< .001	.81	.36	.53	.68
Female	446 (36.5)	.68 [.63-.73]	0.03	< .001	.56	.69	.51	.73
Below 18	139 (37.4)	.63 [.53-.73]	0.05	.011	.17	.93	.60	.65
Above 18	1678 (45.1)	.63 [.61-.66]	0.01	< .001	.80	.41	.52	.71
Māori	735 (52.5)	.60 [.56-.64]	0.02	< .001	.85	.29	.57	.64
European	860 (42.2)	.61[.57-.65]	0.02	< .001	.71	.45	.49	.68

Note SN = sensitivity; SP = specificity, threshold = 1.5

[a]The number of aggressors in the sub-sample and the percentage of aggressors with a recurrence in the sub-sample.

[b]Interpretive criteria for AUCs from Rice & Harris (2005): .56-.63 = small, .64-.70 = moderate, .71-1.0 = large

Similar to the result we observed with the DYRA, the poor performance of the SAFVR can be attributed to it making many false positive predictions, although these were present to a lesser degree than they were with the DYRA. In fact, the SAFVR's specificity statistics were statistically significantly superior to those of the DYRA because it generated more true negative and fewer false positive predictions (McNemar's $p_{3\text{ days}} = .008$, $p_{24\text{ weeks}} < .001$, for the no recurrence group). Overall, despite statistically significant differences, the picture was relatively similar across the two instruments; like the DYRA, the SAFVR correctly detected most aggressors who had a recurrence, but only a small proportion of aggressors predicted to have a recurrence actually did.

Discriminative Ability Across Different FV Sub-Samples

The third question we set out to answer was: to what extent do the discriminative abilities of the DYRA and SAFVR differ for FV sub-samples? To answer this question, we repeated AUC analyses with different sub-samples, and present the results for the 24-week follow-up period in Table 5. The first column of Table 5 shows the number of aggressors in the sub-sample and the percentage of aggressors in that sub-sample with a recurrence by 24-weeks. Although statistically significant, AUCs were close to chance when the DYRA was applied to all episodes involving IPV, and episodes of IPV where children were involved. AUCs were not statistically significant for other types of FV, or for IPV episodes where children were not present. On the other hand, for all FV subsamples the AUCs for the SAFVR were all statistically significant, and small to moderate (Rice & Harris, 2005); they were largest for other types of FV where children were not present, and for cases with female aggressors. Confidence intervals were wider for aggressors under 18 years of age, indicating greater error among this group, possibly due to the smaller sample size. All AUCs for the 3-day follow-up period were not statistically significant, and at 12 weeks, the pattern of results

was similar to the pattern shown in Table 5. Tables for the 3-day and 12-week follow-up periods can be found in the supplemental materials.

Discussion

In this paper we reported on one aspect of predictive validity for FV recurrence—discriminative ability—for the DYRA and SAFVR risk assessment instruments in current use by New Zealand police. Our primary research question was: How well does the DYRA discriminate between FV aggressors with and without a FV recurrence? Taken together, the results from the analyses in this study indicate the DYRA displayed a poor ability to discriminate between aggressors who had a recurrence and those who did not. In fact, AUCs were either not significant or were close to chance (.54–.55), and pseudo R^2 calculations indicated the model did not improve the fit to the data compared to a null (or constant-only) model.

Although AUCs are relatively robust to base rates, part of the DYRA's poor performance on this metric in the 3-day follow-up period can be attributed to the low rates of recurrence that occurred, which made it very difficult to identify positive cases (Helmus & Babchishin, 2016). Even with this disadvantage, at 3 days the recurrence rate for the DYRA low-risk category was approximately half of that for the moderate- and high-risk categories, and the DYRA significantly incremented the predictive ability of the SAFVR. The similarity between base rates of recurrence and the recurrence rates for the high-risk category indicates the DYRA was over-inclusive; providing positive predictions for many aggressors who did not have a recurrence. At 24 weeks, the recurrence rate for high-risk was much closer to the base rate than is desirable for a group labeled as 'high-risk' (see Table 2). A potential solution to the DYRA's poor discriminative ability may be to recalibrate the instrument and find optimal score thresholds, so that its risk categories match the expected rates of

recurrence for each group (Hanson et al., 2017). In addition, further investigation of the DYRA's high-risk 'flag' items is needed to determine whether they *should* be used as flag factors. If these items are not found to be significantly and positively related to FV recurrence, then their usage may be one of the reasons for poor performance of the DYRA in this study.

Our second research question was: how does the discriminative validity of the DYRA compare to that of the SAFVR? The SAFVR had significantly larger AUCs in the 12- and 24-week follow-up periods and had better specificity than the DYRA; however, the SAFVR still displayed a relatively poor ability to discriminate between aggressors with a recurrence and those without. The AUCs calculated for the SAFVR were similarly unimpressive; they were small at 3 and 12 weeks (.55–.61), and moderate at 24 weeks (.64; Rice & Harris, 2005). Again, the main problem affecting the SAFVR's discriminative ability was that it was over-inclusive. There was good separation between the high-risk category and the other categories, and the 24-week rate of recurrence for this group was further from the base rate than was observed for the high-risk category of the DYRA. However, the recurrence rate for the moderate-risk category was lower than the base rate. This result suggests that aggressors in the moderate-risk category of the SAFVR should be 'screened out' with the low-risk category, rather than 'screened in' with the high-risk category (i.e., by using a threshold of 2.5 on AUC analyses, rather than 1.5).

To answer our third research question, we evaluated the discriminative ability of the DYRA and SAFVR for different FV sub-samples at the 24-week follow-up period. We again found most AUCs were small. The AUCs for the DYRA when applied to non-IPV types of FV were not statistically significant, and this result may be due in part to the structure of the DYRA instrument. Two DYRA items only apply in intimate relationships, and four apply when children are present; therefore, there were more 'opportunities' for aggressors who

harm their intimate partners while children are present to receive higher risk ratings, and it is possible that the DYRA may not be valid for cases that do not involve IPV.

AUCs for the SAFVR were similar to the DYRA's for most FV sub-samples (i.e., small; Rice & Harris, 2005), although some were moderate, and all were statistically significant. The reason why the SAFVR risk categories were a significant predictor for all sub-samples may be because the SAFVR instrument includes criminal history items that are not specific to certain types of relationships: these kinds of variables are good predictors for criminal behavior in general (Hilton et al., 2001; Hilton & Eke, 2016).

Comparison to Other FV Risk Assessment Instruments

Although only limited conclusions can be drawn in interpreting comparisons of AUCs across samples due to differences in samples and outcome criteria; overall, the AUCs calculated for the DYRA in this study were poorer than those calculated for most widely used IPV risk assessment instruments, and are comparable to the lower performing instruments in the van der Put (2019) review. Moreover, the DYRA's AUCs were poorer than those calculated for the FVRAT in an Australian study, admittedly with a slightly different outcome criterion: repeat FV related police calls for service for *victims* within 6 months (Dowling, & Morgan, 2019).

Most of the AUCs generated for the SAFVR in this study were small, and only the AUC for the 24 week follow-up period was comparable to those generated for other risk assessment instruments included in meta analyses and reviews (Hanson et al. 2007; Messing & Thaller, 2013; Svalin & Levander, 2019; van der Put, 2019). We observed smaller AUCs than were reported in the validation of the ODARA in New Zealand, and in the development and reanalysis of the SAFVR (Bissielo & Knight, 2016; Heister, 2018, 2019; Mossman, 2014). However, differences in the definition for recurrence employed in this study and the original outcome criterion used in the SAFVR may be an explanation for why it displayed

relatively poorer discriminative ability here (Heister, 2018, 2019). The development of the SAFVR used the outcome criterion of a FV offense rather than repeat calls for service, and over a much longer follow-up period of 2 years.

Alternative Explanations

The ability of the DYRA and SAFVR to discriminate between aggressors with and without a recurrence may have been mediated by the effects of police action and the possible interventions put in place by the multi-agency FV response pilot operating in the sample districts (Mossman, et al., 2017); possibilities we were not able to investigate in this study. The finding that there was little difference between recurrence rates for the moderate and high-risk categories of the DYRA, and the low and moderate categories of the SAFVR, may provide support for this notion. At higher risk levels, the level of intervention delivered by police and other agencies increases (Integrated Safety Response, 2018). Therefore, even though rates of recurrence are high, it is still possible that the increased amount of support delivered to these cases is affecting the instrument's ability to detect cases that remain at higher risk.

It is also possible that prevention advice is working; for example, police and other agencies encourage victims to call police as a de-escalation method (Walton & Brooks, 2019). Therefore, repeat calls for police service may not indicate that more harm is occurring; but rather, it may suggest families are seeking help, and perhaps at an earlier stage. To investigate whether outcomes are improving or not, further analysis could examine whether a greater volume of recurrence or more serious physical harm over time is perpetrated by aggressors in higher risk tiers.

The results of this study may have been affected by differences between the initial sample and the final sample used for analysis. We could only include cases with an available DYRA and SAFVR risk category; almost all exclusions were due to the DYRA not being

completed. We were unable to determine in this study why DYRAs were missing, but we did find a number of correlates of non-completion that suggest it warrants further examination. For example, cases with Māori or Pasifika aggressors or victims, aggressors or victims less than 18 years old, where the aggressor was a current rather than former partner, and where police reported an offense was detected or they issued a safety order all were less likely than expected to have a completed DYRA. These differences suggest that DYRA results for cases with these characteristics may be less representative of the original sample from which they are drawn, than for European and Asian aggressors, IPV cases where the partners are separated, and so on. There could be a number of different reasons for these discrepancies, and it would be valuable to understand more about what they are.

The discriminative ability of the DYRA and SAFVR in this study may also have been affected by the process of selecting the aggressor in each episode (Hamel, 2011). While this decision is clear cut in instances where one individual has committed a FV-related offense, it is more difficult to select an aggressor in "argument-only" episodes, which were common in our sample, or in other cases of apparent bidirectional violence. In the absence of information about the FV history of the parties involved, police make this determination based on the physical capacity to cause harm, leading to a gender bias toward male adults in heterosexual relationships (Hamel, 2011), and possibly in other types of relationships, too. Logically, selecting the incorrect individual on which to complete a risk assessment instrument should affect the predictive validity of that instrument. Deeper investigation is required to ascertain the extent to which the selection of the aggressor impacts the predictive validity of the risk assessment instruments in use by frontline police (Hamel, 2011).

Another factor that is likely to have affected the discriminative ability of the DYRA and SAFVR is the use of risk categories rather than scores. Because actual scores were not available for the SAFVR, using the risk categories was the only way could analyze the data,

and this was a limitation of our study. By using categorical data, we had less variance for statistical testing than other studies that evaluate risk assessment instruments with continuous data (Helmus & Babchishin, 2016), and we could not investigate whether other cut-off scores for the risk categories would improve the discriminative ability of the two instruments. However, this was a field study, and risk categories are what police use in practice to inform their decision making; therefore this study has contributed to the important practical aim of exploring the extent to which these categories can be used to discriminate between aggressors with and without a recurrence (Singh et al., 2013). That said, the use of field ratings is also a potential limitation, because despite police training and quality assurance, we could not quantify the reliability of police ratings, which may have been lower than for research ratings.

Practical Implications

The results from this study suggest that the DYRA in its current form is not performing adequately as a frontline triage instrument, and the SAFVR, while a little better, is at the lower end of acceptable accuracy. Both are substantially overidentifying aggressors as likely to be involved in further FV-related calls for service. On the face of it, it may be argued that these high rates of false positives cause less harm than false negatives. These instruments are not used as part of a comprehensive risk assessment protocol, but instead, mainly for deciding on the level of crisis response, prior to a fuller, interagency-based risk assessment for higher risk cases. Because the purpose of these instruments is to screen and inform police responses, rather than inform a plan for treatment, false positives may lead to police taking situations more seriously, and a greater number of families being flagged for extra intervention. However, as we noted earlier, responding to calls for service is estimated to consume close to half of all police frontline time. With volumes still continuing to increase in association with public education campaigns (Fyfe & Walker, 2019), an increase in the

number of cases triaged as requiring intervention puts considerable strain on the ability of police to respond to calls in a timely fashion, and also puts pressure on the downstream services who complete fuller risk assessments and provide interventions. In the wider criminal justice system there is also evidence that failing to screen out lower risk cases can itself be criminogenic for those cases (Bonta & Andrews, 2016), and it is possible that this finding applies here as well. Therefore further development and validation of the DYRA and SAFVR to reduce the number of cases referred for more comprehensive risk assessment and intervention planning is justified.

Future Directions

There are clear next steps that could build on the knowledge generated in this study. The first of these is to investigate distribution of actual scores on the DYRA (with and without the high-risk 'flag' items) and SAFVR to explore the probabilities associated with more optimal risk 'bins', improve discriminative ability and lower the rate of false positives (Hanson et al., 2017). It is also important to explore the ability of these instruments to predict different types of recurrence or recidivism outcome criteria, such as episodes containing physical violence, offenses, and convictions, or the presence of multiple repeat episodes within a given timeframe (Heckert & Gondolf, 2005; Jones et al., 2010).

Another direction for future research is to investigate the other side of the predictive validity coin: calibration. In this study, we evaluated discriminative ability, but another, often overlooked indicator of predictive performance is to compare expected recidivism or recurrence rates for risk categories generated from an instrument against the actual rates (Hanson, 2017; Singh et al. 2013). It is important too, to calibrate instruments across different sub-samples, to evaluate whether the instrument performs equivalently in predicting recurrence across demographic groups (e.g., for both Māori and European New Zealanders). Moreover, such analyses should account for the level of intervention delivered to the families

in question, to ensure differences observed between groups are not an artefact of differences in the services they receive.

Conclusion

This study was an important first step into investigating and comparing the discriminative ability of two instruments designed for use by police on the front line to triage risk of recurrence for diverse types of FV. Although there is clear room for improvement, the relatively consistent performance of the SAFVR instrument across FV sub-types and follow-up periods suggests that assessment based on the aggressor's criminal history remains an adequate approach to identifying cases that warrant further assessment and intervention services. But, prediction based more on family living circumstances or recent non-criminal behavior by the aggressor needs further investigation if it is to be a useful stand-alone or additional source of risk assessment information.

References

Adams, A. (2016). Social investment in the criminal justice system. *Beehive.* https://www.beehive.govt.nz/speech/social-investment-criminal-justice-system

Bissielo, A., & Knight, G. (2016). *Family violence risk assessment redevelopment: Static risk score*. (New Zealand Police internal report). New Zealand Police.

Bonta, J., & Andrews, D. A. (2016). *The psychology of criminal conduct* (6th ed.). Routledge. https://doi.org/10.4324/9781315677187

Chan, K. L. (2012). Evaluating the risk of child abuse: The Child Abuse Risk Assessment Scale (CARAS). *Journal of Interpersonal Violence, 27*(5), 951–973. https://doi.org/10.1177/0886260511423252

Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy and Law, 11*(3), 347–383. https://doi.org/10.1037/1076-8971.11.3.347

Dowling, C., & Morgan, A. (2019). *Predicting repeat domestic violence: Improving police risk assessment*. Australian Institute of Criminology. https://www.aic.gov.au/publications/tandi/tandi581

Family Violence Act (2018). http://www.legislation.govt.nz/act/public/2018/0046/latest/DLM7159322.html

Fyfe, J. & Walker, L. (2019, November 26). NZ's shame: The regions where family violence is highest. *Newshub.* https://www.newshub.co.nz/home/new-zealand/2019/11/nz-s-shame-the-regions-where-family-violence-is-highest.html

Gondolf, E. W. (2002). *Batterer intervention systems: Issues, outcomes, and recommendations*. Sage.

Graham, L. M., Sahay, K. M., Rizo, C. F., Messing, J. T., & Macy, R. J. (2019). The validity and reliability of available intimate partner homicide and reassault risk assessment

tools: A systematic review. *Trauma, Violence and Abuse, 22*(1), 18–40. https://doi.org/10.1177/1524838018821952

Grant, S. C. (2009). *Family violence risk assessment: An early study of police officers' experiences at the frontline* [Masters thesis, Victoria University of Wellington]. Victoria University of Wellington Research Archive. https://researcharchive.vuw.ac.nz/xmlui/handle/10063/1088

Hamel, J. (2011). In dubious battle: The politics of mandatory arrest and dominant aggressor laws. *Partner Abuse, 2*(2), 224-245. https://doi.org/10.1891/1946-6560.2.2.224

Hanson, R. K. (2017). Assessing the calibration of actuarial risk scales: A primer on the E/O index. *Criminal Justice and Behavior, 44*(1), 26–39. https://doi.org/10.1177/0093854816683956

Hanson, R. K., Bourgon, G., McGrath, R. J., Kroner, D., D'Amora, D. A., Thomas, S. S., & Tavarez, L. P. (2017). *A five-level risk and needs system: Maximizing assessment results in corrections through the development of a common language*. National Reentry Resource Center. https://saratso.org/pdf/A_Five_Level_Risk_and_Needs_System_Report.pdf

Hanson, R. K., Helmus, L., & Bourgon, G. (2007). *The validity of risk assessments for intimate partner violence: A meta-analysis*. Public Safety Canada. http://www.publicsafety.gc.ca/res/cor/rep/_fl/vra_ipv_200707_e.pdf

Heckert, D. A., & Gondolf, E. W. (2005). Do multiple outcomes and conditional factors improve prediction of batterer reassault? *Violence and Victims, 20* (1), 3–24. https://doi.org/10.1891/vivi.2005.20.1.3

Heister, J. (2018). *SAFVR Reanalysis of Morris and Mossman (2015)*. (New Zealand Police internal report). New Zealand Police.

Heister, J. (2019). *The role of age, ethnicity and gender*. (New Zealand Police internal report). New Zealand Police.

Helmus, L. M., & Babchishin, K. M. (2016). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior, 44*(1), 8–25. https://doi.org/10.1177/0093854816678898

Hilton, N. Z., & Eke, A. W. (2016). Non-specialization of criminal careers among intimate partner violence offenders. *Criminal Justice and Behavior, 43*(10), 1347–1363. https://doi.org/10.1177/0093854816637886

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2001). Predicting violence by serious wife assaulters. *Journal of Interpersonal Violence, 16*(5), 408–423. https://doi.org/10.1177/088626001016005002

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2010). *Risk assessment for domestically violent men: Tools for criminal justice, offender intervention, and victim services*. American Psychological Association. https://doi.org/10.1037/12066-000

Integrated Safety Response (2018). *12 week review of cases referred to the Integrated Safety Response (ISR) pilot*. (Integrated Safety Response internal report). Author.

Jones, A. S., Heckert, D. A., Gondolf, E. D., Zhang, Q., & Ip, E. H. (2010). Complex behavioral patterns and trajectories of domestic violence offenders. *Violence and Victims, 25*(1), 3–17. https://doi.org/10.1891/0886-6708.25.1.3

Kahui, S., & Snively, S. (2014). *Measuring the economic costs of child abuse and intimate partner violence to New Zealand*. The Glenn Inquiry. http://img.scoop.co.nz/media/pdfs/1411/ECONOMIC_COSTS_OF_CHILD_ABUSE__INTIMATE_PARTNER_ABUSE.pdf

Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) guide: Reliability and validity in adult male offenders. *Law and Human Behavior, 24*(1), 101–118. https://doi.org/10.1023/a:1005430904495

López-Ossorio, J. J., González-Álvarez, J. L., Vicente, J. M. M., Cortés, C. U., & Andrés-Pueyo, A. (2019). Validation and calibration of the Spanish police intimate partner violence risk assessment system (VioGén). *Journal of Police and Criminal Psychology*, *34*(4), 439–449. https://doi.org/10.1007/s11896-019-09322-9

Magdol, L., Moffitt, T. E., Caspi, A., & Silva, P. A. (1998). Developmental antecedents of partner abuse: A prospective longitudinal study. *Journal of Abnormal Psychology, 107*(3), 375–389. https://doi.org/10.1037/0021-843X.107.3.375

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some propsals on the nature of psychologically meaningful risk factors. *Sexual Abuse, 22*(2), 191–217. https://doi.org/10.1177/1079063210366039

McEwan, T. E., Shea, D. E., & Ogloff, J. R. P. (2018). An acturial instrument for police triage of Australian family violence reports. *Criminal Justice and Behavior, 46*(4), 590–607. https://doi.org/10.1177/0093854818806031

Messing, J. T., & Thaller, J. (2013). The average predictive validity of intimate partner violence risk assessment instruments. *Journal of Interpersonal Violence, 28*(7), 1537–1558. https://doi.org/10.1177/0886260512468250

Mossman, E. (2014). *Evaluation of the use of ODARA in New Zealand*. (New Zealand Police internal report). New Zealand Police.

Mossman, E., Paulin, J., & Wehipeihana, N. (2017). *Evaluation of the family violence Integrated Safety Response pilot*. SUPERU. https://thehub.sia.govt.nz/assets/documents/ISR_pilot_evaluation_FINAL.pdf

New Zealand Crime and Victims Survey (2018). *Topical report: Offences against New Zealand adults by family members*. Ministry of Justice. https://www.justice.govt.nz/assets/Documents/Publications/9ZU3Q-NZCVS-topical-report-Offences-by-family-members-Cycle-1-2018.pdf

New Zealand Family Violence Clearinghouse (2017). Data summaries 2017: Snapshot. *New Zealand Family Violence Clearinghouse.* https://nzfvc.org.nz/sites/nzfvc.org.nz/files/Data-summaries-snapshot-2017.pdf

New Zealand Police (n.d.-a). *Police approach to family harm.* https://www.police.govt.nz/advice-services/family-violence/family-harm-approach-resources

New Zealand Police (n.d.-b). *Police safety orders.* https://www.police.govt.nz/advice-services/family-violence/police-safety-orders

Nicholls, T. L., Pritchard, M. M., Reeves, K. A., & Hilterman, E. (2013). Risk assessment in intimate partner violence: A systematic review of contemporary approaches. *Partner Abuse, 4*(1), 76–168. https://doi.org/10.1891/1946-6560.4.1.76

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7

SafeLives (2020). *Resources for MARAC meetings.* https://safelives.org.uk/practice-support/resources-marac-meetings

Singh, J. P., Desmarais, S. L., & Dorn, R. A. V. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences and the Law, 31*(1), 55–73. https://doi.org/10.1002/bsl.2053

Spivak, B., McEwan, T., Luebbers, S., & Ogloff, J. (2020). Implementing evidence-based practice in policing family violence: The reliability, validity and feasibility of a risk

assessment instrument for prioritising police response. *Policing and Society*, *31*(4), 483-502. https://doi.org/10.1080/10439463.2020.1757668

Steiner, J. J., Johnson, L., Hetling, A., Lin, H.-F., & Postmus, J. L. (2019). Creating a tool for assessing domestic violence risk and impact among TANF clients. *Advances in Social Work*, *19*(1), 157–180. https://doi.org/10.18060/22606

Svalin, K., & Levander, S. (2019). The predictive validity of intimate partner violence risk assessment conducted by practitioners in different settings: A review of the literature. *Journal of Police and Criminal Psychology*, 35(2), 115–130. https://doi.org/10.1007/s11896-019-09343-4

van der Put, C. E., Assink, M., van Solinge, N. F. B. (2017). Predicting child maltreatment: A meta-analysis of the predictive validity of risk assessment instruments. *Child Abuse and Neglect, 73*, 71–88. https://doi.org/10.1016/j.chiabu.2017.09.016

van der Put, C. E., Gubbels, J., & Assink, M. (2019). Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools. *Aggression and Violent Behavior*, *47*, 100–116. https://doi.org/10.1016/j.avb.2019.03.008

Walton, D. & Brooks, R. (2019). *Whāngaia Ngā Pā Harakeke pilot: Counties Manukau District outcomes evaluation*. (New Zealand Police internal report). New Zealand Police.

Williams, K. R. (2012). Family violence risk assessment: A predictive cross-validation study of the Domestic Violence Screening Instrument-Revised (DVSI-R). *Law and Human Behavior*, *36*(2), 120–129. https://doi.org/10.1037/h0093977

Chapter Three: Manuscript Two

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2023). Unpacking multiagency structured professional judgement risk assessment for family violence. *Journal of Interpersonal Violence.* Advance online publication. https://doi.org/10.1177/08862605221147069

Abstract

Assessing the risk for future harm is a crucial task for agencies managing Family Violence (FV) cases. The Integrated Safety Response (ISR) is a multi-agency collaboration of such agencies operating in two areas of New Zealand, and one of the first steps in their process is to perform a risk assessment. However, in these assessments, it is unclear whether the factors ISR triage team members select are the basis for their overall risk categorization (low, medium, or high), and whether those factors are risk factors (i.e., empirical predictors of outcomes). Therefore, in this study we documented the factors ISR triage teams recorded during their risk assessments for 842 FV cases and examined the relationship of those factors with the risk categories. We then investigated whether those factors and the risk categories were indeed capable of predicting FV-related outcomes (recurrence and physical recurrence). We found most of the triage teams' recorded factors were associated with the risk categories, but fewer than half of the factors were associated with FV-related outcomes. Moreover, the risk categories predicted FV-related outcomes better than chance, but their predictive ability varied across sub-groups, performing poorly for aggressors who were Māori or women, and for non-intimate partner cases. We concluded the ISR triage teams' risk assessment protocol may benefit from increased structure and validation.

Key words: family violence; intimate partner violence; risk factors, responsivity, field validity, risk triage

Unpacking multiagency structured professional judgement risk assessments for family violence

Agencies responding to Family Violence (FV) cases commonly use risk assessments because they aid practitioner judgements about the likelihood of future harm and can be used to inform their efforts to prevent that harm (Garrington & Boer, 2020; Henning et al., 2021; Robinson & Clancy, 2020; Stanley & Humphreys, 2014). The Integrated Safety Response (ISR) is a multi-agency collaboration of organizations responsible for managing FV cases reported to Police in two areas of New Zealand (Integrated Safety Response, 2018). Every FV episode reported to New Zealand Police is referred to ISR for further case management, and one of the first actions the ISR takes is to perform a risk assessment. However, it is unclear whether the factors ISR triage team members identify are the basis for the categorical risk level they allocate to cases (low, medium, or high), as well as whether those factors are risk factors (i.e., empirical predictors of criminal outcomes; Bonta & Andrews, 2016). Therefore, in this study we document the factors ISR triage teams recorded for 842 cases and examine the relationship of those factors with the risk categories. We also analyze whether those factors and the risk categories were indeed capable of predicting two FV-related outcomes (recurrence and physical recurrence).

The RNR Model and Risk Assessment by FV Practitioners

Agencies providing FV services often use risk assessments to guide their decision making (Henning et al., 2021; Robinson & Clancy, 2020; Spivak et al., 2020; Stanley & Humphreys, 2014). In criminal justice settings, risk assessments are designed to predict the likelihood of an individual committing a criminal act in the future (Prins & Reich, 2021). This process improves the ability of law enforcement and other agencies to protect victims and target limited resources at the people who need them most. The risk principle of the Risk Needs Responsivity (RNR) model for correctional treatment states that treatment intensity

should be matched to the level of risk offenders pose; with high-risk offenders having the greatest 'room for improvement', thus also receiving the greatest level of intervention (Bonta & Andrews, 2016; Polaschek, 2012). In correctional settings the RNR principles are commonly used to design and evaluate treatment programs for individual offenders, but the model is also relevant for a wider range of interventions, including family-based interventions (Polaschek, 2016). Translated into the FV context, the risk principle simply indicates that families or whānau[7] with the greatest risk-related needs should be prioritized for effective interventions, and a higher level of service and support (Polaschek, 2016).

Risk assessment instruments are comprised of various risk factors, with the instruments' ability to predict the criminal outcome being drawn from the statistical association those factors have with that outcome. Importantly, this association is not necessarily causal (Kropp, 2004). Risk factors are understood as measurable characteristics that can be used to predict peoples' likelihood of engaging in criminal behavior (Bonta & Andrews, 2016; Polaschek, 2012). In addition, victim-focused risk factors can be used to predict peoples' likelihood of experiencing victimization (Spencer et al., 2019). Many of these characteristics are known as static risk factors, which cannot be changed easily once they have occurred (e.g., age at first offense).

On the other hand, some characteristics used in risk assessments to predict criminal behavior are dynamic risk factors; changeable characteristics of the person or their circumstances (e.g., current substance use; Bonta & Andrews, 2016), with changes in these offending-related needs being followed by changes in the measured criminal outcome (Douglas & Skeem, 2005; Polaschek, 2012). These factors are often psychologically meaningful; they may be causal or act as proxies for the underlying causal mechanisms that

[7] The word whānau refers to extended family or the wider family unit in Te Reo Māori, the Māori language.

lead to criminal behavior (Mann et al., 2010); thus, they are often chosen as targets for change during treatment (Prins et al., 2021).

In line with this idea, a second purpose of risk assessments is to identify information that can be used to inform service provision and case management (Butters et al., 2021; Garrington & Boer, 2020; Kropp, 2008). The responsivity principle from the RNR model states that interventions should be delivered in a way that maximizes people's ability to engage and benefit from them (Bonta & Andrews, 2016). Agencies play a pivotal role in selecting which interventions families receive (e.g., alcohol or drug rehabilitation, parenting support, financial planning, or non-violence programs), and providing those interventions in an engaging and accessible manner. The process of assessing risk, which involves reviewing case information, identifying changeable needs that are linked to recurrence, and thinking through future scenarios where harm may occur, serves the dual function of assisting practitioners to plan their response (Butters et al., 2021; Garrington & Boer, 2020; Kropp, 2008).

Types of Risk Assessments

Risk assessment processes differ in the extent to which they are structured. Actuarial instruments are the most structured; they contain a list of risk factors selected for their ability to predict the outcome. These risk factors are scored and combined algorithmically to produce an overall risk score (Geurts et al., 2021). Contrastingly, in unstructured clinical judgement risk assessments practitioners subjectively evaluate case information and use their expertise in a non-transparent way to identify what they regard as important information, before forming an intuitive impression of the likelihood of future harm (Hilton et al., 2006). Structured Professional Judgement (SPJ) risk assessments sit between these two contrasting methods of assessing risk. SPJ risk assessments typically include empirically derived risk factors, but allow discretion (Hart et al., 2016); practitioners can include information they

deem relevant from outside the instrument's list of risk factors and decide themselves how all the pertinent information (i.e., the identified risk factors) should be weighted to influence the final assessment (Hilton et al., 2006; Kropp, 2004; Kropp, 2008).

Within SPJ, assessment protocols vary in their procedures and in their predictive validity (Garrington & Boer, 2020; Geurts et al., 2021). Some SPJ risk assessment instruments contain a list of items that are rated (e.g., 0, 1, or 2; absent, somewhat present, or present, respectively; Douglas et al., 2013; Garrington & Boer, 2020; Kropp & Hart, 2015). After completing these ratings, practitioners select a categorical label to communicate the level of risk they think is present in the case (e.g., low, moderate, or high). For research, ratings on SPJ instruments are often summed to produce a total score, and when used this way (i.e., actuarially), their predictive validity is similar to that of actuarial instruments (van der Put et al., 2019). When the risk categories produced using SPJ instruments are used to predict outcomes they tend to have poorer predictive validity than actuarial instruments (e.g., DASH; Richards, 2009; Turner et al., 2019), but this may be in part due to the reduced variance inherent in a three-item scale (e.g., low = 1, medium = 2, or high = 3; Jolliffe Simpson et al., 2021).

Some of the most well-known actuarial or SPJ risk assessment instruments were created to predict reassault, rearrest, or reconviction in Intimate Partner Violence (IPV) cases, and include the Revised Domestic Violence Screening Instrument (DVSI-R; Stansfield & Williams, 2014), the Ontario Domestic Assault Risk Assessment (Hilton & Eke, 2017; Hilton et al., 2004) and the Spousal Assault Risk Assessment Guide (Kropp & Hart, 2015), and other instruments are used to predict the risk of intimate partner homicide (e.g., Danger Assessment; Messing et al., 2017). Few validated risk assessment instruments exist for other types of family relationships; a small number have been created for child maltreatment (e.g., California Family Risk Assessment, Dankert & Johnson, 2014; Checklist of Child Safety, van

der Put et al., 2016), and some others can be used with all types of FV (e.g., New Zealand Police Dynamic Risk Assessment; Jolliffe Simpson et al., 2021; Victoria Police Screening Assessment for Family Violence Risk; Spivak et al., 2020). In addition to assessing FV risk, some instruments (e.g., SARA; Kropp & Hart, 2015) explicitly inform case management. But despite the availability of these instruments, practitioners often rely on their expertise when assessing risk for FV cases, opting instead to use a less-structured approach (Walklate & Mythen, 2011).

The Integrated Safety Response (ISR)

Practitioners working in New Zealand's Integrated Safety Response (ISR) multi-agency crisis response service for FV routinely perform risk assessments based on the SPJ approach. The ISR includes representatives from organizations such as Police, District Health Boards, and non-governmental FV service providers (e.g., whānau support, non-violence programs, and Women's refuge). Importantly, Indigenous service providers are partner agencies in the ISR and deliver many of the interventions offered to families. Māori (Indigenous people of New Zealand) are over-represented in ISR cases, making up around one-fifth of ISR cases but one-tenth of the population in the Canterbury area, and over half of ISR cases but one-fifth of the population in the Waikato area[8] (Dobbs & Eruera, 2014; Mossman et al., 2017). Accordingly, responsiveness to Māori is an important indicator of the ISR's performance (Mossman et al., 2017).

One of the first steps in the ISR's process is for a triage team to assess risk for future FV and assign interventions. Triage teams meet daily to discuss recent cases reported in the ISR catchment area, assess the likelihood and severity of future FV, and collaboratively make a risk categorization (low, medium, or high) using a risk assessment guide (see Method and

[8] Waikato and Canterbury were the two areas where the ISR operated during this period.

supplemental materials; Integrated Safety Response, 2018). During their assessments, the triage teams do not score the items in the risk assessment guide; instead, they freely select risk factors from the guide and from information held by their respective agencies and record the factors they think are most relevant for the case in a free-text field in a centralized database. In essence, they use SPJ to perform an idiographic formulation of what underpins risk for the case in question.

This Study

In this study we seek to better understand both the process and the predictive validity of the ISR triage teams' risk assessments. Importantly, triage teams record factors *after* agreeing upon a risk categorization for a case, introducing the possibility that they may retrospectively justify their categorization, rather than using the factors prospectively to inform their judgement of risk. Put simply, it is unclear whether the factors that triage teams select are the basis for the risk category they allocate. Therefore, we will document the factors triage teams recorded for 842 cases and examine their frequencies and the relationship of those factors with the risk categories the teams assigned.

In addition, because triage teams can freely select factors, it is unclear whether the factors chosen are risk factors that can be used to predict the likelihood of future FV (Bonta & Andrews, 2016), and the risk categories' predictive validity has not previously been explored in research. Accordingly, we will examine the association of the factors with FV-related outcomes (recurrence and physical recurrence) and appraise the discriminative ability of the risk categories for those same outcomes. Because FV cases involve heterogeneous aggressors and family relationships, we will explore the risk categories' discriminative ability for different sub-groups of cases according to the relationship of the aggressor to the victim and by aggressor characteristics.

Method

Data Source

This study used archival data sourced from the ISR's Family Safety System database for a larger study on FV in New Zealand (Jolliffe Simpson et al., 2021). The ISR is a FV multi-agency crisis response service including representatives from New Zealand Police, Ara Poutama Aotearoa (Department of Corrections), District Health Boards, Oranga Tamariki (child protection services), Accident Compensation Corporation, and non-governmental organizations that provide services to families and whānau experiencing FV (Mossman et al., 2017). All FV episodes reported to police are referred on to ISR for further triaging, risk assessment, and case management, enabled by information sharing between the organizations involved. A FV episode was defined as an event involving any form of harm to a person who is in, or has been in, a family relationship with the aggressor (Integrated Safety Response, 2019).

Procedure

New Zealand Police extracted the dataset a containing information about each FV episode reported in November and December 2018 in Waikato and Canterbury from the Family Safety System database. We coded the risk factors recorded in ISR case notes for episodes reported in the two-week period between 1–14 November 2018, which covered 842 cases. We also coded FV-related recurrence outcomes using criteria (see below) from information in the archive about episodes reported in the 24 weeks following the index episode where the index aggressor was again recorded as an aggressor or mutual participant[9] (i.e., was not the reported victim).

[9] Police used the role "mutual participant" when they were unable to determine a single aggressor at the scene.

Risk Factors

Triage teams assessed each case and made a risk categorization (low, medium, or high) assisted by a risk assessment guide (Integrated Safety Response, 2018). The guide included 46 risk factors relating to the victim, aggressor, relationship, children or young people, and practice considerations (see supplemental materials; Integrated Safety Response, 2018). The ISR risk assessment guide is based on risk factors and lethality indicators drawn from the international research literature, but it has not yet been validated or evaluated. Triage teams did not score items per se; instead, team members identified relevant risk factors from the information held by their respective agencies and shared that information to guide group decision making. After a triage team agreed on a risk category, they recorded the risk factors present in the case in a free-text field and used those risk factors to form a case-management plan. Although it is likely the triage teams prioritized some risk factors over others, it was not possible for us to establish risk factors' weights from the recorded information, only their presence. It is the presence of these risk factors that we focus on in the present study.

We initially developed a coding scheme to organize into 74 dichotomous variables the risk factors listed in the free-text field in the Family Safety System database. As expected, many of these variables came from the risk assessment guide; but some recorded factors were not listed in the guide, and we coded these as well. We then removed the variable *any mention of child* because it overlapped two other variables (*vulnerable child* and *non-biological child*). The first author coded the presence or absence of each risk factor for all cases, and to examine inter-rater reliability a postgraduate student with experience both with FV practice and with the Family Safety System independently coded 199 cases (23.5% of the sample). We calculated intraclass correlation coefficients and used guidelines from Koo and Li (2015) to interpret them: < .50 = poor, .50-.74 = moderate, .75-.89 = good, .90-1.0 =

excellent. Of the 73 risk factors we examined, 10 had perfect agreement, 21 had excellent agreement, 24 had good agreement, 11 had moderate agreement, and two had poor agreement (see supplemental materials for the intraclass correlations). A further five variables were low in frequency and absent in all cases coded for inter rater reliability, but we left them in because infrequent variables still have the potential to be good risk predictors, and to avoid artificially reducing the heterogeneity of the risk factors. We therefore chose to retain all variables except those with poor agreement, leaving 71 variables for analysis.

FV-Related Outcomes

We recorded two FV-related outcomes. *Recurrence* was a binary indicator defined as any subsequent FV-related call for police service within 24 weeks after the index episode, where the index aggressor was an aggressor or mutual participant (i.e., not the victim). *Physical recurrence* was a binary indicator for whether physical harm was recorded in the police report for any episode reported during the 24 weeks following the index episode.

Analytic Plan

We conducted analyses in R (R Core Team, 2013) and in IBM SPSS Statistics v. 27. We used descriptive statistics and cross tabulations to describe the sample and identify statistically significant associations between demographic characteristics and the ISR triage teams' risk categories. Next, we used descriptive statistics and cross tabulations to identify associations between the recorded factors and risk categories.

To discern whether the factors were risk factors (i.e., empirically associated with the FV-related outcomes), we analyzed their associations with recurrence and physical recurrence with point biserial correlations. Then we calculated the Area Under the receiver operating characteristics Curve (AUC) for the risk categories predicting recurrence or physical recurrence over the following 24 weeks. AUC statistics communicate the probability that a randomly selected person with a given outcome (i.e., recurrence or physical recurrence)

would have a higher risk classification than a randomly selected person without that outcome. AUCs can range from 0-1; values near .5 indicate discriminative ability similar to chance, and values closer to 1 indicate better accuracy (Helmus & Babchishin, 2016). The AUC is one of many statistics used to evaluate predictive accuracy and has several advantages, including being relatively resilient to the base rate of the outcome, and being non-parametric (thus appropriate for ordinal categories; Helmus & Babchishin, 2016). We repeated the AUC analyses for different sub-groups of cases according to the relationship of the aggressor to the victim and by aggressor characteristics.

Results

The average age of aggressors was about 33 years with a small proportion being under 18 years old (n = 67, 8.0%), whereas the average age of victims was about 35 and a similarly small proportion of victims were under 18 years old (n = 55, 6.5%). Around three quarters of aggressors were men, and around three quarters of victims were women. Most aggressors and victims were Māori or New Zealand European. Aggressors were most commonly the current partners of victims (half of the sample) followed by former partners, parents, siblings, or other family members. Almost 60% of cases in the sample were in the Waikato region, with the remainder in Canterbury.

Table 1 shows that over half of cases were classified by an ISR triage team as being at medium risk of future FV outcomes; a third were low-risk, and a small proportion were classified as high-risk (6.7%). The final two columns of the table show chi square statistics for the relationship between demographic variables and the risk categories, with statistically significant associations at α =.05 present for almost all variables. For example, one in three aggressors were women in the low risk category, and this proportion became smaller in the higher risk categories. The proportion of victims who were men followed a similar pattern.

Table 1

Descriptive Statistics for the Sample Overall (N= 842), and by ISR Risk Category

		Risk category f (%)				
Variable	*f (%)*	*Low* (*n* = 316)	*Medium* (*n* = 470)	*High* (*n* = 56)	χ^2	*p*
Aggressor gender [a]					30.67	< .001
Gender diverse	2 (0.2)	1 (0.3)	1 (0.2)	0 (0.0)		
Female	203 (24.1)	107 (33.9)	92 (19.6)	4 (7.1)		
Male	637 (75.7)	208 (65.8)	377 (80.2)	52 (92.9)		
Aggressor ethnicity					11.70	.020
European	360 (42.8)	144 (45.6)	193 (41.1)	23 (41.1)		
Māori	373 (44.3)	119 (37.7)	227 (48.3)	27 (48.2)		
Other	109 (12.9)	53 (16.8)	50 (10.6)	6 (10.7)		
Relationship to victim					30.49	< .001
Current partner	418 (49.6)	139 (44.0)	242 (51.5)	37 (66.1)		
Former partner	177 (21.0)	56 (17.7)	108 (23.0)	13 (23.2)		
Parent	42 (5.0)	18 (42.9)	24 (57.1)	0 (0.0)		
Child	108 (12.8)	51 (47.2)	51 (47.2)	6 (5.6)		
Sibling	51 (6.1)	26 (8.2)	25 (5.3)	0 (0.0)		
Other	46 (5.5)	26 (8.2)	20 (4.3)	0 (0.0)		
Victim gender					24.54	< .001
Female	648 (77.0)	219 (69.3)	375 (79.8)	54 (96.4)		
Male	194 (23.0)	97 (30.7)	95 (20.2)	2 (3.6)		
Victim ethnicity					17.12	.002
European	388 (46.1)	150 (47.5)	210 (44.7)	28 (50.0)		
Māori	351 (41.7)	111 (35.1)	216 (46.0)	24 (42.9)		
Other	103 (12.2)	55 (17.4)	44 (9.4)	4 (7.1)		
ISR area					16.50	< .001
Waikato	480 (57.0)	198 (41.3)	263 (54.8)	19 (4.0)		
Canterbury	362 (43.0)	118 (32.6)	207 (57.2)	37 (10.2)		
Recurrence	389 (46.2)	108 (34.2)	248 (52.8)	33 (58.9)	30.18	< .001
Physical recurrence	174 (20.7)	41 (13.0)	113 (24.0)	20 (35.7)	22.41	< .001
		Risk category M (SD)				
	M (SD)	*Low*	*Medium*	*High*		
Aggressor age	33.03 (12.79)	30.54 (12.75)	31.50 (10.65)	31.12 (10.01)		
Victim age	34.57 (14.07)	37.55 (16.10)	33.54 (12.29)	32.12 (11.52)		

Note. f = frequency.

[a] Gender diverse aggressor was excluded from chi-square test.

There was also an association between the risk categories and ethnicity, whereby a greater proportion of aggressors who were Māori were in the medium and high risk categories, a greater proportion of victims who were Māori were in the medium risk category, and a greater proportion of victims who were European were in the high risk category, compared with other ethnicities. Cases in the high risk category almost exclusively involved IPV, and other family relationships were over-represented in the low risk category. A greater proportion of cases in the Canterbury region had higher risk categorizations than in Waikato.

Finally, table 1 shows the proportion of cases with recurrence or physical recurrence in the 24 weeks following the index episode. Overall, almost half of cases had a recurrence reported to police during this period, with one in five cases experiencing a recurrence in which the aggressor physically harmed the victim. The proportion of cases with recurrence and physical recurrence differed by risk category, with those proportions increasing as the risk category increased.

Table 2 shows the 71 factors coded from the triage teams' risk assessments, which we organized inductively, based on the sections in the ISR risk assessment guide (see supplemental materials; Integrated Safety Response, 2018). Unsurprisingly, the most common factors recorded were types of harm in the index episode: *verbal abuse* (43.2%), *physical harm* (31.2%), followed by *aggressor FV/criminal history* (27.4%). Of note, many other factors were low in frequency, with only 14 being present in more than 15% of cases. On average, there were 6.51 factors present in each case (range = 1-23, *SD* = 3.50).

The final two columns of Table 2 show chi square statistics for the relationship between the presence of each factor and the risk categories. More than half of the chi square statistics were statistically significant at $\alpha = .05$, indicating most of the factors appeared in different proportions across the three risk categories, with the overall trend being that factors were more commonly present in higher-risk cases compared with low-risk cases.

Table 2

ISR Factor Frequency by ISR Risk Category

		Risk category f (%)				
	f (%)	*Low*	*Medium*	*High*	χ^2	*p*
Aggressor						
Alcohol use	229 (27.2)	85 (26.9)	129 (27.4)	15 (26.8)	0.03	.983
Breaching/non-complying with conditions/orders	78 (9.3)	4 (1.3)	65 (13.8)	9 (16.1)	38.79	< .001
Drug use	151 (17.9)	31 (9.8)	108 (23.0)	12 (21.4)	22.76	< .001
FV/criminal history	231 (27.4)	40 (12.7)	168 (35.7)	23 (41.1)	56.19	< .001
Mental health issues	134 (15.9)	48 (15.2)	76 (16.2)	10 (17.9)	0.31	.859
Suicide/self-harm	90 (10.7)	23 (7.3)	63 (13.4)	4 (7.1)	8.22	.016
Victim						
Alcohol use	99 (11.8)	43 (13.6)	52 (11.1)	4 (7.1)	2.41	.300
Drug use	39 (4.6)	10 (3.2)	26 (5.5)	3 (5.4)	2.47	.291
Fear	97 (11.5)	10 (3.2)	76 (16.2)	11 (19.6)	35.24	< .001
FV/criminal history	212 (25.2)	39 (12.3)	150 (31.9)	23 (41.1)	46.47	< .001
Mental health issues	64 (7.6)	16 (5.1)	41 (8.7)	7 (12.5)	5.65	.059
Suicide/self-harm	24 (2.9)	4 (1.3)	18 (3.8)	2 (3.6)	4.60	.100
Relationship						
Cohabiting	44 (5.2)	17 (5.4)	23 (4.9)	4 (7.1)	0.54	.765
Custody issues	38 (4.5)	16 (4.7)	22 (4.7)	1 (1.8)	1.04	.595
Denial of end of relationship	29 (3.4)	10 (3.2)	18 (3.8)	1 (1.8)	0.75	.688
Financial stress	92 (10.9)	32 (10.1)	56 (11.9)	4 (7.1)	1.50	.472
Gang affiliations	40 (4.8)	6 (1.9)	26 (5.5)	8 (14.3)	17.57	< .001
Housing problems	70 (8.3)	23 (7.3)	40 (8.5)	7 (12.5)	1.76	.416
Infidelity	30 (3.6)	5 (1.6)	23 (4.9)	2 (3.6)	6.03	.049
Isolation	64 (7.6)	19 (6.0)	36 (7.7)	9 (16.1)	6.86	.032
On/off relationship	39 (4.6)	7 (2.2)	29 (6.2)	3 (5.4)	6.76	.034
Property dispute	28 (3.3)	14 (4.4)	14 (3.0)	0 (0.0)	3.30	.192
Protection order	13 (1.5)	3 (0.9)	9 (1.9)	1 (1.8)	1.18	.554
Separation	157 (18.6)	54 (17.1)	98 (20.9)	5 (8.9)	5.50	.064
Situational stress	163 (19.4)	72 (22.8)	84 (17.9)	7 (12.5)	4.73	.094
Unemployment	42 (5)	11 (3.5)	31 (6.6)	0 (0.0)	7.02	.030
Child/family						
Breakdown in family dynamics	29 (3.4)	13 (4.1)	16 (3.4)	0 (0.0)	2.43	.297
Collusive family	6 (0.7)	0 (0.0)	6 (1.3)	0 (0.0)	4.78	.091
Complex family dynamics	21 (2.5)	5 (1.6)	15 (3.2)	1 (1.8)	2.14	.344
History of victimization/cumulative harm	35 (4.2)	2 (0.6)	29 (6.2)	4 (7.1)	15.89	< .001
Pregnancy or recent birth	40 (4.8)	6 (1.9)	30 (6.4)	4 (7.1)	9.16	.010
Non-biological children	79 (9.4)	20 (6.3)	52 (11.1)	7 (12.5)	5.67	.059
Vulnerable children	100 (11.9)	24 (7.6)	66 (14.0)	10 (17.9)	9.56	.008
Vulnerable elderly person	11 (1.3)	3 (0.9)	6 (1.3)	2 (3.6)	2.54	.280

Table 2 Continued

		Risk category f (%)				
	f (%)	*Low*	*Medium*	*High*	χ^2	*p*
Harm						
Controlling behavior	176 (20.9)	40 (12.7)	117 (24.9)	19 (33.9)	23.27	< .001
Electronic/social media harm	25 (3)	5 (1.6)	17 (3.6)	3 (5.4)	3.90	.142
Harassment	9 (1.1)	2 (0.6)	6 (1.3)	1 (1.8)	1.03	.597
Harmed/threatened to harm, intimidated children, family members, or pets	28 (3.3)	1 (0.3)	26 (5.5)	1 (1.8)	16.43	< .001
Injuries	86 (10.2)	13 (4.1)	65 (13.8)	8 (14.3)	20.54	< .001
Intimidating behavior	123 (14.6)	30 (9.5)	87 (18.5)	6 (10.7)	13.05	< .001
Items thrown	42 (5)	8 (2.5)	33 (7.0)	1 (1.8)	9.34	.009
Non-fatal strangulation	43 (5.1)	0 (0.0)	25 (5.3)	18 (32.1)	101.52	< .001
Physical harm	263 (31.2)	47 (14.9)	187 (39.8)	29 (51.8)	66.40	< .001
Property damage	132 (15.7)	40 (12.7)	82 (17.4)	10 (17.9)	3.49	.174
Psychological abuse	33 (3.9)	6 (1.9)	23 (4.9)	4 (7.1)	6.16	.046
Sexual assault	5 (0.6)	0 (0.0)	4 (0.9)	1 (1.8)	3.76	.152
Stalking	31 (3.7)	8 (2.5)	17 (3.6)	6 (10.7)	8.99	.011
Threats	106 (12.6)	22 (7.0)	67 (14.3)	17 (30.4)	26.34	<.001
Verbal abuse	364 (43.2)	175 (55.4)	171 (36.4)	18 (32.1)	30.79	< .001
Weapons	73 (8.7)	13 (4.1)	52 (11.1)	8 (4.3)	13.92	.001
Episode						
Aggressor not spoken to	106 (12.6)	34 (10.8)	68 (14.5)	4 (7.1)	3.98	.137
Breach of protection order	28 (3.3)	4 (1.3)	20 (4.3)	4 (7.1)	7.97	.019
Conflicting stories	19 (2.3)	0 (0.0)	18 (3.8)	1 (1.8)	12.63	.002
Delay in police attendance	30 (3.6)	13 (4.1)	17 (3.6)	0 (0.0)	2.35	.308
Episode occurred in public	61 (7.2)	16 (5.1)	37 (7.9)	8 (14.3)	6.65	.036
Left address/not located	57 (6.8)	14 (4.4)	37 (7.9)	6 (10.7)	5.03	.081
Refusing to leave	40 (4.8)	17 (5.4)	23 (4.9)	0 (0.0)	3.09	.213
Uncooperative with police	88 (10.5)	22 (7.0)	56 (11.9)	10 (17.9)	8.47	.014
Victim not spoken to	26 (3.1)	8 (2.5)	17 (3.6)	1 (1.8)	1.08	.582
Practice considerations						
Cultural barriers	9 (1.1)	1 (0.3)	6 (1.3)	2 (3.6)	5.20	.074
Denial of FV	18 (2.1)	5 (1.6)	9 (1.9)	4 (7.1)	7.28	.026
Duration of episodes	59 (7)	10 (3.2)	44 (9.4)	5 (8.9)	11.48	.003
Escalation	94 (11.2)	19 (5.7)	70 (14.9)	6 (10.7)	16.13	< .001
Frequent episodes	82 (9.7)	9 (2.8)	68 (14.5)	5 (8.9)	29.07	< .001
Intergenerational FV	22 (2.6)	2 (0.6)	19 (4.0)	1 (1.8)	8.79	.012
Lack of information	184 (21.9)	44 (13.9)	124 (26.4)	16 (28.6)	18.76	< .001
Lack of support	23 (2.7)	2 (0.6)	16 (3.4)	5 (8.9)	14.13	.001
No formal orders	39 (4.6)	14 (4.4)	23 (4.9)	2 (3.6)	0.25	.885
Non-engagement	70 (8.3)	10 (3.2)	47 (10.0)	13 (23.2)	29.06	< .001
Normalization	65 (7.7)	6 (1.9)	52 (11.1)	7 (12.5)	24.21	< .001
Unreported FV	141 (16.7)	20 (6.3)	100 (21.3)	21 (37.5)	48.82	< .001

Note. f = frequency; FV = Family Violence; ISR = Integrated Safety Response. See supplemental materials for factor definitions.

These chi square statistics should be interpreted with caution because there is a high chance of false positive results given the large number tests conducted, and we include them simply to illustrate the overall pattern of results.

Table 3 shows relatively few of the factors triage teams recorded were associated with recurrence or physical recurrence. In order from largest to smallest associations, *frequent episodes, financial stress, housing problems, non-engagement, aggressor drug use, controlling behavior, escalation, items thrown, victim FV or criminal history, victim mental health issues,* and *victim drug use* were positively associated with recurrence. On the other hand, *injuries, physical harm,* and *aggressor alcohol use* were negatively associated with recurrence. *Frequent episodes, financial stress, denial of FV, episode occurred in public, items thrown, on/off relationship, housing problems, escalation, lack of information, non-engagement, normalization, unemployment,* and *victim suicide or self-harm* were positively associated with physical recurrence, whereas *aggressor alcohol use* and *custody issues* were negatively associated with physical recurrence.

Next, we examined the predictive validity of the ISR triage teams' risk categories. Table 4 shows that overall, the risk categories predicted recurrence and physical recurrence within the six-month follow-up period better than chance, but the AUCs were small in magnitude (Rice & Harris, 2005). Table 4 also shows that the risk categories predicted recurrence or physical recurrence for all IPV cases, and for cases between current partners; as well as for aggressors who were male, aged above 18 years, or European. Further, the risk categories predicted recurrence for cases between former partners, and when aggressors were aged below 18 years; and predicted physical recurrence for cases where parents were aggressors towards their children. No other AUCs were statistically significant.

Table 3

Correlations Between the ISR Factors and Recurrence and Physical Recurrence

	Recurrence		*Physical Recurrence*	
	r [95% CI]	*p*	*r* [95% CI]	*p*
Aggressor				
Alcohol use	-.07 [-.14, .00]	.047	-.07 [-.13, .00]	.048
Breaching/non-complying with conditions/orders	.06 [-.01, .12]	.097	.03 [-.04, .10]	.398
Drug use	.08 [.01, .14]	.027	.01 [-.05, .08]	.691
FV/criminal history	.03 [-.04, .10]	.414	-.03 [-.10, .04]	.367
Mental health issues	.04 [-.03, .11]	.250	-.01 [-.08, .05]	.694
Suicide/self-harm	.04 [-.03, .11]	.226	-.03 [-.10, .03]	.322
Victim				
Alcohol use	-.04 [-.11, .03]	.219	-.03 [-.10, .04]	.361
Drug use	.07 [.00, .13]	.049	.03 [-.04, .09]	.433
Fear	.04 [-.03, .11]	.262	-.02 [-.09, .05]	.586
FV/criminal history	.08 [.02, .15]	.016	.03 [-.04, .10]	.412
Mental health issues	.08 [.01, .14]	.028	.05 [-.01, .12]	.125
Suicide/self-harm	.03 [-.04, .09]	.428	.07 [.00, .14]	.039
Relationship				
Cohabiting	.02 [-.05, .09]	.604	.01 [-.06, .08]	.729
Custody issues	-.03 [-.10, .04]	.395	-.07 [-.14, .00]	.047
Denial of end of relationship	-.02 [-.09, .05]	.597	-.03 [-.10, .04]	.353
Financial stress	.14 [.07, .21]	< .001	.14 [.07, .21]	< .001
Gang affiliations	.02 [-.05, .08]	.622	.02 [-.04, .09]	.488
Housing problems	.09 [.02, .16]	.008	.07 [.00, .14]	.044
Infidelity	-.01 [-.08, .06]	.749	-.02 [-.09, .05]	.582
Isolation	.07 [.00, .13]	.053	.01 [-.06, .08]	.804
On/off relationship	.06 [-.01, .12]	.102	.08 [.02, .15]	.016
Property dispute	.01 [-.05, .08]	.682	.00 [-.06, .07]	.919
Protection order	.02 [-.05, .09]	.578	-.06 [-.13, .00]	.064
Separation	.00 [-.07, .06]	.925	.03 [-.03, .10]	.320
Situational stress	.03 [-.04, .10]	.412	-.03 [-.09, .04]	.428
Unemployment	.06 [-.01, .13]	.076	.07 [.00, .14]	.038
Child/family				
Breakdown in family dynamics	.01 [-.06, .08]	.820	-.05 [-.12, .02]	.163
Collusive family	.01 [-.06, .07]	.852	-.04 [-.11, .02]	.210
Complex family dynamics	-.04 [-.11, .03]	.232	-.06 [-.13, .00]	.068
History of victimization/cumulative harm	.03 [-.03, .10]	.328	.03 [-.04, .09]	.452
Pregnancy or recent birth	.05 [-.02, .12]	.142	.00 [-.07, .06]	.915
Non-biological children	.01 [-.06, .08]	.722	.05 [-.02, .11]	.173
Vulnerable children	.03 [-.04, .10]	.417	-.02 [-.09, .04]	.484
Vulnerable elderly person	.02 [-.05, .09]	.577	.04 [-.02, .11]	.196

Table 3 Continued

	Recurrence		*Physical Recurrence*	
	r [95% CI]	*p*	*r* [95% CI]	*p*
Harm				
Controlling behavior	.08 [.01, .15]	.020	.00 [-.07, .06]	.938
Electronic/social media harm	.02 [-.05, .09]	.555	-.05 [-.12, .01]	.113
Harassment	.00 [-.07, .06]	.916	-.05 [-.12, .01]	.124
Harmed/threatened to harm, intimidated children, family members, or pets	-.04 [-.11, .03]	.258	-.03 [-.10, .04]	.397
Injuries	-.10 [-.17, -.03]	.004	-.03 [-.09, .04]	.437
Intimidating behavior	.00 [-.07, .07]	.973	.02 [-.05, .09]	.534
Items thrown	.07 [.00, .14]	.036	.10 [.03, .17]	.004
Non-fatal strangulation	-.01 [-.08, .06]	.786	.00 [-.07, .07]	.965
Physical harm	-.08 [-.15, -.01]	.021	.02 [-.05, .08]	.627
Property damage	.01 [-.06, .07]	.847	-.04 [-.11, .03]	.217
Psychological abuse	.02 [-.05, .09]	.533	-.01 [-.08, .06]	.720
Sexual assault	-.01 [-.08, .06]	.781	.00 [-.07, .07]	.971
Stalking	-.02 [-.08, .05]	.628	-.02 [-.09, .05]	.526
Threats	.01 [-.05, .08]	.673	-.02 [-.08, .05]	.625
Verbal abuse	-.01 [-.08, .06]	.763	.02 [-.05, .08]	.634
Weapons	-.01 [-.08, .05]	.672	.01 [-.06, .08]	.782
Episode				
Aggressor not spoken to	.01 [-.06, .07]	.831	.00 [-.07, .07]	.981
Breach of protection order	.07 [.00, .13]	.051	-.01 [-.08, .05]	.709
Conflicting stories	.05 [-.02, .12]	.134	.06 [-.01, .13]	.078
Delay in police attendance	.05 [-.01, .12]	.123	.06 [-.01, .13]	.081
Episode occurred in public	.04 [-.03, .10]	.309	.12 [.05, .18]	.001
Left address/not located	.02 [-.05, .08]	.647	-.01 [-.08, .06]	.792
Refusing to leave	-.02 [-.08, .05]	.631	-.05 [-.11, .02]	.192
Uncooperative with police	.05 [-.02, .12]	.152	.06 [-.01, .12]	.106
Victim not spoken to	-.01 [-.08, .05]	.686	.01 [-.06, .08]	.758
Practice considerations				
Cultural barriers	.00 [-.07, .06]	.916	.06 [-.01, .13]	.077
Denial of FV	.04 [-.02, .11]	.200	.13 [.06, .19]	< .001
Duration of episodes	.03 [-.03, .10]	.311	.04 [-.02, .11]	.205
Escalation	.07 [.00, .14]	.036	.07 [.00, .14]	.041
Frequent episodes	.22 [.15, .28]	< .001	.23 [.16, .29]	< .001
Intergenerational FV	.03 [-.04, .09]	.427	.01 [-.06, .08]	.809
Lack of information	.04 [-.03, .11]	.243	.07 [.00, .14]	.040
Lack of support	.01 [-.06, .07]	.874	-.05 [-.12, .02]	.151
No formal orders	.02 [-.05, .09]	.515	.04 [-.03, .11]	.234
Non-engagement	.09 [.02, .16]	.008	.07 [.00, .14]	.044
Normalization	.04 [-.03, .10]	.304	.07 [.00, .14]	.036
Unreported FV	-.01 [-.07, .06]	.833	.02 [-.05, .09]	.515

Note. FV = Family Violence; ISR = Integrated Safety Response. See supplemental materials for factor definitions.

Table 4

AUCs for the ISR Risk Categories Predicting Recurrence and Physical Recurrence, by Aggressor-Victim Relationship, Aggressor Characteristics, and ISR Area

		Recurrence			*Physical Recurrence*		
	n	*f* (%)	AUC [95% CI]	*p*	*f* (%)	AUC [95% CI]	*p*
Overall	842	389 (46.2)	.60 [.56, .63]	< .001	174 (20.1)	.60 [.56, .65]	< .001
Relationship to victim							
IPV	595	292 (49.1)	.61 [.56, .65]	< .001	139 (23.4)	.60 [.55, .65]	< .001
Current partner	418	204 (48.8)	.60 [.54, .65]	< .001	105 (25.1)	.61 [.54, .67]	.001
Former partner	177	88 (49.7)	.63 [.55, .71]	.002	34 (19.2)	.58 [.48, .69]	.125
Non-IPV	247	97 (39.3)	.54 [.47, .61]	.285	35 (14.2)	.57 [.47, .67]	.184
Parent	42	10 (23.8)	.58 [.38, .79]	.409	3 (7.1)	.73 [.52, .94]	.030
Child	108	51 (47.2)	.54 [.43, .65]	.493	19 (17.6)	.62 [.49, .75]	.078
Sibling	51	23 (45.1)	.61, .45, .77]	.178	8 (15.7)	.51 [.29, .73]	.959
Other	46	13 (28.2)	.41 [.23, .59]	.338	5 (10.9)	.37 [.13, .61]	.285
Aggressor characteristic							
Male	637	313 (49.1)	.60 [.55, .64]	< .001	144 (22.6)	.61 [.56, .66]	< .001
Female	203	75 (36.9)	.57 [.48, .65]	.117	29 (14.3)	.54 [.43, .65]	.486
Below 18 years	67	29 (43.3)	.64 [.51, .78]	.034	14 (20.9)	.57 [.41, .73]	.370
More than 18 years	775	360 (46.5)	.59 [.55, .63]	< .001	160 (20.6)	.60 [.56, .65]	< .001
Māori	373	199 (53.4)	.55 [.49, .61]	.107	93 (24.9)	.55 [.48, .62]	.155
European	360	159 (44.2)	.59 [.53, .65]	.004	64 (17.8)	.61 [.54, .69]	.003
ISR area							
Waikato	480	232 (48.3)	.58 [.53, .63]	.001	104 (21.7)	.59 [.53, .65]	.004
Canterbury	362	157 (43.4)	.62 [.56, .68]	< .001	70 (19.3)	.62 [.55, .69]	.001

Note. AUC = Area Under the Curve; CI = Confidence Interval; *f* = frequency; ISR = Integrated Safety Response; IPV = Intimate Partner Violence.

Notably, the risk categories were no better than chance at predicting recurrence or physical recurrence for aggressors who were women or Māori; and for non-IPV cases, particularly cases between siblings or people with other types of familial relationships. We speculated that one reason why the risk categories performed better for IPV cases than non-IPV cases was because IPV cases may be more likely to recur with the same victim than non-IPV cases, meaning the victim-related factors recorded in assessments remain relevant. Indeed, of the 292 IPV cases with a recurrence, 266 (91.1%) of the first recurrences recorded were again IPV episodes, with most of those (87.7%) involving the same victim as the index episode. In contrast, of the 97 non-IPV cases with a recurrence, 76 (81.4%) were again non-IPV episodes, but with only 47 (48.5%) involving the same victim.

Discussion

In this study we sought to better understand both the process and the predictive validity of the ISR triage teams' risk assessments. To fulfil this aim we first documented the factors triage teams listed during their risk assessments and examined the proportions of those factors across the risk categories the triage teams assigned. Overall, most of the 71 factors were associated with the risk categories, with many factors being more common among higher-risk cases than low-risk cases. This result may suggest the factors triage teams record are the basis for their risk categorizations, but because factors were recorded after the categorization, it could also suggest teams retrospectively assign a greater number of risk factors to cases in higher risk categories.

We then examined the association between these factors and two outcomes — recurrence and physical recurrence — to discern whether the identified factors can be used to predict the likelihood of future FV (Bonta & Andrews, 2016). We found that *frequent episodes, financial stress, items thrown, housing problems, escalation,* and *non-engagement*

were risk factors for recurrence and physical recurrence, suggesting they may be particularly important. For example, when these factors cooccur they may indicate a particularly volatile family environment in need of intervention to prevent further harm. Moreover, for recurrence, *aggressor* and *victim drug use, victim FV or criminal history, victim mental health issues* and *controlling behavior* were significant predictors. There were also several significant predictors for physical recurrence; *denial of FV, on/off relationship, unemployment, lack of information, normalization,* and *episode occurred in public*. But overall, fewer than half of the factors triage teams recorded were associated with recurrence or physical recurrence, in line with previous research showing that many of the items in risk assessment instruments did not predict violence (Coid et al., 2011). The findings from this study also support other research suggesting that specific outcomes may have distinct risk factors (Capaldi et al., 2019; Robinson et al., 2018).

Some of the factors triage teams recorded had negative associations with recurrence or physical recurrence, suggesting they could be protective factors or act as proxies for other mechanisms that prevented recurrence (Eisenberg et al., 2022). For physical recurrence, there were several factors with negative associations, including *injuries* and *physical harm*, which seem counterintuitive here. We speculate that their presence in the index episode may have led to ISR agencies to prioritize these cases for interventions that prevented recurrence. In addition, one of the possibilities for why *aggressor alcohol use* was negatively associated with physical recurrence may be that it was a situational factor that helped to precipitate the index episode and was not a stable characteristic of the aggressor (whereas chronic substance use is a well-established risk factor for FV; Capaldi et al., 2019; Timshel et al., 2017). Furthermore, the factor *custody issues* was a unique protective factor for physical recurrence, with one explanation being that this factor is commonly present in cases where reported episodes are limited to verbal disagreements over parenting arrangements.

Many of the factors recorded in the triage teams' assessments may add value to those assessments by highlighting barriers to responsivity and assisting practitioners to deliver interventions in ways that maximize engagement and behavior change (Bonta & Andrews, 2016, Polaschek, 2012; Polaschek, 2016), in addition to (or rather than) contributing to the assessments' ability to predict recurrence (Butters et al., 2021; Kropp, 2008). For example, *non-engagement* could indicate a family or whānau has not engaged with FV services in the past, causing practitioners to consider alternative ways to foster that engagement. Some factors that may also measure responsivity (e.g., *unreported FV, lack of support*) were more common in cases with higher risk categorizations, perhaps due to concerns about managing those cases successfully; high-risk cases tend to be more difficult to manage (Tomkins, 2020).

Finally, we examined the discriminative ability of the ISR triage teams' risk categories and found they predicted recurrence or physical recurrence better than chance with small effect sizes (Rice & Harris, 2005). These AUCs were comparable to the average predictive ability of risk categorizations made using SPJ risk assessment instruments reported in an international meta-analysis of instruments used to predict IPV (mean AUC of .58; van der Put et al., 2019). However, when we investigated the predictive ability of the risk categories for different sub-groups based on aggressor-victim relationship and aggressor characteristics, we found they performed poorly for aggressors who were Māori, or were women, and for non-IPV cases. This result echoes previous research on the predictive validity of New Zealand Police's FV risk assessment instruments, which also performed poorly for women, Māori and non-IPV cases of FV (Jolliffe Simpson et al., 2021).

The finding that the risk categories performed more poorly for non-IPV cases than IPV cases was not surprising. By far the most dominant and widely researched form of FV is physical IPV within heterosexual adult couples, with women as the primary victims; this was

also the most common type of FV in our sample. While coding factors, we noticed the triage teams' assessments were geared towards identifying risk factors for IPV. The risk assessment guide exemplifies this, with an entire section devoted to 'relationship risks' (Integrated Safety Response, 2018). Indeed, some of the risk factors that triage teams recorded only applied to IPV (e.g., *infidelity, on/off relationship*). This characteristic of the assessments is not necessarily bad; but it suggests that research should test existing risk factors' predictive validity for non-IPV types of FV and answer the question of whether those cases should be assessed using procedures created for IPV cases. Moreover, different types of FV tend to co-occur (Chan et al., 2021); there are additional challenges inherent in assessing risk and responding to families experiencing both IPV *and* violence against other family members, compared with families experiencing IPV *or* non-IPV.

That the risk categories performed more poorly when aggressors were women or Māori may be grounds for concern. This result is also unsurprising given most research on risk assessment for FV focuses on men who commit IPV, and there is less research validating risk assessment instruments for women (van der Put et al., 2019). Moreover, there is a lack of research about culturally specific risk factors (Ashford et al., 2022; Mallory et al., 2016), and few instruments have been validated for specific cultural contexts (e.g., Chan, 2012). The poor performance for Māori is especially problematic when considering the ISR's special commitment to being responsive to Māori communities and including Indigenous service providers (Mossman et al., 2017). However, because we could not account for the level, type or effectiveness of interventions received after the index episode, it is also possible that interventions were sufficiently effective for these groups to statistically decouple the risk categorizations from the outcomes we measured. Therefore, we need to understand more about what interventions families and whānau receive, and how those services are provided, including whether provision adheres to the RNR principles of effective correctional treatment

(Mossman et al., 2017). This examination will enable us to more clearly understand how the triage teams' assessments inform responses for different groups of people experiencing FV and the extent to which those responses moderate the relationship between the risk categories and recurrence.

Limitations

A clear limitation of this study is how risk factors were recorded. Triage teams freely listed the risk factors they considered relevant to each case, rather than systematically scoring the presence or absence of risk factors from the risk assessment guide. Our observations of the triage team meetings around the time these data were collected suggested that most risk factors mentioned in oral discussion by triage team members were recorded in the free-text field in the Family Safety System database; but the teams did not systematically work through and acknowledge or discard each of the risk factors included in the guide. As a result, assessments may not have captured the full range of risk factors present for each case. Because triage teams typically recorded risk factors *after* determining the risk category for the case, they may have retrospectively chosen risk factors to justify their categorization, and while there was an association between the risk categories and many of the factors, we could not rule out this alternative explanation. This characteristic of their risk assessment process suggests the triage teams' risk assessment process is relatively unstructured, even though they have a guide that is intended to impose some structure.

In addition, the ISR's recording procedure meant that to generate data for research we coded the risk factors for each case from a string of text, adding a layer of interpretation and potentially introducing further error that we sought to minimize by measuring inter-rater reliability. Because the triage teams did not score items but simply noted their presence, we also could not infer items' weightings. Many SPJ risk assessment instruments (e.g., HCR-20; Douglas et al., 2013; SARA; Kropp & Hart, 2015) allow practitioners to weight items

(usually from 0-2), indicating the degree to which they are relevant and present for the case in question. Some research shows risk factors indicating the potential for serious harm such as weapons, injuries, escalation, and fear may be weighted more strongly by practitioners when they assess risk, compared with other risk factors like custody issues, unemployment, and non-biological children (Perez Trujillo & Ross, 2008; Robinson & Howarth, 2012; Robinson et al., 2018). This weighting of factors is an important aspect of risk assessments that we could not investigate here. Increased structure in the ISR's risk assessment protocol such as requiring triage teams to formally identify risk factors from the risk assessment guide before their categorization may increase transparency of their decision making and ensure risk factors are used as the basis for risk categorizations. Moreover, greater transparency would enable formal validation of the triage teams' risk assessments and may help identify which factors triage teams consider most strongly in their assessments.

We also could not account for interventions following the index episode that may have prevented FV recurrence, and thus moderated the association between the categorizations and recurrence. For example, a high risk family could engage intensively with effective services and reduce their level of risk substantially. By contrast, a family at medium risk may not be offered or engage with any services. As a result, the initially high risk family may now be at lower risk than the medium risk family, weakening the relationship between the pre-intervention risk category and FV outcomes. The ISR was developed to coordinate interventions delivered to families experiencing FV and all cases are assigned at least one intervention (Integrated Safety Response, 2019), although not all families engage with what is offered. Thus, further research that accounts for the number, intensity, and effectiveness of those interventions is necessary before we can fairly evaluate the triage teams' risk categorizations.

Implications

The results of this research indicate the ISR triage teams' risk assessment protocol may benefit from increased structure and validation. We were not able to investigate the predictive validity of the guide create for triage teams' assessments because it was not used systematically. Therefore, encouraging practitioners to complete the guide may help improve their assessments, by a) helping them to systematically consider the presence of all risk factors in the guide, rather than just those that come to mind, b) creating data that can be used to validate the risk factors more rigorously than was possible here, and c) producing weightings that indicate which factors contributed most strongly to their risk categorization. This information would lend itself to the systematic identification of a subset of risk factors associated with recurrence, and an optimal algorithm for combining those factors; thus creating an improved model to guide ISR triage teams' decision making.

Conclusion

In this study we documented and examined the ISR triage teams' factors and risk categorizations for 842 cases, to better understand the both the process and the predictive validity of the teams' risk assessments. Taken together, the results indicate ISR triage teams consider a wide array of factors when assessing risk in a relatively unstructured way. Moreover, many of the factors triage teams recorded were not associated with recurrence or physical recurrence, and the risk categories' predictive validity varied across sub-groups, performing particularly poorly when aggressors were Māori, or women, and for non-IPV cases. Therefore, the triage teams' risk assessment protocol may benefit from increased structure and validation. Future research should also measure the interventions cases are allocated and engage with, to find whether those interventions mitigate risk for the recurrence of FV.

References

Ashford, L. J., Spivak, B. L., Ogloff, J. R. P., & Shepherd, S. M. (2022). The cross-cultural fairness of the LS/RNR: An Australian analysis. *Law and Human Behavior*, *46*, 214-226. https://doi.org/10.1037/lhb0000486

Bonta, J., & Andrews, D. A. (2016). *The Psychology of Criminal Conduct* (6th ed.). Routledge.

Butters, R. P., Droubay, B. A., Seawright, J. L., Tollefson, D. R., Lundahl, B., & Whitaker, L. (2021). Intimate partner violence perpetrator treatment: Tailoring interventions to individual needs. *Clinical Social Work Journal*, *49*(3), 391-404. https://doi.org/10.1007/s10615-020-00763-y

Capaldi, D. M., Low, S., Tiberio, S. S., & Shortt, J. W. (2019). Intimate partner violence across the lifespan: Dyadic theory and risk and protective factors. In R. Geffner, V. Vieth, V. Vaughan-Eden, A. Rosenbaum, L. Hamberger, & J. White (Eds.), *Handbook of Interpersonal Violence Across the Lifespan* (pp. 1-25). Cham: Springer.

Chan, K. L. (2012). Predicting the risk of intimate partner violence: The Chinese risk assessment tool for victims. *Journal of Family Violence*, *27*(2), 157-164. https://doi.org/10.1007/s10896-012-9418-4

Chan, K. L., Chen, Q., & Chen, M. (2021). Prevalence and correlates of the co-occurrence of family violence: A meta-analysis on family polyvictimization. *Trauma, Violence and Abuse*, *22*(2), 289-305. https://doi.org/10.1177/1524838019841601

Coid, J. W., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Farrington, D., & Rogers, R. (2011). Most items in structured risk assessment instruments do not predict violence. *Journal of Forensic Psychiatry & Psychology*, *22*(1), 3-21. https://doi.org/10.1080/14789949.2010.495990

Dankert, E. W., & Johnson, K. (2014). *Risk assessment validation: A prospective study*. Retrieved from http://www.evidentchange.org/sites/default/files/publication_pdf/risk-assessment-validation.pdf

Dobbs, T., & Eruera, M. (2014). *Kaupapa Māori wellbeing framework: The basis for whānau violence prevention and intervention*. New Zealand Family Violence Clearinghouse.

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20 Assessing Risk for Violence: User Guide*. Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology Public Policy and Law*, *11*, 347-383. https://doi.org/10.1037/1076-8971.11.3.347

Eisenberg, M. J., van Horn, J. E., van der Put, C. E., Stams, G. J. J. M., & Hendriks, J. (2022). Protective factors as uni- or bipolar factors and their incremental validity and accuracy in predicting general recidivism. *International Journal of Law and Psychiatry*, *81*, 101772. https://doi.org/https://doi.org/10.1016/j.ijlp.2021.101772

Garrington, C., & Boer, D. P. (2020). Structured professional judgement in violence risk assessment. In J. S. Wormith, L. A. Craig, & T. E. Hogue (Eds.), *The Wiley Handbook of What Works in Violence Risk Management* (1st ed., pp. 145-162): John Wiley & Sons Ltd.

Geurts, R., Raaijmakers, N., Delsing, M. J. M. H., Spapens, T., Wientjes, J., Willems, D., & Scholte, R. H. J. (2021). Assessing the risk of repeat victimization using structured and unstructured police information. *Crime & Delinquency*, 001112872110475. https://doi.org/10.1177/00111287211047533

Hart, S. D., Douglas, K. S., & Guy, L. S. (2016). The structured professional judgement approach to violence risk assessment: Origins, nature, and advances. In D. P. Boer, A.

R. Beech, T. Ward, L. A. Craig, M. Rettenberger, L. E. Marshall, & W. L. Marshall (Eds.), *The Wiley Handbook on the Theories, Assessment, and Treatment of Sexual Offending* (pp. 643–666). Wiley Blackwell. https://doi.org/10.1002/9781118574003.wattso030

Helmus, L. M., & Babchishin, K. M. (2016). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior, 44*(1), 8–25. https://doi.org/10.1177/0093854816678898

Henning, K., Campbell, C. M., Stewart, G., & Johnson, J. (2021). Prioritizing police investigations of intimate partner violence using actuarial risk assessment. *Journal of Police and Criminal Psychology*. https://doi.org/10.1007/s11896-021-09466-7

Hilton, N. Z., & Eke, A. W. (2017). Assessing risk of intimate partner violence. In J. C. Campbell & J. T. Messing (Eds.), *Assessing Dangerousness: Domestic Violence Offenders and Child Abusers*. Springer. https://doi.org/10.1891/9780826133274.0006

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist, 34*, 400-409. https://doi.org/10.1177/0011000005285877

Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., & Lines, K. (2004). A brief accurate actuarial assessment for the prediction of wife assault recidivism: The Ontario Domestic Assault Risk Assessment. *Psychological Assessment, 16*(3), 267-275. https://doi.org/10.1037/1040-3590.16.3.267

Integrated Safety Response. (2018). *Risk Assessment Quick Guide*. Wellington, New Zealand

Integrated Safety Response. (2019). *12 week review of cases referred to the Integrated Safety Response (ISR) pilot*. https://www.police.govt.nz/sites/default/files/publications/isr-12-week-case-review-report.pdf

Jolliffe Simpson, A. D., Joshi, C., & Polaschek, D. L. L. (2021). Predictive validity of the DYRA and SAFVR: New Zealand Police's family violence risk assessment instruments. *Criminal Justice and Behavior*, *48*(10), 1487-1508. https://doi.org/10.1177/0093854821997525

Koo, T. K., & Li, M. Y. (2015). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*. https://doi.org/10.1016/j.jcm.2017.10.001

Kropp, P. R. (2004). Some questions regarding spousal assault risk assessment. *Violence Against Women*, *10*, 676-697. https://doi.org/10.1177/1077801204265019

Kropp, P. R. (2008). Intimate partner violence risk assessment and management. *Violence and Victims*, *23*(2), 202-220. https://doi.org/10.1891/0886-6708.23.2.202

Kropp, P. R., & Hart, S. D. (2015). SARA-V3: User manual for version 3 of the spousal assault risk assessment guide. Proactive Resolutions.

Mallory, A. B., Dharnidharka, P., Deitz, S. L., Barros-Gomes, P., Cafferky, B., Stith, S. M., & Van, K. (2016). A meta-analysis of cross-cultural risk markers for intimate partner violence. *Aggression and Violent Behavior*, *31*, 116-126. https://doi.org/10.1016/j.avb.2016.08.004

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. Sexual Abuse: A Journal of Research & Treatment, 22, 191-217. https://doi.org/10.1177/1079063210366039

Messing, J. T., Campbell, J. C., & Snider, C. (2017). Validation and adaptation of the danger assessment-5: A brief intimate partner violence risk assessment. *Journal of Advanced Nursing*, *73*, 3220-3230. https://doi.org/10.1111/jan.13459

Mossman, E., Paulin, J., & Wehipeihana, N. (2017). *Evaluation of the family violence Integrated Safety Response pilot*. Wellington, New Zealand: SUPERU Retrieved from https://thehub.sia.govt.nz/assets/documents/ISR_pilot_evaluation_FINAL.pdf

Perez Trujillo, M., & Ross, S. (2008). Police response to domestic violence. *Journal of Interpersonal Violence*, *23*(4), 454-473. https://doi.org/10.1177/0886260507312943

Polaschek, D. L. L. (2012). An appraisal of the risk-need-responsivity (RNR) model of offender rehabilitation and its application in correctional treatment. *Legal and Criminological Psychology*, *17*(1), 1–17. https://doi.org/10.1111/j.2044-8333.2011.02038.x

Polaschek, D. L. L. (2016). *Responding to perpetrators of family violence.* New Zealand Family Violence Clearinghouse. https://nzfvc.org.nz/sites/nzfvc.org.nz/files/NZFVC-issues-paper-11-responding-perpetrators.pdf

Prins, S. J., & Reich, A. (2021). Criminogenic risk assessment: A meta-review and critical analysis. *Punishment & Society*, *23*(4), 578-604. https://doi.org/10.1177/14624745211025751

R Core Team. (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. http://www.R-project.org/

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior*, *29*, 615-620. https://doi.org/10.1007/s10979-005-6832-7

Richards, L. (2009). *Domestic Abuse, Stalking and Harassment and Honour Based Violence (DASH, 2009) Risk Identification and Assessment and Management Model*. https://www.dashriskchecklist.co.uk/wp-content/uploads/2016/09/DASH-2009.pdf

Robinson, A. L. & Clancy, A. (2020). Systematically identifying and prioritising domestic abuse perpetrators for targeted intervention. *Criminology & Criminal Justice*, 174889582091438. https://doi.org/10.1177/1748895820914380

Robinson, A. L. & Howarth, E. (2012). Judging risk. *Journal of Interpersonal Violence, 27*(8), 1489-1518. https://doi.org/10.1177/0886260511425792

Robinson, A. L., Pinchevsky, G. M. & Guthrie, J. A. (2018). A small constellation: Risk factors informing police perceptions of domestic abuse. *Policing and Society*, *28*(2), 189-204. https://doi.org/10.1080/10439463.2016.1151881

Spencer, C. M., Stith, S. M., & Cafferky, B. (2019). Risk markers for physical intimate partner violence victimization: A meta-analysis. *Aggression and Violent Behavior*, *44*, 8-17. https://doi.org/10.1016/j.avb.2018.10.009

Spivak, B., McEwan, T., Luebbers, S. & Ogloff, J. (2020). Implementing evidence-based practice in policing family violence: The reliability, validity and feasibility of a risk assessment instrument for prioritising police response. *Policing and Society*. https://doi.org/10.1080/10439463.2020.1757668

Stanley, N. & Humphreys, C. (2014). Multi-agency risk assessment and management for children and families experiencing domestic violence. *Children and Youth Services Review*, *47*, 78-85. https://doi.org/10.1016/j.childyouth.2014.06.003

Stansfield, R. & Williams, K. R. (2014). Predicting family violence recidivism using the DVSI-R: Integrating survival analysis and perpetrator characteristics. *Criminal Justice and Behavior*, *41*, 163-180. https://doi.org/10.1177/0093854813500776

Timshel, I., Montgomery, E., & Dalgaard, N. T. (2017). A systematic review of risk and protective factors associated with family related violence in refugee families. *Child Abuse & Neglect, 70*, 315-330. https://doi.org/10.1016/j.chiabu.2017.06.023

Tomkins, J. (2020). *High-risk victims of intimate partner violence within the Integrated Safety Response pilot: An examination of psychosocial stressors and repeat victimisation.* University of Waikato. https://researchcommons.waikato.ac.nz/handle/10289/14141

Turner, E., Medina, J., & Brown, G. (2019). Dashing hopes? The predictive accuracy of domestic abuse risk assessment by police. *The British Journal of Criminology, 59*(5), 1013-1034. https://doi.org/10.1093/bjc/azy074

van der Put, C. E., Assink, M., & Stams, G. J. J. M. (2016). Predicting relapse of problematic child-rearing situations. *Children and Youth Services Review, 61*, 288-295. https://doi.org/10.1016/j.childyouth.2016.01.002

van der Put, C. E., Gubbels, J., & Assink, M. (2019). Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools. *Aggression and Violent Behavior, 47*, 100-116. https://doi.org/10.1016/j.avb.2019.03.008

Walklate, S., & Mythen, G. (2011). Beyond risk theory: Experiential knowledge and 'knowing otherwise'. *Criminology & Criminal Justice, 11*, 99-113. https://doi.org/10.1177/1748895811398456

Chapter Four: Latent Classes Among Family Violence Cases

Integrated Safety Response (ISR) triage teams routinely conduct risk assessments for Family Violence (FV) cases, and during these assessments they record various factors to support the risk category they allocate (low, medium, or high). These factors are purportedly selected because they are risk factors that can be used to predict recurrence of FV. But, in addition to risk assessments' primary purpose of estimating risk, those assessments are also an opportunity for practitioners to identify information that helps them deliver services in an accessible and engaging way. Therefore, despite being identified during risk assessments, the factors ISR triage teams record may add value to the assessments by highlighting barriers to responsivity and informing case management in addition to (or rather than) contributing to the assessments' ability to predict recurrence (Jolliffe Simpson et al., 2023). Accordingly, in this study we use latent class analysis of the ISR triage teams' recorded factors to identify and describe the types of cases the ISR manages. These classes may have implications for families' treatment and case management, including how practitioners can maximize engagement and behavior change (Bonta & Andrews, 2016).

Considering Risk, Need, and Responsivity

The Risk Need Responsivity (RNR) model for correctional treatment puts forward that risk factors are measurable characteristics that can be used to predict the likelihood a person will engage in criminal behavior (Bonta & Andrews, 2016; Polaschek, 2012). This principle is embedded in structured risk assessments that orient practitioners towards factors empirically related to the outcome of interest (e.g., recurrence of FV). It is well established that guiding practitioner decisions in this way improves the predictive validity of their risk assessments (Garrington & Boer, 2020; Svalin & Levander, 2019; van der Put et al., 2019). Hence, risk assessments are a salient procedure for FV practitioners who are often

responsible for important decisions about safety and are required to allocate limited resources to the people who need them most (Robinson, 2006; Robinson & Clancy, 2020; Robinson & Howarth, 2012).

Performing risk assessments entails identifying risk factors and allocating risk categories (e.g., low, medium, or high), that subsequently inform case management decisions like the type and intensity of treatment services offered to aggressors, victims, and whānau[10]. The RNR model's risk principle explains that risk level should dictate treatment intensity, because higher risk cases are more likely to benefit from treatment than lower risk cases, and low risk cases sometimes fare worse after receiving unnecessary treatment (Bonta & Andrews, 2016; Polaschek, 2016). In addition, the RNR model's need principle states that treatment targets should align with the changeable characteristics currently contributing to a person's offending (e.g., employment status; Bonta & Andrews, 2016). Taken together, these two principles summarize why the risk factors and risk categories from risk assessments inform decisions about the type and intensity of treatments people receive.

Yet there is a third, equally important RNR principle that elucidates how risk assessments inform case management. The responsivity principle of the RNR model states that treatment services should be delivered in a way that maximizes people's willingness to engage and ability to benefit from them (Bonta & Andrews, 2016). For example, if a person does not have access to a vehicle, they should be offered services they can feasibly reach by other means, or that can be brought to them. Responsivity is divided into general responsivity, or the overarching principles of effective treatment; and specific responsivity, which captures individual factors that may prevent people from engaging in treatment that practitioners should endeavor to accommodate (e.g., motivation to receive treatment,

[10] Whānau is the Te Reo Māori (Māori language) word for extended family or a family-like group.

cognitive ability, personality, and maturity; Bonta & Andrews, 2016; Eisenberg et al., 2022). Performing risk assessments involves reviewing case information, identifying factors that may be linked to recurrence, and visualizing scenarios where the potential for harm may arise. While designed to structure estimations of risk; this process can also help practitioners plan their response by prompting them to anticipate factors that could act as barriers to responsivity (Eisenberg et al., 2022; Kropp, 2008).

The Integrated Safety Response (ISR)

Practitioners working in the ISR regularly conduct risk assessments and make decisions about how to manage FV cases. The ISR is a multi-agency[11] collaboration of governmental and Non-Governmental Organizations (NGOs) that coordinates service provision for FV cases in two areas of New Zealand. Māori (Indigenous New Zealanders) are disproportionately affected by FV (Dobbs & Eruera, 2014), and outcomes for Māori are an important indicator of the ISR's success. Accordingly, many ISR agencies employ Kaupapa Māori approaches in their service provision (i.e., practices that are Māori-led and based in a Māori worldview; Mossman et al., 2017). The ISR triage teams' risk assessments are also distinctive because they consider information at the level of the whānau or family unit, rather than focusing on the individual, as is the norm in correctional treatment settings (Bonta & Andrews, 2016; Integrated Safety Response, 2018). Representatives from the organizations involved in ISR make up triage teams that share and review information about cases with recent FV episodes reported to police, then select a risk category (low, medium, or high), before listing factors to support that categorization, and allocating interventions to prevent further FV (Integrated Safety Response, 2018).

[11] The structure and purpose of the ISR mirrors many other international multi-agency teams (e.g., Contini & Wilson, 2019; Robinson & Tregidga, 2007; Shorrock et al., 2020).

In a previous study we examined the predictive validity of ISR triage teams' risk factors and found fewer than half of the factors the triage teams recorded were related to FV recurrence (Jolliffe Simpson et al., 2023). In considering the wide array of factors, we posited that some of those factors may add value to the risk assessments by informing case management in addition to (or rather than) contributing to the risk assessments' ability to predict recurrence (Jolliffe Simpson et al., 2023). After all, at the conclusion of each assessment triage teams allocate intervention services and nominate an organization to lead the family's case management according to their identified needs (Integrated Safety Response, 2018). This means that engaging with families is at the forefront of team members' minds when they perform these assessments, and it logically follows that some of the factors they record could reflect their consideration of specific needs or barriers to responsivity.

In the same study we also found the predictive ability of the ISR triage teams' risk categorizations varied widely among classes, performing more poorly for aggressors who were Māori or women, and for non-Intimate Partner Violence (IPV) cases (e.g., episodes between siblings or parents and children; Jolliffe Simpson et al., 2023). We do not know why the risk categories performed poorly for these groups; for example, it is possible they received effective interventions that decoupled the risk categorizations from the outcomes we measured. Though we cannot directly address this question here, the exploration of the triage teams' factors in this study may help shed light on some reasons for the risk categories' varying discriminative ability across classes and generate questions for further research.

This Study

We previously examined the predictive validity of the ISR triage teams' recorded factors and risk categories (Jolliffe Simpson et al., 2023), but have not explored the factors' utility for case management. In this study we will use latent class analysis of the ISR triage teams' recorded factors for 842 FV cases to create classes of cases with similar factors to identify and describe the types of cases the ISR manages. These classes may have implications for families' treatment and case management, such as highlighting barriers to responsivity and agencies' abilities to maximize engagement and behavior change (Bonta & Andrews, 2016).

Research Questions

1. To what extent can we identify classes of families from the factors ISR triage teams recorded during their risk assessments?
2. How do the triage teams' recorded factors differ between the identified classes?
3. What are the demographic and case characteristics of families assigned to the identified classes?

Method

Data Source

We used archival data from the ISR's Family Safety System database for FV episodes reported to police in the Waikato and parts of Canterbury from 1–14 November 2018 (N = 842; Jolliffe Simpson et al., 2023). The data were collected for a larger study on FV in New Zealand (Jolliffe Simpson et al., 2021). Variables in the dataset included demographic information about the aggressor and victim, and descriptors about the reported episode (e.g., whether children were present at the scene of the episode) and case management (e.g., the type of organization leading the management of each case). The dataset also included New

Zealand Police's Static Assessment for Family Violence Recidivism (SAFVR; Bissielo & Knight, 2016) and Dynamic Risk Assessment (DYRA; Jolliffe Simpson et al., 2021) for the index episode; and the ISR triage teams' risk categorization made after the index episode (Integrated Safety Response, 2018).

ISR Triage Teams' Factors

For a previous study, we coded 71 factors that ISR triage teams recorded when performing their risk assessments (Jolliffe Simpson et al., 2023). ISR triage teams discussed each case and made a risk categorization (low, medium, or high) before recording factors in their case notes to support that risk categorization. The factors were loosely based on the ISR's risk assessment guide that contains 46 items about the victim, aggressor, relationship, children or young people, and practice considerations (Integrated Safety Response, 2018; Jolliffe Simpson et al., 2023); however, triage teams did not score items on the risk assessment guide and could include factors external to the guide. For more detail about the coding process, see the article by Jolliffe Simpson and colleagues (2023).

FV Outcomes

For the same previous study, we also recorded two binary indicators for FV outcomes (Jolliffe Simpson et al., 2023). Recurrence was defined as a further FV-related call for police service, and physical recurrence was defined as a further FV-related call for police service for an episode involving physical harm. These outcomes were coded as present if recurrence or physical recurrence was reported to police in the 24 weeks following the index episode, and the index aggressor was again recorded as the aggressor or a mutual participant in the episode[12] (i.e., not the victim).

[12] Police used the role "mutual participant" when they were unable to determine a single aggressor at the scene.

Plan for Analysis

We conducted analyses in R (R Core Team, 2013), and in IBM SPSS Statistics v. 27. To address our first research question about whether we could identify classes of families from the factors ISR triage teams recorded during their risk assessments, we used the 71 factors to perform latent class analysis. Latent class analysis is a useful technique for describing heterogeneity within a population and identifying sub-groups (Lanza & Cooper, 2016). This approach uses binary or categorical variables to identify patterns in the data and group together cases with similar characteristics (Cavanaugh et al., 2012).

We used the poLCA package (Linzer & Lewis, 2011) to group cases with similar factors together in solutions with one to six classes, then used a variety of information sources to select the best model for the data (Nylund-Gibson & Choi, 2018; Weller et al., 2020). These sources included model fit statistics such as the Approximate Weight of Evidence (AWE), Bayesian Information Criterion (BIC), Consistent Akaike Information Criterion (CAIC), and Log Likelihood (LL; Nylund-Gibson & Choi, 2018), where lower values indicate better model fit. We also considered the interpretability of the class sizes (Nylund-Gibson & Choi, 2018) and calculated the Lo–Mendell–Rubin adjusted likelihood ratio test (LMRT; Nylund et al., 2007). A p value < .05 for the LMRT indicates the current solution improves model fit compared with a solution containing one fewer class (Nylund-Gibson & Choi, 2018). There were no missing data, and no variables other than the ISR triage teams' recorded factors were used to form the classes.

After selecting a latent class model, we examined how well it differentiated between cases in the sample. First, we reported entropy, an index for model classification accuracy (Weller et al., 2020). There are no established criteria for interpreting entropy values, but .8 is often used to indicate good case classification (Nylund-Gibson & Choi, 2018) and .6 is used as a minimum value for good class separation (Asparouhov & Muthén, 2014). We also

examined the average posterior class membership probabilities for each of the classes. An average posterior probability of 1 indicates all cases in a class were perfectly assigned to that class, therefore values closer to 1 indicate better class assignment, and values over .8 are preferred (Nagin, 2005; Weller et al., 2020). Then we examined the conditional probabilities for the triage teams' recorded factors by class (Schreiber, 2017). Conditional item probabilities below .3 and above .7 can indicate a greater degree of similarity between cases in each of the classes (i.e., class homogeneity; Nylund-Gibson & Choi, 2018).

To answer our second research question, we used χ^2 crosstabulations to describe how the frequency of the factors and the demographic and case characteristics of families differed between the identified classes. Finally, to answer our third research question, we used χ^2 crosstabulations to describe the characteristics of families assigned to each class. In this step, we used information separate to the factors that we used to create the classes to provide a sense of external validity.

Results

Table 1 describes latent class models with one to six classes. We chose to proceed with the three-class model. The two- and three-class models had the smallest AWE, BIC, and CAIC statistics (Nylund-Gibson & Choi, 2018), and the LMRT indicated adding a third class significantly improved model fit compared with the two-class model (Nylund et al., 2007). In addition, all classes in the three-class model were sizeable enough to interpret, compared with the four and five-class models that had classes representing <1% of the sample.

Table 1

Model Fit and Diagnostic Criteria

Model fit criteria				
Classes (n)	*LL*	*BIC*	*CAIC*	*AWE*
1	-16137.86	32753.95	32824.95	32860.45
2	-15667.83	32298.87	32441.87	32513.37
3	-15511.12	32470.44	32685.44	32792.94
4	-15479.89	32892.96	33179.96	33323.46
5	-15374.5	33167.15	33526.15	33705.65
6	-15193.38	33289.88	33720.88	33936.38
Diagnostic criteria				
Classes (n)	*Smallest class (n)*	*Smallest class (%)*	*Entropy*	*LMRT p*
1	842	100.0%	-	-
2	340	40.4%	.74	< .001
3	149	17.7%	.74	< .001
4	0	0.0%	.75	.853
5	1	0.01%	.76	< .001
6	69	8.2%	.82	< .001

Note. N = 842. AWE = Approximate Weight of Evidence; BIC = Bayesian Information Criterion; CAIC = Consistent Akaike Information Criterion; LL = Log Likelihood; LMRT = Lo–Mendell–Rubin adjusted likelihood ratio test.

Class 1 was the smallest identified class, comprised of 17.7% of the sample (n = 149). Class 2 included approximately a quarter (24.7%, n = 208) of the sample, and class 3 was the largest class, accounting for more than half of families in the sample (57.6%, n = 485). On average, triage teams recorded 9.38 factors for families in class 1, 9.52 factors for class 2, and 4.34 factors for class 3.

Table 1 shows the entropy value for the three-class model was .74, below the value of .8 used to indicate good classification (Nylund-Gibson & Choi, 2018), but above the minimum value of .6 for good class separation (Asparouhov & Muthén, B., 2014). The average posterior probabilities for the classes in the three-class model were .88, .88 and .92, respectively, indicating good class separation (Nagin, 2005). Figure 1 shows conditional item probabilities for the factors in each class, grouped in sections based on the sections in the ISR risk assessment guide (see supplemental materials; Integrated Safety Response, 2018). Most conditional item probabilities were below .3, which could indicate cases in each of the classes were homogeneous. No conditional item probabilities exceeded .7, but this is unsurprising because most of the factors were low in frequency.

Our second research question concerned how the triage teams' recorded factors differed between the identified classes. To answer this question, we used χ^2 crosstabulations to identify relationships between the factors and the classes. We again present these results in themed sections based on the ISR risk assessment guide (Integrated Safety Response, 2018).

Figure 1

Conditional Item Probabilities for Factors by Class

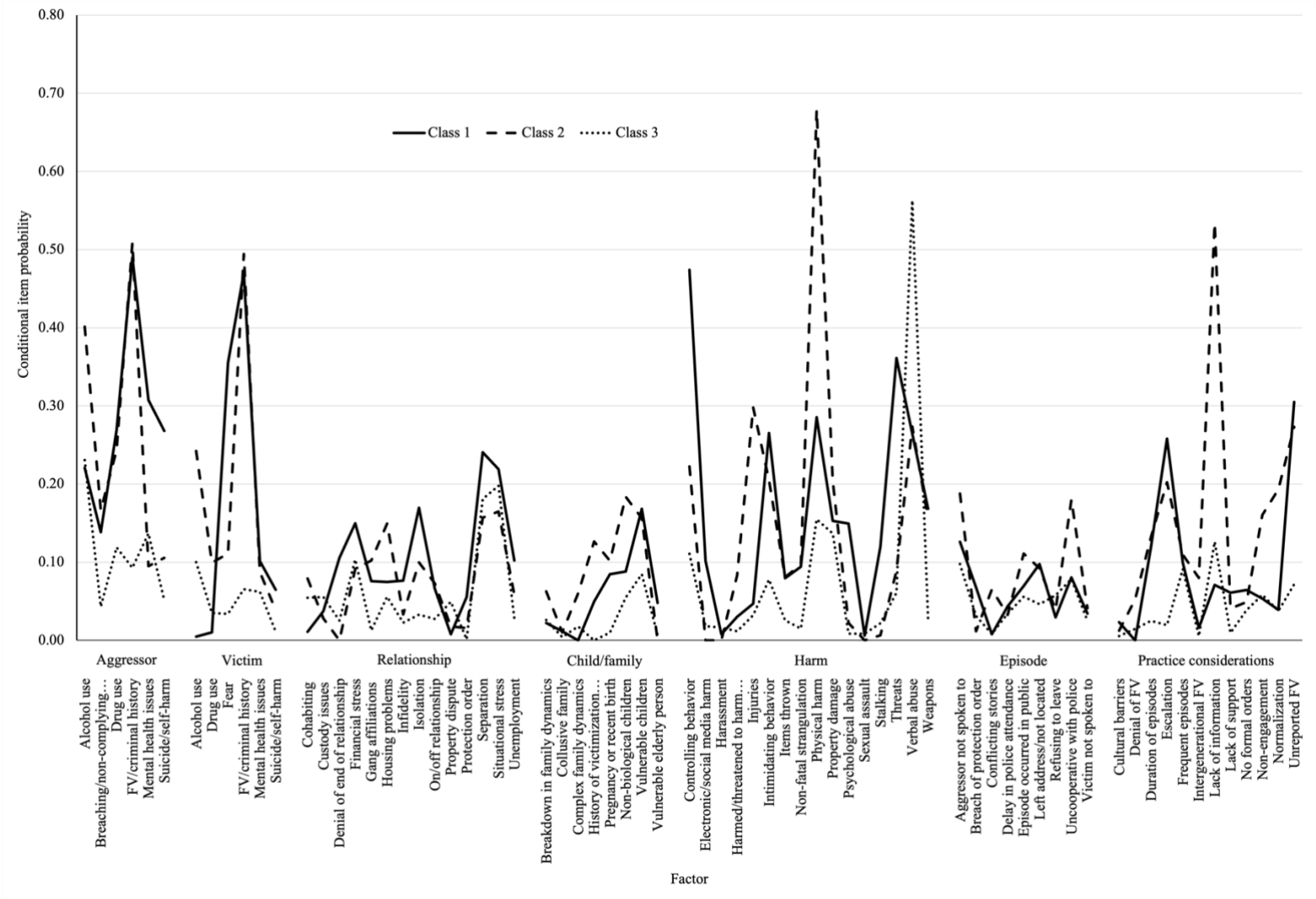


Table 2

Frequency of ISR Triage Teams' Recorded Factors By Class

		Class				
		1	*2*	*3*		
Factor	*f* (%)	(*n* = 149)	(*n* = 208)	(*n* = 485)	χ^2	*p*
Aggressor						
Alcohol use	229 (27.2)	34 (22.8)	84 (40.4)	111 (22.9)	24.26	< .001
Breaching/non-complying with conditions/orders	78 (9.3)	23 (15.4)	34 (16.3)	21 (4.3)	33.21	< .001
Drug use	151 (17.9)	43 (28.9)	48 (23.1)	60 (12.4)	26.02	< .001
FV/criminal history	231 (27.4)	77 (51.7)	109 (52.4)	45 (9.3)	189.44	< .001
Mental health issues	134 (15.9)	47 (31.5)	17 (8.2)	70 (14.4)	37.31	< .001
Suicide/self-harm	90 (10.7)	43 (28.9)	21 (10.1)	26 (5.4)	66.03	< .001
Victim						
Alcohol use	99 (11.8)	1 (0.7)	52 (25.0)	46 (9.5)	55.22	< .001
Drug use	39 (4.6)	1 (0.7)	21 (10.1)	17 (3.5)	20.75	< .001
Fear	97 (11.5)	58 (38.9)	23 (11.1)	16 (3.3)	142.00	< .001
FV/criminal history	212 (25.2)	76 (51.0)	105 (50.5)	31 (6.4)	214.31	< .001
Mental health issues	64 (7.6)	15 (10.1)	19 (9.1)	30 (6.2)	3.37	.185
Suicide/self-harm	24 (2.9)	11 (7.4)	8 (3.8)	5 (1.0)	[a]	[a]
Relationship						
Cohabiting	44 (5.2)	1 (0.7)	17 (8.2)	26 (5.4)	9.91	.007
Custody issues	38 (4.5)	5 (3.4)	6 (2.9)	27 (5.6)	2.99	.224
Denial of end of relationship	29 (3.4)	16 (10.7)	0 (0.0)	13 (2.7)	32.11	< .001
Financial stress	92 (10.9)	23 (15.4)	19 (9.1)	50 (10.3)	3.99	.136
Gang affiliations	40 (4.8)	11 (7.4)	23 (11.1)	6 (1.2)	33.80	< .001
Housing problems	70 (8.3)	10 (6.7)	34 (16.3)	26 (5.4)	23.66	< .001
Infidelity	30 (3.6)	13 (8.7)	7 (3.4)	10 (2.1)	14.76	.001
Isolation	64 (7.6)	28 (18.8)	21 (10.1)	15 (3.1)	42.45	< .001
On/off relationship	39 (4.6)	12 (8.1)	13 (6.3)	14 (2.9)	8.53	.014
Property dispute	28 (3.3)	1 (0.7)	3 (1.4)	24 (4.9)	[a]	[a]
Protection order	13 (1.5)	9 (6.0)	3 (1.4)	1 (0.2)	[a]	[a]
Separation	157 (18.6)	34 (22.8)	31 (14.9)	92 (19.0)	3.66	.160
Situational stress	163 (19.4)	33 (22.1)	35 (16.8)	95 (19.6)	1.61	.446
Unemployment	42 (5.0)	17 (11.4)	11 (5.3)	14 (2.9)	17.52	< .001

Table 2 continued

		Class				
		1	*2*	*3*		
Factor	*f* (%)	(*n* = 149)	(*n* = 208)	(*n* = 485)	χ^2	*p*
Child/family						
Breakdown in family dynamics	29 (3.4)	3 (2.0)	14 (6.7)	12 (2.5)	9.05	.011
Collusive family	6 (0.7)	2 (1.3)	2 (1.0)	2 (0.4)	[a]	[a]
Complex family dynamics	21 (2.5)	0 (0.0)	13 (6.3)	8 (1.6)	[a]	[a]
History of victimization/ cumulative harm	35 (4.2)	8 (5.4)	27 (13.0)	0 (0.0)	62.24	< .001
Pregnancy or recent birth	40 (4.8)	14 (9.4)	21 (10.1)	5 (1.0)	35.07	< .001
Non-biological children	79 (9.4)	12 (8.1)	41 (19.7)	26 (5.4)	35.64	< .001
Vulnerable children	100 (11.9)	27 (18.1)	32 (15.4)	41 (8.5)	13.43	.001
Vulnerable elderly person	11 (1.3)	8 (5.4)	0 (0.0)	3 (0.6)	[a]	[a]
Harm						
Controlling behavior	176 (20.9)	75 (50.3)	46 (22.1)	55 (11.3)	105.08	< .001
Electronic/social media harm	25 (3.0)	16 (10.7)	0 (0.0)	9 (1.9)	[a]	[a]
Harassment	9 (1.1)	1 (0.7)	0 (0.0)	8 (1.6)	[a]	[a]
Harmed/threatened to harm, intimidated children, family members, or pets	28 (3.3)	4 (2.7)	19 (9.1)	5 (1.0)	[a]	[a]
Injuries	86 (10.2)	7 (4.7)	64 (30.8)	15 (3.1)	127.60	< .001
Intimidating behavior	123 (14.6)	43 (28.9)	40 (19.2)	40 (8.2)	43.55	< .001
Items thrown	42 (5.0)	13 (8.7)	17 (8.2)	12 (2.5)	15.31	< .001
Non-fatal strangulation	43 (5.1)	15 (10.1)	21 (10.1)	7 (1.4)	31.68	< .001
Physical harm	263 (31.2)	44 (29.5)	148 (71.2)	71 (14.6)	216.71	< .001
Property damage	132 (15.7)	22 (14.8)	43 (20.7)	67 (13.8)	5.29	.071
Psychological abuse	33 (3.9)	23 (15.4)	5 (2.4)	5 (1.0)	64.50	< .001
Sexual assault	5 (0.6)	1 (0.7)	0 (0.0)	4 (0.8)	[a]	[a]
Stalking	31 (3.7)	19 (12.8)	1 (0.5)	11 (2.3)	43.31	< .001
Threats	106 (12.6)	59 (39.6)	17 (8.2)	30 (6.2)	120.53	< .001
Verbal abuse	364 (43.2)	38 (25.5)	56 (26.9)	270 (55.7)	72.20	< .001
Weapons	73 (8.7)	26 (17.4)	34 (16.3)	13 (2.7)	51.96	< .001

Table 2 continued

		Class				
		1	*2*	*3*		
Factor	*f* (%)	(*n* = 149)	(*n* = 208)	(*n* = 485)	χ^2	*p*
Episode						
Aggressor not spoken to	106 (12.6)	20 (13.4)	39 (18.8)	47 (9.7)	10.97	.004
Breach of protection order	28 (3.3)	11 (7.4)	2 (1.0)	15 (3.1)	[a]	[a]
Conflicting stories	19 (2.3)	1 (0.7)	13 (6.3)	5 (1.0)	[a]	[a]
Delay in police attendance	30 (3.6)	7 (4.7)	6 (2.9)	17 (3.5)	0.84	.656
Episode occurred in public	61 (7.2)	10 (6.7)	24 (11.5)	27 (5.6)	7.80	.020
Left address/not located	57 (6.8)	16 (10.7)	18 (8.7)	23 (4.7)	8.05	.018
Refusing to leave	40 (4.8)	3 (2.0)	8 (3.8)	29 (6.0)	4.46	.107
Uncooperative with police	88 (10.5)	12 (8.1)	40 (19.2)	36 (7.4)	22.80	< .001
Victim not spoken to	26 (3.1)	4 (2.7)	9 (4.3)	13 (2.7)	[a]	[a]
Practice considerations						
Cultural barriers	9 (1.1)	4 (2.7)	2 (1.0)	3 (0.6)	[a]	[a]
Denial of FV	18 (2.1)	0 (0.0)	11 (5.3)	7 (1.4)	[a]	[a]
Duration of episodes	59 (7.0)	18 (12.1)	28 (13.5)	13 (2.7)	33.12	< .001
Escalation	94 (11.2)	44 (29.5)	40 (19.2)	10 (2.1)	104.84	< .001
Frequent episodes	82 (9.7)	15 (10.1)	22 (10.6)	45 (9.3)	0.30	.860
Intergenerational FV	22 (2.6)	2 (1.3)	18 (8.7)	2 (0.4)	[a]	[a]
Lack of information	184 (21.9)	9 (6.0)	116 (55.8)	59 (12.2)	188.58	< .001
Lack of support	23 (2.7)	10 (6.7)	9 (4.3)	4 (0.8)	[a]	[a]
No formal orders	39 (4.6)	11 (7.4)	9 (4.3)	19 (3.9)	3.16	.206
Non-engagement	70 (8.3)	7 (4.7)	34 (16.3)	29 (6.0)	23.63	< .001
Normalization	65 (7.7)	6 (4.0)	41 (19.7)	18 (3.7)	55.78	< .001
Unreported FV	141 (16.7)	50 (33.6)	57 (27.4)	34 (7.0)	80.12	< .001

Note. *f* = frequency in total sample. FV = Family Violence.
[a] At least 1 cell had an expected count less than 5, violating χ^2 assumptions.

Aggressor-related Factors

Table 2 shows the factors' frequency overall and among cases in each class. Paired with Figure 1, Table 2 shows that aggressor-related factors were more often recorded for cases in classes 1 and 2 than in class 3. Around half of cases in classes 1 and 2 had the factor *FV/criminal history*, and the factors *breaching/non-complying with conditions/orders* and *drug use* were also more often recorded for cases in these classes compared with class 3. Furthermore, *mental health issues* and *suicide/self-harm* were more often recorded for cases in class 1, and *alcohol use* was more often recorded for cases in class 2, than for the other classes. All aggressor-related factors were recorded less often for cases in class 3, except for *mental health issues*, which was least often recorded for cases in class 2.

Victim-related Factors

Around half of cases in classes 1 and 2 had the victim-related factor *FV/criminal history,* compared with fewer than a tenth of cases in class 3. In addition, the factors *fear* and *suicide/self-harm* [a] were more often recorded for cases in class 1, and *alcohol* and *drug use* were more often recorded for cases in class 2, compared with cases in the other classes. None of the victim-related factors were recorded more often for cases in class 3, except for *drug use*, which was least often recorded for cases in class 1.

Relationship-related Factors

Table 2 and Figure 1 show many relationship-related factors were more often recorded for cases in class 1, including *isolation, denial of end of relationship, unemployment, infidelity,* and *financial stress.* The factors *housing problems* and *gang affiliations* were more often recorded for cases in class 2, whereas *denial of end of relationship* was not recorded for any cases in class 2. The factor *property dispute* [a] was more often recorded for cases in class 3 than in classes 1 or 2.

Child and Family-related Factors

The child and family-related factors *vulnerable children* and *pregnancy or recent birth* were more often recorded for cases in classes 1 and 2 than class 3. Table 2 and Figure 1 show many other child and family-related factors were more often recorded for cases in class 2 than the other classes, including *non-biological children, history of victimization/ cumulative harm,* and *breakdown in family dynamics*. None of the child and family-related factors were recorded more often for cases in class 3, except for *complex family dynamics*, which was not recorded for any cases in class 1.

Harm-related Factors

The harm-related factors *intimidating behavior, weapons, non-fatal strangulation,* and *items thrown* were more often recorded for cases in classes 1 and 2 compared with class 3. Furthermore, several other harm-related factors including *controlling behavior, psychological abuse, stalking,* and *threats* were more often recorded for cases in class 1, and factors including *physical harm* and *injuries* were more often recorded for cases in class 2, compared with cases in the other classes. Most harm-related factors were less commonly recorded for cases in class 3, except for *verbal abuse*, which was recorded for around half of cases in the class, compared with around a quarter of cases in classes 1 and 2.

Episode-related Factors

The episode-related factors *uncooperative with police, aggressor not spoken to*, and *episode occurred in public* were more often recorded for cases in class 2 compared with cases in the other classes. The factors *breach of protection order* and *left address/not located* were more often recorded for cases in class 1. Most episode-related factors were less commonly recorded for cases in class 3 compared with the other classes, except *refusing to leave*.

Table 3

Demographic and Case Characteristics Overall (N= 842) and by Class

		Class				
		1	2	3		
	f (%)	(*n* = 149)	(*n* = 208)	(*n* = 485)	χ^2	*p*
Aggressor gender [a]					15.22	< .001
Gender diverse	2 (0.2)	1 (0.7)	0 (0.0)	1 (0.2)		
Female	203 (24.1)	22 (14.8)	41 (19.7)	140 (28.9)		
Male	637 (75.7)	126 (84.6)	167 (80.3)	344 (70.9)		
Aggressor ethnicity						
European	360 (42.8)	65 (43.6)	63 (30.3)	232 (47.8)	31.41	< .001
Māori	373 (44.3)	65 (43.6)	126 (60.6)	182 (37.5)		
Other	109 (12.9)	19 (12.8)	19 (9.1)	71 (14.6)		
Relationship to victim					46.5	< .001
Current partner	418 (49.6)	66 (44.3)	130 (62.5)	224 (46.2)		
Former partner	177 (21.0)	44 (29.5)	21 (10.1)	112 (23.1)		
Parent	42 (5.0)	6 (4.0)	14 (6.7)	22 (4.5)		
Child	108 (12.8)	24 (16.1)	12 (5.8)	70 (14.4)		
Sibling	51 (6.1)	6 (4.0)	18 (8.7)	27 (5.6)		
Other	46 (5.5)	3 (2.0)	13 (6.3)	30 (6.2)		
Relationship type					1.85	.397
IPV	595 (70.7)	110 (73.8)	151 (72.6)	334 (68.9)		
Non-IPV	247 (29.3)	39 (26.2)	57 (27.4)	151 (31.1)		
Victim gender					12.23	.002
Female	648 (77.0)	127 (85.2)	168 (80.8)	353 (72.8)		
Male	194 (23.0)	22 (14.8)	40 (19.2)	132 (27.2)		
Victim ethnicity					23.13	< .001
European	388 (46.1)	72 (48.3)	72 (34.6)	244 (50.3)		
Māori	351 (41.7)	59 (39.6)	116 (55.8)	176 (36.3)		
Other	103 (12.2)	18 (12.1)	20 (9.6)	65 (13.4)		
Children involved	419 (49.8)	87 (58.4)	108 (51.9)	224 (46.2)	7.31	.026
Reported by					29.96	< .001
Family member	198 (23.5)	30 (20.1)	57 (27.4)	111 (22.9)		
Neighbor	118 (14.0)	13 (8.7)	33 (15.9)	72 (14.8)		
Other	155 (18.4)	15 (10.1)	46 (22.1)	97 (20.0)		
Victim	368 (43.7)	91 (61.1)	72 (34.6)	205 (42.3)		

Table 3 continued

		Class				
		1	2	3		
	f (%)	(*n* = 149)	(*n* = 208)	(*n* = 485)	χ^2	*p*
ISR area					60.19	< .001
Waikato	480 (57.0)	96 (64.4)	160 (76.9)	224 (46.2)		
Canterbury	362 (43.0)	53 (35.6)	48 (23.1)	261 (53.8)		
Lead organization					60.84	< .001
Ara Poutama	66 (7.8)	14 (9.4)	18 (8.7)	34 (7.0)		
Lifeline	165 (19.6)	15 (10.1)	21 (10.1)	129 (26.6)		
District Health Boards	34 (4.0)	12 (8.1)	10 (4.8)	12 (2.5)		
NGO	292 (34.7)	52 (34.9)	72 (34.6)	168 (34.6)		
Oranga Tamariki	46 (5.5)	10 (6.7)	14 (6.7)	14 (2.9)		
Police	45 (5.3)	17 (11.4)	14 (6.7)	14 (2.9)		
Women's Refuge	194 (23.0)	29 (19.5)	59 (28.4)	106 (21.9)		
Kaupapa Māori organization	207 (24.6)	33 (22.1)	82 (39.4)	92 (19.0)	22.43	< .001
SAFVR risk category					14.63	.023
No score	122 (14.5)	19 (12.8)	19 (9.1)	84 (17.3)		
Low	168 (20.0)	26 (17.4)	35 (16.8)	107 (22.1)		
Moderate	109 (12.9)	20 (13.4)	28 (13.5)	61 (12.6)		
High	443 (52.6)	84 (56.4)	126 (60.6)	233 (48.0)		
DYRA risk category					109.34	< .001
Not answered	121 (14.4)	9 (6.0)	51 (24.5)	61 (12.6)		
Low	200 (23.8)	12 (8.1)	40 (19.2)	148 (30.5)		
Moderate	207 (24.6)	25 (16.8)	43 (20.7)	139 (28.7)		
High	314 (37.3)	103 (69.1)	74 (35.6)	137 (28.2)		
ISR risk category					164.67	< .001
Low	316 (37.5)	23 (15.4)	26 (12.5)	267 (55.1)		
Medium	470 (55.8)	104 (69.8)	159 (76.4)	207 (42.7)		
High	56 (6.7)	22 (14.8)	23 (11.1)	11 (2.3)		
Outcomes						
Recurrence	389 (46.2)	80 (53.7)	100 (48.1)	209 (43.1)	5.54	.063
Physical recurrence	174 (20.7)	28 (18.8)	47 (22.6)	99 (20.4)	0.81	.667

Note. f = frequency in total sample. IPV = Intimate Partner Violence, ISR = Integrated Safety Response, NGO = Non-Governmental Organization, SAFVR = Static Assessment for Family Violence Recidivism, DYRA = Dynamic Risk Assessment.

[a] Gender diverse aggressors were excluded from chi-square test.

Practice-related Factors

Finally, Table 2 and Figure 1 show many practice-related factors, including *lack of information, normalization,* and *non-engagement,* were more often recorded for cases in class 2 than for cases in the other classes. In addition, the factors *duration of episodes, escalation,* and *unreported FV* were more often recorded for cases in classes 1 and 2 compared with class 3. Again, most practice-related factors were less commonly recorded for cases in class 3 than the other classes.

Demographic and Case Characteristics

To answer our third research question, we used χ^2 crosstabulations to describe the demographic and case characteristics of families in the classes. Table 3 shows approximately three quarters of aggressors were men overall, but this proportion differed between the classes; being largest in class 1 and smallest in class 3. Approximately three quarters of victims were women, and this proportion also differed between the classes following the opposite pattern to aggressor gender. The table also shows most aggressors and victims were Māori or European, but this differed between the classes. The proportion of aggressors and victims who were Māori was largest in class 2 and smallest in class 3, and proportion of aggressors and victims who were European was largest in class 3 and smallest in class 2. We used a one-way ANOVA for the association between the classes and age, and found victims tended to be younger in class 2 compared with classes 1 and 3 ($F(2) = 11.95$, $p < .001$), but there was no relationship with aggressor age.

The relationship of aggressors to victims also differed across the classes. Overall, half of cases involved current intimate partners, and a further fifth involved former intimate partners, meaning just over two-thirds of cases were categorized as Intimate Partner Violence (IPV). The proportion of cases categorized as IPV did not differ significantly between the classes, but the proportion of cases involving current partners was higher in class 2 compared

with classes 1 and 3, and the proportion of cases involving former partners was lower in class 2 compared with classes 1 and 3. Furthermore, class 2 had a smaller proportion of cases where the aggressor was the victim's child, and a larger proportion of cases where the aggressor was a parent, sibling or other family member of the victim, compared with the other classes. Class 1 had the largest proportion of episodes where children were involved (i.e., present at the scene of the episode preceding the ISR triage teams' assessment), and the class 3 had the smallest proportion of these episodes.

Further, Table 3 shows about two fifths of the episodes preceding the ISR triage teams' assessments to police were reported by victims, with around a fifth being reported by family members or others[13], and the remainder being reported by neighbors. Again, proportions differed across the classes, with the proportion of cases with episodes reported by victims being largest in class 1 and smallest in class 2. The proportion of cases with episodes reported by family members was largest in class 2, and the proportion of cases with episodes reported by neighbors was largest in classes 2 and 3. Finally, the proportion of cases with episodes reported by others (which could include the aggressor themselves) was largest in classes 2 and 3, and smallest in class 1.

Moreover, there was a relationship between the classes and location; just under three fifths of cases in the entire sample were in the Waikato region, but more than three-quarters of cases in class 2 were in the Waikato region, and the proportion of cases in Canterbury was highest in class 3. The *lead organization* section of Table 3 refers to which organization ISR triage teams nominated as the most appropriate to manage each case. This organization is usually nominated because it has existing oversight or connections with the aggressor, victim, or whānau; or may be responsible for interventions allocated in response to the most recent

[13] People who were not the victim, a neighbour, or a family member (e.g., bystanders or aggressors).

episode. There was an association between class membership and lead organization; the proportion of cases Police or the District Health Boards managed was highest in class 1, the proportion of cases managed by a Women's Refuge organization was highest in class 2, the proportion of cases Ara Poutama (Department of Corrections) or Oranga Tamariki managed was highest in classes 1 and 2, and the proportion of cases Lifeline[14] managed was highest in class 3. The proportion of cases managed by NGOs remained similar across the classes, but because many of the NGOs involved in the ISR were kaupapa Māori, we also examined the proportion of cases with kaupapa Māori lead organizations. We found the proportion of cases managed by a kaupapa Māori organization was much higher than average in class 2, and lowest in class 3.

The next section of table 3 shows the police SAFVR and DYRA risk categories, and the ISR triage teams' risk categories. Overall, around half of cases were categorized as high risk on the SAFVR and this proportion neared three fifths in classes 1 and 2. For the DYRA, almost two fifths of cases in the overall sample were high risk; compared with more than two thirds of cases in class 1. Allocation of cases to the high risk category occurred much less often than for the other ISR risk categories; overall, triage teams categorized 6.7% of cases as high risk, this proportion was greater among classes 1 and 2, and smaller in class 3. Finally, Table 3 shows the rates of recurrence and physical recurrence within 24 weeks after the index episode. Just under half of cases in the sample had a recurrence, and one fifth had a recurrence involving physical harm. Interestingly, there was no significant difference in the proportion of cases with recurrence or physical recurrence between the classes.

[14] Lifeline is a free helpline operating in New Zealand. In the context of ISR, Lifeline calls people in low-risk cases to offer support.

Discussion

In this chapter we used latent class analysis of the factors ISR triage teams recorded during their risk assessments to identify and describe the types of cases the ISR manages. We identified three classes of cases with similar factors and described the demographic and case characteristics of families in each. Engaging with families is at the forefront of ISR triage team members' minds when they perform risk assessments, and in this study, we were interested in whether the factors triage teams recorded could add value to the assessments by informing case management (Jolliffe Simpson et al., 2023). The broad classes we identified could be useful for understanding the types of cases the ISR manages, and in this section, we label the classes and discuss their implications for families' treatment and case management (Bonta & Andrews, 2016; Lanza & Cooper, 2016).

Class 1: High Needs

We labelled class 1 as "High Needs"; it involved families with aggressors who had a history of detected criminality (e.g., *breaching/non-complying with conditions/orders* and *FV/criminal history*), experiencing *mental health issues* and *drug use*. Indeed, compared with the other classes, cases in the High Needs class were more often led by Police—which could indicate people had current or previous involvement with law enforcement—and had higher risk categories on the SAFVR, DYRA, and ISR risk assessments. The SAFVR measures static risk factors, and the DYRA measures dynamic risk factors; thus, the larger than average proportion of cases with high-risk categories from both instruments indicates the High Needs class involved aggressors who had criminal or FV histories and who may have been experiencing acute stressors at the time of the reported episode (Bissielo & Knight, 2016; Jolliffe Simpson et al., 2021).

We also labelled this class "High Needs" due to the presence of vulnerable victims (e.g., *fear, vulnerable children, pregnancy or recent birth*) experiencing barriers to

responsivity (e.g., *isolation, unreported FV*). Compared with the other classes, cases in the High Needs class were more often managed by District Health Boards, which could indicate family members had previous connections with health or mental health services. Furthermore, the High Needs class had the largest proportion of aggressors who were the former partner of the victim, triage teams recorded a variety of relationship-related factors (e.g., *denial of end of relationship, infidelity, financial stress*), and forms of harm including *controlling behavior*, *psychological abuse*, and *stalking*. Taken together, these results suggest some families in the High Needs class were experiencing conflict or unwanted contact after the end of an intimate relationship.

The overarching aim of this study was to explore the utility of the ISR triage team's recorded factors for case management; we were interested in whether the factors added value to the teams' risk assessments by highlighting barriers to responsivity and informing practitioners' efforts to maximize engagement and behavior change (Jolliffe Simpson et al., 2023). The profile of the High Needs class indicated the presence of mental health problems including suicidality or self-harm and substance use, which has been as identified as a risk factor for IPV and child maltreatment, and should be addressed during intervention (Stith et al., 2004; Stith et al., 2009). Moreover, due to the dynamic of fear and controlling behavior some families experienced, agencies likely needed to account for those needs and implement additional strategies to reach and engage with victims (e.g., by providing independent victim support advocates to improve their prospects of fostering engagement and preventing harm).[15]

[15] The case management recommendations we describe in this discussion reflect our interpretation of necessary actions based on the ISR triage teams' recorded factors, and these steps may have already occurred in practice

Class 2: Complex Needs

We labelled class 2 as "Complex Needs" because the dominant theme that emerged from the triage teams' recorded factors consisted of barriers to responsivity (e.g., *lack of information, normalization, gang affiliations, uncooperative with police, unreported FV, non-engagement,* and *housing problems*). Compared with the other classes, the Complex Needs class had the largest proportion of aggressors who were the current partner of the victim, the smallest proportion of episodes reported by victims, and a larger proportion of some victim- and child and family-related factors indicating vulnerability (e.g., *pregnancy or recent birth, history of victimization/cumulative harm, vulnerable children, victim alcohol* and *drug use*). These factors indicate families in the Complex Needs class were unlikely to contact—and engage with—law enforcement and other organizations providing support for people experiencing FV.

In addition to experiencing barriers to engagement, the Complex Needs class had the largest proportion of cases managed by refuge organizations and involved harm-related factors such as *physical harm* and *injuries,* which may indicate elevated severity of IPV and the need for safe housing. Like the High Needs class, the triage team's recorded factors suggested the Complex Needs class involved with aggressors with criminal or FV histories (e.g., *aggressor FV/criminal history* and *breaching/non-complying with conditions/orders*), and compared with the other classes, the Complex Needs class had the largest proportion of cases involving aggressor *alcohol use* and high risk categories on the SAFVR.

Taken together, the profile of the Complex Needs class involved physical harm, vulnerabilities including substance use, and non-engagement with services. Practitioners managing cases in this class may have needed to refer people to specialist intervention services (e.g., substance use programs) and likely experienced problems fostering engagement and supporting behavior change due to the many and varied barriers to

responsivity (Bonta & Andrews, 2016). Importantly, the Complex Needs class had the largest proportion of families where the aggressor or victim was Māori. When paired with the factors suggesting barriers to responsivity, the larger proportion of Māori among this class indicates agencies providing responses should deliver them in a way that facilitates Māori engagement and wellbeing. The ISR was initially criticized for its lack of responsiveness to Māori (Mossman et al., 2017), but has since increased investment in Māori positions and the number of Māori service providers included. This investment was illustrated here by the fact that Kaupapa Māori organizations were the lead agencies for almost half of cases in this class.

Class 3: Low Needs

We labelled class 3 as "Low Needs" because few of the ISR triage teams' factors were more common among cases this class compared with the other classes, and those that were more common (e.g., *verbal abuse*) suggest families were experiencing relatively low-level harm that may have been precipitated by acute and/or temporary stressors. The characteristics more common among cases in the Low Needs class support this observation; above all, this class involved the largest proportion of cases managed by Lifeline, indicating a low-risk response. Overall, the profile of the Low Needs class suggests that families may not have required intensive intervention following the reported episode, but nevertheless, should be encouraged to continue to seek help and report future FV. Based on this appraisal, triage teams may have triaged out families as low risk, and instead focused their intervention efforts on cases in the High and Complex Needs classes.

Limitations

The factors we used to construct the classes were measured cross sectionally (i.e., they were taken from a single risk assessment). This means we could not account for cases' involvement with the ISR and its partner agencies prior to the index episode, including

whether they had previously received interventions. In addition, we could not account for interventions assigned following the risk assessment in question because that information was not systematically recorded. The finding that there was no difference in rates of recurrence between the classes could suggest the triage teams' factors cannot be used to meaningfully distinguish cases according to their risk of recurrence. However, because we could not account for interventions that were delivered after the assessments took place, this same finding could also indicate some cases received effective interventions that moderated the relationship between the recorded factors and recurrence. Because ISR triage teams assign interventions to each case immediately after identifying purported risk factors, this is a reasonable alternative explanation that requires further investigation. Finally, in this study we analyzed practitioners' recorded factors and did not measure the actual characteristics of cases. It is possible that factors recorded for some cases were present in other cases, but not recorded by triage teams. Therefore, the results should be interpreted as an analysis of the triage teams' description of cases, rather than of the cases themselves.

Implications

The results from this research have several implications. First, our results provide support for the idea that in addition to appraising future risk, during risk assessments practitioners also find information relevant to how they plan their response (Kropp, 2008). ISR triage team members recorded various factors, many of which reflected parties' willingness to engage with service providers, and barriers to engagement present among the family unit. This finding suggests that ISR triage teams' risk assessments are case management planning sessions, which is not undesirable but suggests evaluations of those assessments' performance should factor in their validity for case management alongside their validity for predicting recurrence. Further efforts such as operationalizing measurable outcomes indicating good case management and improving day-to-day recording of

interventions and engagement are required before this type of evaluation can be systematically performed. Moreover, the ISR would benefit from requiring triage teams to use a structured risk assessment instrument designed to guide case management, such as the LS/CMI (Andrews at al., 2004). This change would require triage teams to systematically consider factors related to risk, need, and responsivity and could improve both the predictive validity and utility of their assessments for case management (Andrews at al., 2004).

The results from this study show practitioners understand the risk, need and responsivity principles—which were derived from meta-analysis based on individual people—as applying more widely (Bonta & Andrews, 2016). Triage teams described stressors and vulnerabilities affecting the family as a whole, which makes sense because human behavior is intimately connected to social and physical environments (Wortley, 2011). But, this approach contrasts to the individual focus typical of research applying Bonta & Andrews' RNR model for correctional treatment, where the focus is on individual offenders (Bonta & Andrews, 2016; Hilton & Ennis, 2020; Polaschek, 2012). Research supports the RNR model's relevance for FV; its central eight predictors of criminal offending are established predictors of FV, and many of the people who commit FV also offend more generally (Hilton & Eke, 2016; Timshel et al., 2017; Polaschek, 2016). Moreover, despite its dominance in individual-based correctional treatment, hypothetically the RNR model can be applied to a wide range of interventions including family-based interventions (Bonta & Andrews, 2016). Therefore, it makes sense for the RNR model to inform intervention services for FV cases, but to our knowledge research has not considered what adaptations are necessary for the model and its principles to adequately fulfil this purpose (Polaschek, 2016).

Moreover, in the interest of focusing on what works, the RNR model provides principles for correctional treatments that reduce recidivism rather wellbeing or agreeableness (Bonta & Andrews, 2016). That focus on criminal offending excludes factors precipitating

sub-criminal, but harmful —and often chronic—behavior. Fewer than one third of FV cases reported to police in New Zealand involve physical harm and even fewer result in the detection of offenses or arrests (Jolliffe Simpson et al., 2021). Therefore, in the context of FV, the RNR model's responsivity principle could widen from its focus on aggressor characteristics that require accommodations in treatment (e.g., literacy) to include family-wide factors that act as barriers to responsivity. Moreover, the need principle should widen from its focus on aggressor characteristics that predict subsequent criminal behavior to include chronic external stressors that precipitate to conflict and cause cumulative harm (Polaschek, 2016).

Conclusion

Overall, in this study we showed that we could create classes from the factors ISR triage teams recorded during their risk assessments. We labelled and described the classes as involving High Needs, Complex Needs, and Low Needs; and posited some of the factors' implications for agencies' abilities to engage with families. The results provide support for the idea that risk assessments are useful for case management, and have practical implications for how the ISR manages cases, with the implementation of a structured instrument to guide ISR triage teams' risk assessment and case management being a clear next step. Finally, the results of this study have implications for how principles from the RNR model can be applied to FV case management, and raise questions about whether and how those principles should be extended to apply to social units.

Chapter Five: Manuscript Three

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2023). Modeling behavioral patterns of family violence aggressors in Aotearoa New Zealand [Manuscript submitted for publication].

Abstract

Purpose: The presumption that Family Violence (FV) will continue and escalate without intervention is embedded in aspects of FV practice, including risk assessment and case management. However, there is limited evidence that further episodes are inevitable after an episode is reported, or that subsequent episodes will increase in severity. Therefore, we need to better understand temporal patterns in aggressor behavior to inform how risk is conceptualized in practice. **Method:** This study sought to understand the behavioral patterns of FV aggressors in Aotearoa New Zealand. For a sample of 2115 FV aggressors, we entered information New Zealand Police routinely collected between 2018-2020 into a hidden Markov model to estimate the latent (i.e., unmeasurable) states behind the information reported to police, and model aggressors' movement between those states over a two-year period. **Results:** We identified three latent states. The first contained low or no harm, the second contained small probabilities of harm, and the third state involved a high probability of verbal abuse and a moderate probability of physical violence. We identified four pathways through the latent states over the two-year follow-up period, which we called *No harm*, *De-escalation*, *High harm*, and *Low harm*. **Conclusions:** Information that police routinely collect about FV cases can be used to estimate aggressors' latent behavioral states and model pathways through those states over time. The results have implications that could contribute to improved FV risk assessment and practice.

Keywords family violence, intimate partner violence, longitudinal, hidden Markov, latent states, typologies, risk assessment, police

Modeling behavioral patterns of family violence aggressors in Aotearoa New Zealand

Family Violence (FV) is a heterogeneous problem. It involves harm to family members in the broadest sense, encompassing Intimate Partner Violence (IPV), child maltreatment, elder, child to parent, and sibling abuse; and acts ranging from verbal abuse to physical violence (Family Violence Act 2018, § 9). People have long presumed that FV aggressors will continue and escalate their harmful acts without intervention (Feld & Straus, 1989; Pagelow, 1981). One of the many reasons for this assumption is that FV can result in homicide; between 2009 and 2018 there were 125 IPV-related deaths (New Zealand Family Violence Death Review Committee, 2021a), and 70 child maltreatment-related deaths in New Zealand (New Zealand Family Violence Death Review Committee, 2021b). The assumption that harm will continue and escalate without intervention is embedded in aspects of FV practice—such as risk assessment and case management—that seek to prevent serious harm and FV-related deaths (Henning et al., 2021; Richards, 2009). However, there is limited evidence that further episodes are inevitable after people report a FV episode to police, or that those episodes will escalate in severity (Barnham et al., 2017; Bland, 2015; Bland & Ariel, 2015). Therefore, we need to better understand temporal patterns in aggressor behavior to inform how FV risk is conceptualized in practice.

The term FV amalgamates various harmful acts committed in the context of familial relationships, but research on FV overwhelmingly focuses on IPV (McEwan et al., 2018) and child maltreatment (Moody et al., 2018; Stoltenborgh et al., 2015). There is a smaller body of evidence about elder abuse (Jackson, 2016; Storey, 2020), child to parent (Arias-Rivera et al., 2020; Condry & Miles, 2014; Ibabe, 2020), and sibling violence (Tucker et al., 2013); most studies focus on one type of FV in isolation. Yet, when one type of FV occurs, it is often accompanied by others (Chan et al., 2021; Dixon et al., 2007) and when police respond to FV calls for service, they are often faced with multiple, complex relationships between family

members (McEwan et al., 2018; Saxton et al., 2022). Accordingly, New Zealand legislation defines FV as physical, sexual, or psychological violence inflicted against a person 'whom [the aggressor] is, or has been, in a family relationship with' (Family Violence Act 2018, § 9), and in the New Zealand context, family relationships are interpreted with reference to the concept of whānau[16].

FV also encompasses heterogeneous types of harm. Acts constituting FV range in severity from non-physical harm (e.g., verbal abuse, property damage, threats, and coercive control; Robinson et al., 2018; Wiener, 2017) to physical and sexual violence, and homicide (Kim & Merlo, 2021). People have long presumed that FV aggressors continue and escalate harmful acts without intervention from external parties (e.g., law enforcement; Bland & Ariel, 2015; Feld & Straus, 1989). This idea was first expressed by researchers examining IPV who observed that men's physical violence towards their wives would occur more than once and increase in severity until the relationship ended due to either death or separation (Pagelow, 1981). Importantly, escalation can refer to an increase in either the severity or frequency of harmful acts (Barnham et al., 2017; Bland & Ariel, 2015), but is most often conceptualized as increases in the severity of harm (e.g., a progression from verbal abuse to physical violence).

Retrospective analysis of the histories of people involved in FV-related homicides adds some support for the escalation hypothesis (New Zealand Family Violence Death Review Committee, 2021a, 2021b). In some cases, the perpetrators or victims of these homicides had previous contact with FV-response agencies, leading to questions about whether those agencies could have detected the likelihood of escalation and intervened before it was too late (New Zealand Family Violence Death Review Committee 2021a). As a result,

[16] Whānau is the Te Reo Māori (Māori language) word for extended family or a family-like group.

there has been a widespread push for law enforcement and other agencies to detect risk of continued harm and escalation and mitigate that risk to prevent further FV-related homicides (Henning et al., 2021). Moreover, the idea that agencies can detect and intervene to prevent recurrence and escalation of FV that is otherwise inevitable is part of public discourse; evidenced by news media. For example, New Zealand's Integrated Safety Response (ISR) for FV is credited by media with saving lives through its risk assessment and case management practices (e.g., Ensor & Cooke, 2019).

On the other hand, longitudinal research about patterns in the behavior of FV aggressors has highlighted that the severity and frequency of harmful acts over time differs among cases (Barnham et al., 2017; Bland & Ariel, 2015). Cases range from experiencing one-off events to chronic recurring episodes (Barnham et al., 2017). Moreover, among cases that do experience recurrence, behavioral patterns are non-linear. For example, Piquero and colleagues (2006) found FV aggressors were a heterogeneous mix—escalating or de-escalating their harmful acts, or both, within short follow up periods—and Jones and colleagues (2010) found that most aggressors stably committed low-level harm, while others alternated between different types of harm, and only 2% persistently used physical violence. Taken together, longitudinal research suggests the most common FV patterns involve low-level harm or desistance (Bland & Ariel, 2015; Heckert & Gondolf, 2005; Jones et al., 2010; Piquero et al., 2006; Swartout et al., 2012). Therefore, the preponderance of evidence is that further episodes are not *inevitable* and do not necessarily escalate if they do repeat (Barnham et al., 2017; Bland & Ariel, 2015).

With that said, longitudinal research does indicate that a small proportion of cases experience high levels of harm or escalation (Bland & Ariel, 2015; Robinson & Clancy, 2020). Indeed, an estimated 62% of FV-related harm is experienced by 1% of adults in New Zealand (Mossman et al., 2017). But, searching for these cases with risk assessments

designed to triage out low-risk cases is like searching for a needle in a haystack. This fact is recognized by the advent of targeted instruments that seek to identify the highest risk aggressors (e.g., Robinson & Clancy, 2020). To support these efforts, we need to better understand temporal patterns in aggressor behavior; such research may inform how FV risk is conceptualized in practice.

This Study

In this study, we modeled the behavioral patterns of 2115 FV aggressors from information police in Aotearoa New Zealand collected when responding to FV-related calls for service between November 2018 and October 2020. Previous research about temporal patterns in the behavior of FV aggressors has several limitations that we sought to ameliorate in this study. For example, most research about FV focuses on IPV (e.g., Jones et al., 2010) and when other types of FV are examined they are most often examined in isolation. Therefore, there is a need for research that describes temporal patterns in aggressor behavior that encompasses all types of FV (Dixon & Browne, 2003; Heckert & Gondolf, 2005; Jones et al., 2010; McEwan et al., 2018).

In addition, research about FV tends to focus on single outcomes that capture physical violence or criminal offending (Heckert & Gondolf, 2005). Yet, only around a third of the 165,039 FV investigations New Zealand conducted in 2020 involved behaviors that could be classified as criminal offenses (e.g., physical harm or property damage) and even fewer resulted in arrest (New Zealand Police, 2022). More than half of episodes reported to police in other jurisdictions also involve other forms of antisocial behavior that are not criminalized like verbal abuse or psychological harm, but which nevertheless cause long term cumulative consequences (Ansara & Hindin, 2011; Dichter et al., 2018; Wiener, 2017). These behaviors are not adequately accounted for by single outcomes or measures of severity such as the Crime Harm Index, which weights the severity of acts based on sentencing days for criminal

offenses (Bland & Ariel, 2015; Heckert & Gondolf, 2005). Based on the work of Jones and colleagues (2010), in this study we instead estimated latent (i.e., unmeasurable) states from the information reported to police about aggressors' harmful behaviors. For each state, we described the probability of different types of harm, which provided a more complete picture of aggressor behavior compared with a single outcome or the Crime Harm Index.

In addition, research that intends to inform FV practice should be based on information that agencies already routinely collect, to ensure the findings are relevant to those agencies and can be acted upon (Boxall et al., 2015). Hence, we used information police in New Zealand routinely collected when responding to FV-related calls for service for the sample of 2115 aggressors. Of course, relying on reported information has several limitations, including that many FV episodes are not reported to police (New Zealand Crime and Victims Survey, 2018). This consideration is an additional reason why we chose to estimate latent states: a procedure that explicitly recognizes the unmeasured nature of the construct in question (Jones et al., 2010; Zhang et al., 2010; Zucchini et al., 2016). Finally, research about temporal patterns in aggressor behavior should use non-linear modeling to reflect the non-linear nature of human behavior (Bland & Ariel, 2015; Jones et al., 2010; Piquero et al., 2006). Accordingly, we modeled aggressors' movements between those latent states at three-month intervals over two years using a non-linear Markov model. Then we identified common patterns in aggressor behavior and examined the demographic and risk characteristics of groups of aggressors with similar pathways, to explore whether the findings could inform how FV risk is conceptualized in practice.

Research Questions

1. Can we estimate latent states from information reported to police about FV aggressors' harmful behaviors?

2. To what extent are there common patterns in FV aggressors' movements between latent states over two years?

Method

Data source

The archival data for this study came from a larger project on FV in New Zealand (Jolliffe Simpson et al., 2021). The index sample contained all FV episodes reported to New Zealand Police in the Integrated Safety Response (ISR) catchment areas of Waikato and parts of Canterbury between 1 November and 9 December 2018 (N = 2115). The ISR is a multi-agency FV response including representatives from organizations such as Police, District Health Boards, Oranga Tamariki (Ministry for Children), and Ara Poutama (Department of Corrections), among others. All FV episodes reported to police in the ISR's catchment areas are referred to the ISR for triaging, risk assessment, and case management.

Archival Information

Police Reports

Police recorded a narrative description of the episode, demographic characteristics, risk categories, and descriptive variables about the episode itself, including five dichotomous indicators for the presence of harm: verbal abuse (present in 92.2% of index episodes), physical harm (30.9%), sexual harm (0.3%), threats of harm (13.8%), and property damage (14.2%). The risk categories came from the Static Assessment of Family Violence Recidivism (SAFVR) and Dynamic Risk Assessment (DYRA); New Zealand police's two FV risk assessment instruments. We identified the predominant aggressor in each index episode based on the role police assigned to people in their report (e.g., perpetrator, suspect; New Zealand Police, n.d.). In cases where two or more individuals were labelled as 'mutual

participants', we identified the aggressor as the 'person posing risk' in the DYRA completed by attending officers (described below; Jolliffe Simpson et al., 2021).

SAFVR

The SAFVR is an actuarial risk assessment instrument designed to predict the likelihood an aggressor will be convicted of a FV-related offense within the next two years (Bissielo & Knight, 2016). The SAFVR contains eight static variables held by New Zealand Police and the Ministry of Justice and is updated daily by a computerized algorithm. The SAFVR has good predictive ability with an AUC of .77 for predicting FV offenses within two years (Bissielo & Knight, 2016) and .64 for predicting FV-related calls for police service within six months (Jolliffe Simpson et al., 2021). Raw SAFVR scores were unavailable; instead, risk was communicated through three categories: 'low', 'moderate', or 'high'.

DYRA

The DYRA (New Zealand Police, n.d.) is an actuarial instrument designed to estimate the likelihood of an aggressor committing further FV-related harm and inform the 3-day safety plan police compose following each call out. Police complete the DYRA by entering victim responses to a series of questions into an iPhone app that calculates the total score and risk bin ('low', 'moderate', or 'high') for the episode. The DYRA consists of ten questions asked in all cases, a further two questions asked if the episode contains IPV, and a further four questions asked if children normally reside at the scene address (New Zealand Police, n.d.). A recent study found the predictive ability of the DYRA was statistically inferior to the SAFVR for FV-related calls for police service within six months (AUC of .54; Jolliffe Simpson et al., 2021). We could access DYRA scores but chose to use the risk categories instead to remain consistent with the SAFVR and ISR risk categories.

ISR Risk Categories

We also had access to risk categories from the ISR. Triage teams, comprised of representatives from the ISR's partner organizations, meet daily to undertake collaborative structured professional judgement risk assessments for cases recently reported to police in their catchment area (Integrated Safety Response, 2019). For each case, team members read the police report for the most recent FV episode, share relevant information from their organization's organizations' databases, identify risk factors, and collaboratively determine a categorical risk level: 'low', 'medium', or 'high' (Integrated Safety Response, 2019). This risk category is then used to determine the level of concern for future episodes and the harm expected within them, and guides the intensity of interventions offered to the family. A recent study found the ISR triage teams' risk categories have an AUC of .60 for predicting FV-related calls for police service within six months; statistically inferior to the SAFVR, but superior to the DYRA (Jolliffe Simpson et al., 2021).

Sample Characteristics

Table 1 shows that around three quarters of aggressors were male, and most were identified as Māori or New Zealand European. Half of index episodes were between current intimate partners, and a further fifth were between former intimate partners, meaning that over two-thirds of aggressors in the sample had intimate partner violence index cases. A further one-fifth of index episodes were between parents and children, with the remainder being between siblings or people in other familial relationships. More than half of aggressors were categorized as high-risk on the SAFVR, with similar proportions having no score, or a low or moderate risk category. A third of aggressors were categorized as high-risk on the DYRA for the index episode. In contrast, the most common ISR risk category was medium, applying to more than half of aggressors. The ISR rated fewer than one in ten aggressors as high-risk.

Table 1

Descriptive Statistics for Sample (N = 2115)

Variable	*n*	%
Aggressor gender		
Gender diverse	3	0.1
Female	513	24.3
Male	1599	75.6
Aggressor ethnicity		
European	929	43.9
Māori	930	44.0
Other	256	12.1
Relationship to index victim		
Current partner	1051	49.7
Former partner	421	19.9
Parent or child	386	18.3
Sibling	128	6.1
Other	129	6.1
SAFVR		
No score [a]	315	14.9
Low	417	19.7
Moderate	275	13.0
High	1108	52.4
DYRA		
Not answered	297	14.0
Low	568	26.9
Moderate	493	23.3
High	757	35.8
ISR		
Low	876	41.4
Medium	1113	52.6
High	126	6.0

Note. DYRA = Dynamic Risk Assessment; SAFVR = Static Assessment for Family Violence Recidivism; ISR = Integrated Safety Response.
[a] People received no SAFVR score when they lacked the necessary criminal history for a score to be calculated (D. Scott, personal communication with first author, January 28, 2020).

Procedure

We collected all FV-related episodes reported to police during the two years following the index episode for the index aggressor, where they were again in the role of perpetrator, suspect, or mutual participant. We divided the two-year follow-up period into eight three-month intervals and recorded the presence of each of the five indicators of harm—

verbal abuse, physical, sexual, threats of harm, property damage—during each interval. If an aggressor had no episodes reported, they scored a 0 for all five types of harm. Each type of harm could only be scored once, even if an aggressor had multiple episodes reported during the same interval (e.g., a person with two episodes in one interval, one containing verbal abuse, and one containing verbal abuse and physical harm, would score 1 for verbal abuse and 1 for physical harm for that three-month interval).

Analysis

Recall our first research question: can we estimate latent states from information reported to police about FV aggressors' harmful behaviors? To answer this question, we used the five harm indicators from police reports for FV episodes (verbal abuse, physical harm, sexual harm, threats of harm, and property damage) to create hidden Markov models with two to eight latent states, using the LMest package in R (Bartolucci et al., 2017). We selected the best model for the data using the Bayesian Information Criterion (BIC; Jones et al., 2010; Zhang et al., 2010; Zucchini et al., 2016). For each state we described the probability of aggressors committing each of the five types of harm while in that state (Heckert & Gondolf, 2005; Jones et al., 2010).

Next, we used the Viterbi algorithm from the LMest package to estimate each aggressor's most likely latent state at each of the eight three-month intervals (Bartolucci et al., 2017), and described aggressors' demographic and risk characteristics by their initial latent state (i.e., their latent state during the first three-month interval). Hidden Markov models allow for non-linear movement between latent states over time (Jones et al., 2010; Zhang et al., 2010). This approach generates *average transition probabilities* that we used to estimate how likely a given aggressor was to move from one state to another, or remain in the same state, over time, given their most recent state (Zucchini et al., 2016).

The second research question was: to what extent are there common patterns in FV aggressors' movements between latent states over two years? To answer this question, we performed k-means clustering in IBM SPSS Statistics version 27 to find the most common pathways between the latent states during the two-year follow-up period. Finally, we examined the demographic and risk characteristics of groups of aggressors with similar pathways.

Results

We compared the BIC for hidden Markov models with two to eight latent states (see supplemental materials) and chose to proceed with the three-state model because it had the smallest BIC statistic (Jones et al., 2010). Figure 1 shows the probability of each of the five harm variables being present in the three latent states. The probabilities of the different types of harm did not add to one because they were separate dichotomous variables; instead, the probability of each type of harm being present or absent, in each state, added to one. The first state had a very small probability of verbal abuse (.02) and virtually no probability of all other types of harm (< .01); thus, it can be viewed as a state containing almost no harm. The second state in the model appeared to be "in between" in terms of severity. It had a small probability of verbal abuse (.18), very small probability of physical harm (.02) and threats (.01), and virtually no probability of property damage or sexual harm (< .01). The third state included an almost certain probability of verbal abuse (.96), moderate probability of physical harm (.43), small probabilities of threats (.23) and property damage (.22); and a very small probability of sexual harm (.01). Thus, this state was viewed as the highest in terms of severity.

Figure 1

Probabilities of the Presence of Five Different Types of Harm in Each of the Latent States in the Three State Model

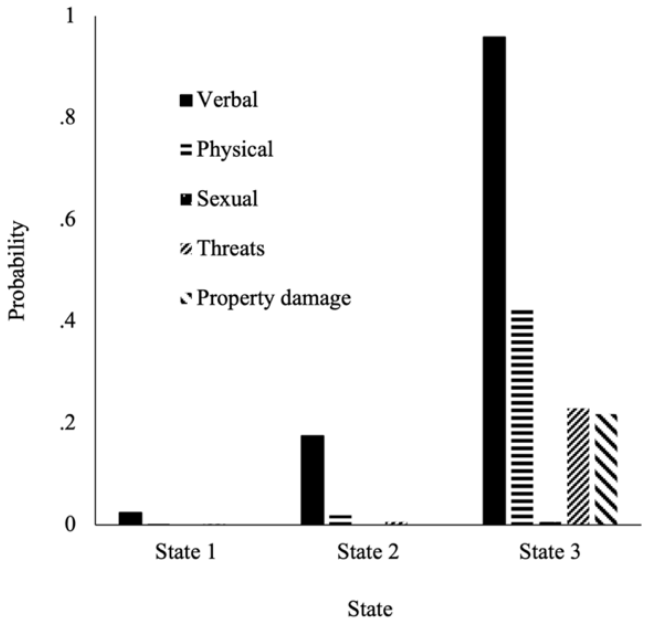


Table 2 describes the demographic and risk characteristics of aggressors by their initial latent state (i.e., in the three months following the index episode). There was a significant association between initial latent state and risk category on the SAFVR, DYRA and ISR measures, whereby state 1 (the least severe) had a greater proportion of aggressors in the lower risk tiers than state 2 and state 3 (the most severe). There was an association between initial latent state and aggressor gender; a greater proportion of aggressors were men in states 2 and 3 compared with state 1. There were also patterns in aggressors' relationship to the index victim, whereby the largest proportion of aggressors who were the current intimate partner of the index victim was in state 2, the largest proportion of aggressors who were former intimate partners was in state 3, and the largest proportion of aggressors with parent/child index episodes was in state 1. There was also an association between initial state and aggressor ethnicity; a greater proportion of Māori were in state 2 and 3 than in state 1 and a greater proportion of New Zealand European aggressors were in states 1 or 3 than in state 2.

Table 2

Descriptive Statistics for Aggressors by Initial Latent State (N = 2115)

	State				
Variable	1	2	3	χ^2	*p*
Aggressor gender [a]				16.44	< .001
Gender diverse	1 (0.1)	0 (0.0)	2 (0.3)		
Female	349 (27.1)	17 (14.9)	147 (20.6)		
Male	936 (72.8)	97 (85.1)	566 (79.2)		
Aggressor ethnicity				63.55	< .001
European	602 (46.8)	37 (32.5)	290 (40.6)		
Māori	488 (37.9)	70 (61.4)	372 (52.0)		
Other	196 (15.2)	7 (6.1)	53 (7.4)		
Relationship to index victim				37.58	< .001
Current partner	597 (46.4)	68 (59.6)	386 (54.0)		
Former partner	246 (19.1)	14 (12.3)	161 (22.5)		
Parent or child	266 (20.7)	14 (12.3)	106 (14.8)		
Sibling	87 (6.8)	7 (6.1)	34 (4.8)		
Other	90 (7.0)	11 (9.6)	28 (3.9)		
SAFVR				117.68	< .001
No score	260 (20.2)	7 (6.1)	48 (6.7)		
Low	277 (21.5)	15 (13.2)	125 (17.5)		
Moderate	183 (14.2)	11 (9.6)	81 (11.3)		
High	566 (44.0)	81 (71.1)	461 (64.5)		
DYRA				15.42	.017
Not answered	185 (14.4)	23 (20.2)	89 (12.4)		
Low	371 (28.8)	27 (23.7)	170 (23.8)		
Moderate	299 (23.3)	26 (22.8)	168 (23.5)		
High	431 (33.5)	38 (33.3)	288 (40.3)		
ISR				86.20	< .001
Low	633 (49.2)	37 (32.5)	206 (28.8)		
Medium	592 (46.0)	65 (57.0)	456 (63.8)		
High	61 (4.7)	12 (10.5)	53 (7.4)		

Note. DYRA = Dynamic Risk Assessment; SAFVR = Static Assessment for Family Violence Recidivism; ISR = Integrated Safety Response. State 1 *n* = 1286, 60.8% of cases; State 2 *n* = 114, 5.4%; State 3 *n* = 715, 33.8%.

[a] Gender diverse aggressors were excluded from chi square test.

[b] 95% confidence intervals based on 1000 bootstrap samples.

Table 3 shows the averaged transition probabilities for the three-state model (see supplemental materials for the unaveraged transition probabilities). The states in the rows represent the current latent state (t), and the states in the column represent the latent state at the next interval (t+1). The table shows that if a case was in state 1 at a given interval, the most likely outcome was that they remained in that state, with a very small probability of escalating directly to state 3. If a case was in state 2, they were most likely to remain in that state at the next interval, with a moderate probability of escalating to state 3, and a small probability of de-escalating to state 1. If a case was in state 3, they were most likely to remain in state 3, followed by lower probabilities of de-escalating to state 1, or state 2.

Table 3

Averaged Transition Probabilities for the Three State Model, with Averaged Standard Errors in Brackets

	t +1		
t	State 1	State 2	State 3
State 1	.95 (.03)	.01 (.02)	.05 (.02)
State 2	.10 (.12)	.53 (.15)	.38 (.10)
State 3	.33 (.06)	.26 (.05)	.41 (.04)

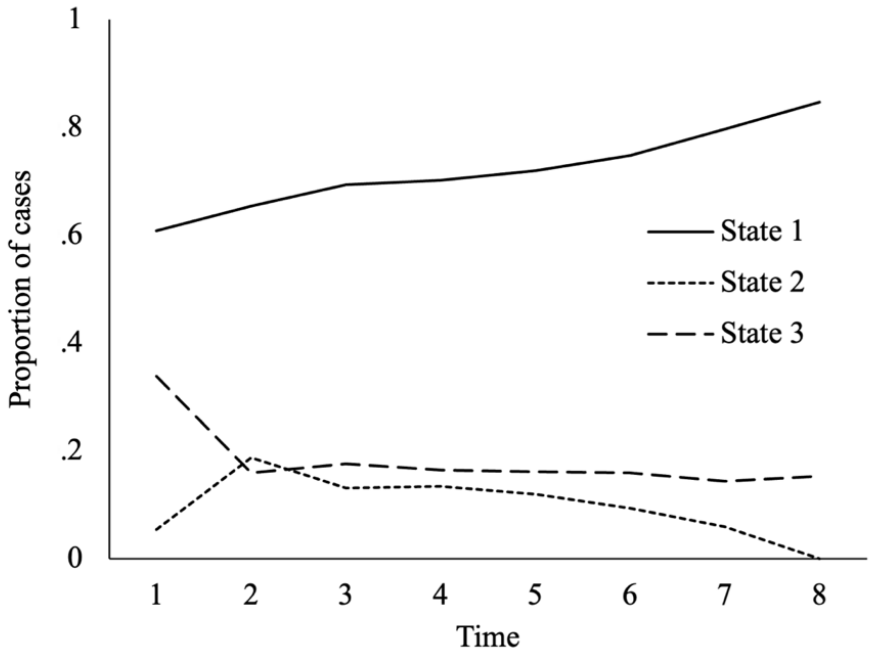
Note. BIC = 29312.19

Next, figure 2 shows the distribution of the three latent states across the two-year follow-up period. The proportion of aggressors in state 3 halved from 34% at time 1, to 16% at time 2 and 15% at time 3. The proportion of aggressors in state 2 increased from 5 to 19%, then steadily decreased, reaching 0% by time 8. As we would expect given the declining proportions of aggressors in states 2 and 3, there was a corresponding increase in the proportion of aggressors in state 1, which reached 85% by time 8. The raw figures support this pattern; the proportion of aggressors with any further episodes reported to police declined

in each 3-month period from 68% at time 1, to 61% at time 2, 56% at time 3, 51% at time 4, 46% at time 5, 39% at time 6, 29% at time 7, and 19% at time 8 (see supplemental materials).

Figure 2

Distribution of Latent States in the Three State Model During the Two-Year Follow-up Period

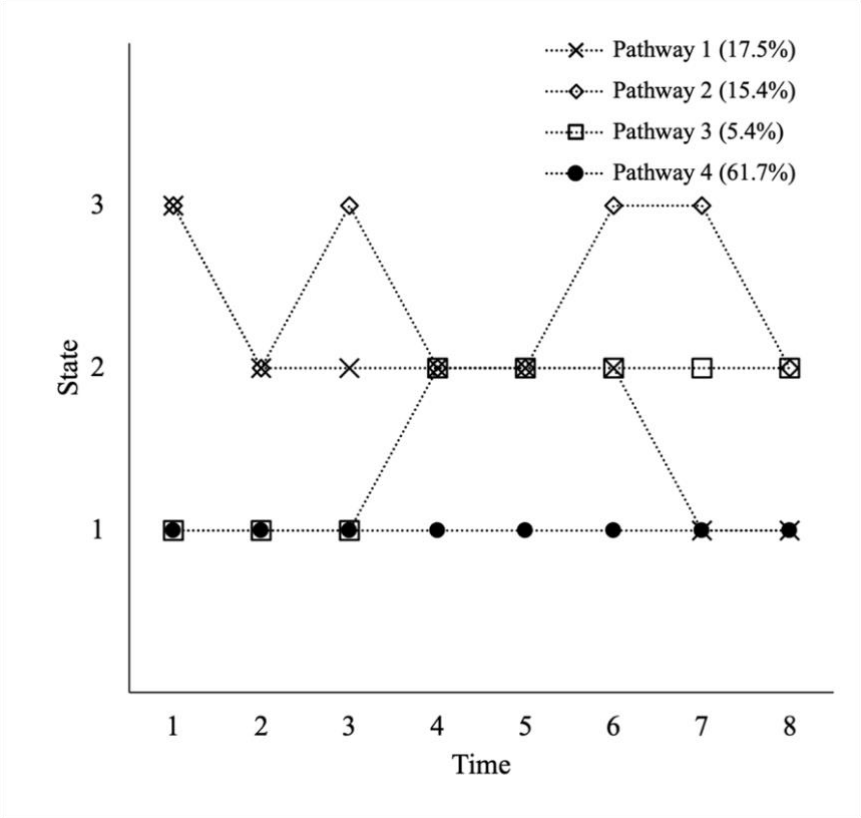


Recall the second research question: to what extent are there common patterns in FV aggressors' movements between latent states over two years? To answer this question, we performed cluster analysis of the aggressors' most likely latent state at each interval to identify the most common pathways between the latent states during the two-year follow-up period. We performed this analysis with two to eight clusters and chose to proceed with the four-cluster solution because the patterns in the cluster centers were sufficiently different to one another to interpret. Figure 3 shows the four pathways between the latent states identified in this cluster analysis. In the first pathway, which we called *De-escalation* (*n* = 371, 17.5%) aggressors were in state 3 in the first interval, then de-escalated to state 2 for five intervals,

then de-escalated further to state 1 for the remaining two intervals. In the second pathway, which we called *High harm* (*n* = 326, 15.4%), aggressors moved between states 3 and 2 continuously during the follow-up period. In the third, least common, pathway *Low harm* (*n* = 114, 5.4%) aggressors were in state 1 for the first three intervals, then escalated to state 2 for the remaining five intervals. We chose to call this pathway *Low harm* rather than *Escalation* because state 2 was still relatively minor in severity; characterized by a small probability of verbal abuse and very small or no probability of other types of harm. Lastly, in the fourth—but most common—pathway of *No harm* (*n* = 1304, 61.7%), aggressors remained in state 1 for the entire follow-up period.

Figure 3

Four Pathways between the Three Latent States During the Two-Year Follow-up Period



Finally, Table 4 describes the demographic and risk characteristics of aggressors by pathway. There were significant associations between the SAFVR, DYRA and ISR risk categories and the pathways, but the associations were not straightforward. For the SAFVR, the greatest proportion of high-risk aggressors was in the *High harm* pathway, followed by the *Low harm* pathway and *De-escalation* pathway. However, almost half of aggressors in the *No harm* pathway were also rated as high-risk. The *High harm* and *De-escalation* pathways had a higher proportion of aggressors assigned medium- and high-risk categories by the ISR after the index episode, compared with the *No harm* and *Low harm* pathways. The *De-escalation* pathway had the greatest proportion of cases assigned the high-risk category of the DYRA at the index episode. With that said, the distribution of aggressors in the DYRA risk categories did not appear to differ greatly between the pathways, and the chi-square statistic for this comparison may only have been statistically significant due to the large sample size.

There were also associations between the pathways and aggressor gender and ethnicity; a greater proportion of men and Māori were in the *Low harm, High harm, and De-escalation* pathways, compared with the *No harm* pathway. Finally, the *High harm* pathway had the largest proportion of aggressors who were the current intimate partner of the index victim, whereas the *No harm* pathway had the largest proportion of aggressors who were former partners of the index victim.

Table 4

Descriptive Statistics for Aggressors by Pathway (N = 2115)

	Pathway					
	1	2	3	4		
Variable	*De-escalation*	*High harm*	*Low harm*	*No harm*	χ^2	*p*
Aggressor gender [a]					33.42	< .001
Gender diverse	2 (0.5)	0 (0.0)	0 (0.0)	1 (0.1)		
Female	65 (175)	53 (16.3)	24 (21.1)	371 (28.5)		
Male	304 (81.9)	273 (83.7)	90 (78.9)	932 (71.5)		
Aggressor ethnicity					108.29	< .001
European	152 (41.0)	106 (32.5)	47 (41.2)	624 (47.9)		
Māori	191 (51.5)	206 (63.2)	60 (52.6)	473 (36.3)		
Other	28 (7.5)	14 (4.3)	7 (6.1)	207 (159)		
Relationship to index victim					29.07	.004
Current partner	192 (51.8)	195 (59.8)	59 (51.8)	605 (46.4)		
Former partner	73 (19.7)	50 (15.3)	19 (16.7)	279 (21.4)		
Parent or child	68 (18.3)	46 (14.1)	19 (16.7)	253 (19.4)		
Sibling	14 (3.8)	24 (7.4)	8 (7.0)	82 (6.3)		
Other	24 (6.5)	11 (3.4)	9 (7.9)	85 (6.5)		
SAFVR					182.69	< .001
No score	26 (0.7)	10 (3.1)	8 (0.7)	271 (20.8)		
Low	73 (19.7)	33 (10.1)	19 (16.7)	292 (22.4)		
Moderate	43 (11.6)	30 (9.2)	13 (11.4)	189 (14.5)		
High	229 (61.7)	253 (77.6)	74 (64.9)	552 (42.3)		
DYRA					17.33	.044
Not answered	53 (14.3)	60 (18.4)	21 (18.4)	163 (12.5)		
Low	81 (21.8)	86 (26.4)	31 (27.2)	370 (28.4)		
Moderate	90 (24.3)	66 (20.2)	22 (19.3)	315 (24.2)		
High	147 (39.6)	114 (35)	40 (35.1)	456 (35.0)		
ISR					67.49	< .001
Low	112 (30.2)	90 (27.6)	50 (43.9)	624 (47.9)		
Medium	233 (62.8)	212 (65)	57 (50)	611 (46.9)		
High	26 (7.0)	24 (7.4)	7 (6.1)	69 (5.3)		

Note. DYRA = Dynamic Risk Assessment; SAFVR = Static Assessment for Family Violence Recidivism; ISR = Integrated Safety Response. Pathway 1 *n* = 371, 17.5% of sample; Pathway 2 *n* = 326, 15.4%; Pathway 3 *n* = 114, 5.4%; Pathway 4 *n* = 1304, 61.7%.

[a] Gender diverse aggressors were excluded from chi square test.

[b] 95% confidence intervals based on 1000 bootstrap sample.

Discussion

In this study we sought to understand the behavioral patterns of FV aggressors in Aotearoa New Zealand. We constructed latent states based on information New Zealand Police recorded about the harm aggressors committed, modeled aggressors' movement between those states during a two-year period, and identified common pathways in aggressor behavior. Our model contained three latent states, each with different probabilities of five types of harm occurring. The first state contained low or no harm, the second contained small probabilities of harm (with verbal abuse being most likely), and the third state involved an almost certain probability of verbal abuse and a moderate probability of physical violence.

When we modeled aggressors' movement between the latent states over two years, we found that those who were in the least severe state at a given point in time were also the least likely to change states, compared with aggressors in the most severe state who were the most likely to change. Moreover, the most common pathway between the latent states involved virtually no harm over the two-year follow-up period, the next most common pathway was characterized by de-escalation, followed by a high harm and a low harm pathway. The *Low harm* pathway did exhibit escalation from state 1 to state 2, but state 2 was still low in severity. Hence, rather than truly indicating escalation, that pattern could indicate families chose to report low-level FV episodes (e.g., verbal arguments). The raw data also indicated the proportion of aggressors with any further episodes reported to police gradually decreased. Taken together, these findings support other research indicating that FV aggressors do not *inevitably* continue or escalate their harmful behavior, that a small subset of cases are responsible for a disproportionately large amount of harm, and that police data suggests many people desist from FV (Bland & Ariel, 2015; Piquero et al., 2006; Walker et al., 2013).

We also examined the demographic and risk characteristics of aggressors by their initial latent state and their pathway during the two-year follow-up period, and found men

and Māori were disproportionately represented in the states and pathways with higher probabilities of harm. In addition, there was an association between aggressors' SAFVR risk categories and their pathways during the follow-up period. The SAFVR categories were calculated on the day of the index episode from information about aggressors' criminal and FV histories, with the instrument being designed to predict FV-related offending within two years (Bissielo & Knight, 2016). The relationship between the SAFVR risk categories and pathways in this research is promising and indicates risk assessment instruments could be used to predict pathways rather than just single dichotomous outcomes (Jones et al., 2010). However, further work is necessary to find whether predicting trajectories also predicts outcomes further in the future.

The latent states in this study contained smaller probabilities of harm than latent states produced in a study by Jones et al. (2010), where the most severe contained high probabilities of all types of harm they measured, including physical violence. This difference in severity is probably because we used a sample of calls for police service, rather than focusing on men referred to non-violence programs (Jones et al., 2010). Our results are in line with other studies showing the most common pathway for people experiencing FV contains low or no harm, and a small subset of cases involve most of the harm that occurs (Bland & Ariel, 2015; Swartout et al., 2012). Our results are also in line with Gulliver and Fanslow's (2015) research on patterns of FV in New Zealand. Based on survey data, they also derived three classes of responses describing women's experiences of IPV victimization. The most common class involved no, or low levels of, harm, and the second class involved less severe and frequent harm than the third class, which involved high rates of severe IPV (Gulliver & Fanslow, 2015).

Implications

This research generates suggestions for changes to how agencies conceptualize and assess risk for FV. Firstly, the results challenge the longstanding presumption of escalation and align with the growing body of evidence showing aggressors do not *inevitably* continue or escalate their harmful behavior (Barnham et al., 2017; Bland & Ariel, 2015). That means further efforts should be undertaken to improve our ability to identify the small subset of families who do experience ongoing episodes and account for a disproportionate amount of harm (Robinson & Clancy, 2020). Many risk assessment instruments in current use are risk-averse; being designed to triage out low-risk cases and otherwise over-predicting the likelihood of recurrence (Jolliffe Simpson, 2021). Indeed, in this study the SAFVR instrument had the strongest relationship with the pathways, but even then, two-fifths of aggressors in the *No harm* pathway had high SAFVR risk categories. Therefore, a need exists for further validation of existing instruments so they can be used to identify the highest-risk cases (Robinson & Clancy, 2020).

In addition, some of the approaches employed in this research could be embedded into existing risk assessment instruments. For example, latent states generated from information about aggressor behavior could be used as outcomes in risk assessment instruments (Heckert & Gondolf, 2005; Jones et al., 2010). In practice, this would entail further work to identify latent states from a larger sample of information about aggressor behavior, and to validate existing risk assessment instruments to predict the likelihood of those states. Conceptualizing risk in this way would mean the outcomes that instruments are designed to predict would align with the outcomes that practitioners responding to FV have in mind (Heckert & Gondolf, 2005). For instance, one state may describe situations where families are likely to experience ongoing conflict and verbal arguments, whereas another may describe a more volatile situation with a greater probability of multiple forms of harm. Alternatively,

instruments could be created to predict a given aggressor's most likely pathway (Jones et al., 2010; Zhang et al., 2010). These changes would preserve the simple design of risk assessment instruments and allow agencies to predict the likelihood of the state or pathway they are most interested in, based on their focus (e.g., ensuring immediate safety versus longer-term case management and preventing cumulative harm).

On the other hand, the results from this study could contribute to an entirely different way of thinking about risk. The risk categories presently used to communicate scores from risk assessments are abstract labels (e.g., low, medium, or high), that are interpreted differently by different people, and applied inconsistently across types of offending (e.g., sex offending versus violence; Hanson et al., 2017). Instead of predicting one outcome, the transition matrix from a hidden Markov model communicates the probability of multiple outcomes (e.g., the situation staying the same, de-escalating, or escalating), given the aggressor's most recent state (Jones et al., 2010; Zhang et al., 2010; Zucchini et al., 2016). Hence, risk categories could be replaced by a description of a given aggressor's most likely state and the types of harm that most probable within that state. Such a description would be both informative and uniformly interpreted; thus, could provide value to practice (Hanson et al., 2017).

For example, an instrument could indicate that based on current circumstances, a given aggressor was recently in a state involving high harm (e.g., physical violence and psychological harm), and has a .2 probability of remaining in that state, a .3 probability of de-escalating to a state involving low-level harm (e.g., verbal abuse only), and a .5 probability of de-escalating to a state involving moderate-level harm (e.g., psychological abuse and threats, with a small probability of physical harm). This type of information may prompt practitioners to consider how to support an aggressor's transition from their current state to a comparatively safer state, based on the probability of that transition happening. Of course,

these probabilities would only form part of the practitioner's decision making and should be supplemented by other information. Nevertheless, this approach could provide FV practitioners with a more complete picture of the risk posed in each case, compared with the risk categories in current use (Hanson et al. 2017).

Limitations

This study has limitations that should be addressed in further research. First, by using police-recorded information, we modeled harm that came to the attention of police and missed information that was not reported. This limitation is the reason why we opted for a hidden Markov model, which explicitly uses measured information to model unmeasurable constructs (Jones et al., 2010; Zhang et al., 2010; Zucchini et al., 2016). In addition, because most risk assessment for FV occurs on the front line and is performed by police officers, the ability to complete those assessments with administrative data is an important aspect of ecological validity. With that said, further research should use the hidden Markov approach to generate latent states from interviews with victims about their day-to-day experiences to reveal whether states generated from police information resemble families' actual experiences.

Another limitation of this research is that the police harm indicators covered verbal abuse, physical harm, sexual harm, threats of harm, and property damage, but not psychological harm, coercive control, or stalking (Robinson et al., 2018). These are important components of FV that detrimentally impact victims but are often sub-criminal and go unmeasured (Ansara & Hindin, 2011; Dichter et al., 2018; Wiener, 2017). Research incorporating these types of harm may improve our understanding of families' experiences; for example, there may be a distinct state involving coercive controlling behaviors, or those behaviors may accompany (or even precede) physical violence. Moreover, by focusing on the harm that individual aggressors committed, we could not account for bi-directional harm, or

dyadic patterns where people switch between the roles of aggressor and victim (Straus, 2015; Straus & Gozjolko, 2016), and we did not know whether episodes reported to police during the follow-up period involved the same victim as the index event. Integrating a wider range of information such as psychological abuse and other family members' characteristics and behavior would further improve the holistic and descriptive nature of the model presented in this study.

Finally, we did not have information about whether aggressors and their families were offered intervention services during the follow up period. This is an important limitation because all cases in the sample were referred to the ISR for risk assessment and case management, and in theory, some cases could have engaged with interventions that changed the pathway they may have followed without treatment. Further research should account for those interventions to show how they align with patterns of harm. Furthermore, the hidden Markov technique could be used to explore whether engagement with a given intervention is associated with a corresponding de-escalation in an aggressor's behavioral state.

Conclusion

In this study, we modeled the behavioral patterns of 2115 FV aggressors using police-recorded information. We estimated three latent states of harm and identified four pathways between those states during a two-year period. Taken together, the results suggest information police routinely collect about FV cases can be used to model latent states and aggressors' behavioral pathways through those states. The results of this study generated ideas for further work that could improve how risk for FV is conceptualized, assessed, and communicated in practice.

References

Ansara, D. L., & Hindin, M. J. (2011). Psychosocial consequences of intimate partner violence for women and men in Canada. *Journal of Interpersonal Violence*, *26*(8), 1628-1645. https://doi.org/10.1177/0886260510370600

Arias-Rivera, S., Hidalgo, V., & Lorence, B. (2020). A scoping study on measures of child-to-parent violence. *Aggression and Violent Behavior*, *52*, 101426. https://doi.org/10.1016/j.avb.2020.101426

Barnham, L., Barnes, G. C., & Sherman, L. W. (2017). Targeting escalation of intimate partner violence: Evidence from 52,000 offenders. *Cambridge Journal of Evidence-Based Policing*, *1*(2-3), 116-142. https://doi.org/10.1007/s41887-017-0008-9

Bartolucci, F., Pandolfi, S., & Pennoni, F. (2017). LMest: An R package for latent Markov models for longitudinal categorical data. *Journal of Statistical Software, 81*(4), 1-38. https://doi.org/10.18637/jss.v081.i04

Bissielo, A., & Knight, G. (2016). *Family violence risk assessment redevelopment: Static risk score* (New Zealand Police internal report). New Zealand Police.

Bland, M. (2015). M.St. Thesis in applied criminology and police management, Institute of Criminology, University of Cambridge.

Bland, M., & Ariel, B. (2015). Targeting escalation in reported domestic abuse: Evidence from 36,000 callouts. *International Criminal Justice Review*, *25*(1), 30-53. https://doi.org/10.1177/1057567715574382

Boxall, H., Rosevear, L., & Payne, J. (2015). Domestic violence typologies: What value to practice? *Trends and Issues in Crime and Criminal Justice*, *494*, 1-9. https://www.aic.gov.au/publications/tandi/tandi494

Chan, K. L., Chen, Q., & Chen, M. (2021). Prevalence and correlates of the co-occurrence of family violence: A meta-analysis on family polyvictimization. *Trauma, Violence and Abuse*, *22*(2), 289-305. https://doi.org/10.1177/1524838019841601

Condry, R., & Miles, C. (2014). Adolescent to parent violence: Framing and mapping a hidden problem. *Criminology & Criminal Justice*, *14*(3), 257-275. https://doi.org/10.1177/1748895813500155

Dichter, M. E., Thomas, K. A., Crits-Christoph, P., Ogden, S. N., & Rhodes, K. V. (2018). Coercive control in intimate partner violence: Relationship with women's experience of violence, use of violence, and danger. *Psychology of Violence*, *8*(5), 596-604. https://doi.org/10.1037/vio0000158

Dixon, L., & Browne, K. (2003). The heterogeneity of spouse abuse: A review. *Aggression and Violent Behavior*, *8*(1), 107-130. https://doi.org/10.1016/S1359-1789(02)00104-0

Dixon, L., Hamilton-Giachritsis, C., Browne, K., & Ostapuik, E. (2007). The co-occurrence of child and intimate partner maltreatment in the family: Characteristics of the violent perpetrators. *Journal of Family Violence*, *22*(8), 675-689. https://doi.org/10.1007/s10896-007-9115-x

Ensor, B. & Cooke, H. (2019). The homicide report: NZ's family violence record 'horrific', says Jacinda Ardern. *Stuff.* https://www.stuff.co.nz/national/crime/112724319/the-homicide-report-nzs-family-violence-record-horrific-says-jacinda-ardern

Family Violence Act (2018). http://www.legislation.govt.nz/act/public/2018/0046/latest/DLM7159322.html

Feld, S. L., & Straus, M. A. (1989). Escalation and desistance of wife assault in marriage, *Criminology*, *27*(1), 141-162. https://doi.org/10.1111/j.1745-9125.1989.tb00866.x

Gulliver, P., & Fanslow, J. L. (2015). The Johnson typologies of intimate partner violence: An investigation of their representation in a general population of New Zealand

women. *Journal of Child Custody*, *12*, 25-46. https://doi.org/10.1080/15379418.2015.1037051

Hanson, R. K., Bourgon, G., McGrath, R. J., Kroner, D., D'Amora, D. A., Thomas, S. S., & Tavarez, L. P. (2017). *A five-level risk and needs system: Maximizing assessment results in corrections through the development of a common language*. National Reentry Resource Center. https://saratso.org/pdf/A_Five_Level_Risk_and_Needs_System_Report.pdf

Heckert, D. A., & Gondolf, E. W. (2005). Do multiple outcomes and conditional factors improve prediction of batterer reassault? *Violence and Victims*, *20*(1), 3-24. https://doi.org/10.1891/vivi.2005.20.1.3

Henning, K., Campbell, C. M., Stewart, G., & Johnson, J. (2021). Prioritizing police investigations of intimate partner violence using actuarial risk assessment. *Journal of Police and Criminal Psychology*, *36*(4), 667-678. https://doi.org/10.1007/s11896-021-09466-7

Ibabe, I. (2020). A systematic review of youth-to-parent aggression: Conceptualization, typologies, and instruments. *Frontiers in Psychology*, *11*, 1-18. https://doi.org/10.3389/fpsyg.2020.577757

Integrated Safety Response. (2019). *12 week review of cases referred to the Integrated Safety Response (ISR) pilot*. https://www.police.govt.nz/sites/default/files/publications/isr-12-week-case-review-report.pdf

Jackson, S. L. (2016). All elder abuse perpetrators are not alike: The heterogeneity of elder abuse perpetrators and implications for intervention. *International Journal of Offender Therapy and Comparative Criminology*, *60*(3), 265-285. https://doi.org/10.1177/0306624X14554063

Jones, A. S., Heckert, D. A., Gondolf, E. D., Zhang, Q., & Ip, E. H. (2010). Complex behavioral patterns and trajectories of domestic violence offenders. *Violence and Victims, 25*(1), 3-17. https://doi.org/10.1891/0886-6708.25.1.3

Kim, B., & Merlo, A. V. (2021). Domestic homicide: A synthesis of systematic review evidence. *Trauma, Violence, & Abuse,* XX(X), 1-18. https://doi.org/10.1177/15248380211043812

McEwan, T. E., Shea, D. E., & Ogloff, J. R. P. (2018). An acturial instrument for police triage of Australian family violence reports. *Criminal Justice and Behavior, 46*(4), 590–607. https://doi.org/10.1177/0093854818806031

Moody, G., Cannings-John, R., Hood, K., Kemp, A., & Robling, M. (2018). Establishing the international prevalence of self-reported child maltreatment: A systematic review by maltreatment type and gender. *BMC Public Health*, *18*(1), 1-15. https://doi.org/10.1186/s12889-018-6044-y

Mossman, E., Paulin, J., & Wehipeihana, N. (2017). *Evaluation of the family violence Integrated Safety Response pilot*. SUPERU. https://thehub.sia.govt.nz/assets/documents/ISR_pilot_evaluation_FINAL.pdf

New Zealand Crime and Victims Survey (2018). *Topical report: Offences against New Zealand adults by family members*. Ministry of Justice. https://www.justice.govt.nz/assets/Documents/Publications/9ZU3Q-NZCVS-topical-report-Offences-by-family-members-Cycle-1-2018.pdf

New Zealand Family Violence Death Review Committee (2021a). *Child abuse and neglect deaths in Aotearoa New Zealand: Fact sheets*. https://www.hqsc.govt.nz/assets/FVDRC/Publications/FVDRC_2021_CAN_English_web2.pdf

New Zealand Family Violence Death Review Committee (2021b). *Intimate partner violence deaths in Aotearoa New Zealand: Fact sheets*. https://www.hqsc.govt.nz/assets/FVDRC/Publications/FVDRC_2021_IPV_English_web.pdf

New Zealand Police (2022). *Daily Occurrences of Crime and Family Violence Investigations*. https://www.police.govt.nz/about-us/statistics-and-publications/data-and-statistics/daily-occurrences-crime

Pagelow, M. D. (1981). Woman-battering: Victims and their experiences. Beverly Hills, CA: Sage. https://doi.org/10.2307/2149761

Piquero, A. R., Brame, R., Fagan, J., & Moffitt, T. E. (2006). Assessing the offending activity of criminal domestic violence suspects: Offense specialization, escalation, and de-escalation evidence from the spouse assault replication program. *Public Health Reports*, *121*(4), 409-418. https://doi.org/10.1177/003335490612100409

Richards, L. (2009). *Domestic Abuse, Stalking and Harassment and Honour Based Violence (DASH, 2009) Risk Identification and Assessment and Management Model*. https://www.dashriskchecklist.co.uk/wp-content/uploads/2016/09/DASH-2009.pdf

Robinson, A. L., & Clancy, A. (2020). Systematically identifying and prioritising domestic abuse perpetrators for targeted intervention. *Criminology & Criminal Justice*, *21*(5), 687-704. https://doi.org/10.1177/1748895820914380

Robinson, A. L., Myhill, A., & Wire, J. (2018). Practitioner (mis) understandings of coercive control in England and Wales. *Criminology & Criminal Justice*, *18*(1), 29-49. https://doi.org/10.1177/1748895817728381

Saxton, M. D., Jaffe, P. G., Dawson, M., Straatman, A.-L., & Olszowy, L. (2022). Complexities of the police response to intimate partner violence: Police officers'

perspectives on the challenges of keeping families safe. *Journal of Interpersonal Violence, 37*(5-6), 2557-2580. https://doi.org/10.1177/0886260520934428

Stoltenborgh, M., Bakermans-Kranenburg, M. J., Alink, L. R., & van IJzendoorn, M. H. (2015). The prevalence of child maltreatment across the globe: Review of a series of meta-analyses. *Child Abuse Review*, *24*(1), 37-50. https://doi.org/10.1002/car.2353

Storey, J. E. (2020). Risk factors for elder abuse and neglect: A review of the literature. *Aggression and Violent Behavior*, *50*, 101339. https://doi.org/10.1016/j.avb.2019.101339

Straus, M. A. (2015). Dyadic concordance and discordance in family violence: A powerful and practical approach to research and practice. *Aggression and Violent Behavior*, *24*, 83-94. https://doi.org/10.1016/j.avb.2015.04.011

Straus, M. A., & Gozjolko, K. L. (2016). Concordance between partners in "intimate terrorism"; A comparison of two typologies. *Aggression and Violent Behavior*, *29*, 55-60. https://doi.org/10.1016/j.avb.2016.06.003

Swartout, K. M., Cook, S. L., & White, J. W. (2012). Trajectories of intimate partner violence victimization. *Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health*, *13*(3), 272-277. https://doi.org/10.5811/westjem.2012.3.11788

Tucker, C. J., Finkelhor, D., Shattuck, A. M., & Turner, H. (2013). Prevalence and correlates of sibling victimization types. *Child Abuse & Neglect*, *37*(4), 213-223. https://doi.org/10.1016/j.chiabu.2013.01.006

Walker, K., Bowen, E., & Brown, S. (2013). Desistance from intimate partner violence: A critical review. *Aggression and Violent Behavior*, *18*(2), 271-280. https://doi.org/10.1016/j.avb.2012.11.019

Wiener, C. (2017). Seeing what is 'invisible in plain sight': Policing coercive control. *The Howard Journal of Crime and Justice*, *56*(4), 500-515. https://doi.org/10.1111/hojo.12227

Zhang, Q., Jones, A. S., Rijmen, F., & Ip, E. H. (2010). Multivariate discrete hidden Markov models for domain-based measurements and assessment of risk factors in child development. *Journal of Computational and Graphical Statistics*, *19*(3), 746-765. https://doi.org/10.1198/jcgs.2010.09015

Zucchini, W., MacDonald, I. L., & Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R* (2nd ed.). CRC Press. https://doi.org/10.1201/b20790

Chapter Six: General Discussion

In this thesis, we examined the risk assessments that New Zealand Police and the Integrated Safety Response (ISR) used to guide their decision making when responding to Family Violence (FV) cases, explored the utility of the ISR's assessments for case management, and modeled the behavioral patterns of FV aggressors over a two-year period. The purpose of this research was to contribute to the evidence base about risk assessment for FV aggressors in Aotearoa New Zealand; and in doing so, identify strengths, weaknesses, and opportunities for improvement. Accordingly, two overarching research questions guided the research:

1. What are the current risk assessment processes used for FV in Aotearoa New Zealand?
2. How can we improve risk assessment for FV in Aotearoa New Zealand?

We also addressed additional aims throughout. For example, we examined differences in the risk categories' performance for sub-groups of cases created according to aggressor characteristics and aggressor-victim relationships, and we sought to highlight relevant implications of the findings for FV practitioners. In this chapter, we integrate and discuss the main findings from across the research presented in chapters two to five. We outline the theoretical and practical implications of those insights and discuss some of the limitations present in this research and sharing suggestions for future research.

What are the current risk assessment processes used for FV in Aotearoa New Zealand?

Risk Assessment for FV Cases by New Zealand Police

New Zealand Police are usually the first organization that responds to reported FV episodes, and once at the scene of an episode, they use the actuarial Static Assessment for Family Violence Recidivism (SAFVR) and Dynamic Risk Assessment (DYRA) instruments

to determine a) the aggressors' risk of committing a FV-related offense within two years, and b) the likelihood of further FV occurring in the 72 hours following the episode. Taken together, these two instruments guide police's safety plan, designed to prevent harm until further risk assessment and intervention service allocation takes place. The findings from manuscript one, presented in chapter two, suggested these instruments had a poor ability to discriminate between aggressors who had a recurrence reported to police and those who did not have a recurrence. Most AUC statistics were better than chance, but had small effect sizes, and both instruments over-predicted recurrence, with the SAFVR outperforming the DYRA. The Area Under the Curve (AUC) statistics for both instruments were inferior to the average AUCs calculated for widely used IPV risk assessment instruments in an international meta-analysis (van der Put, 2019), as well as AUCs generated for the FVRAT, a comparable instrument in Australia (Dowling, & Morgan, 2019).

Risk Assessment for FV Cases by the ISR

Having described New Zealand Police's risk assessment instruments and examined their predictive validity, we turned our attention to the next phase of the FV response and explored the risk assessment process the ISR uses to triage cases and allocate intervention services. Unlike police's actuarial instruments, ISR triage teams use a relatively unstructured approach loosely described as Structured Professional Judgement (SPJ). The findings from manuscript two, presented in chapter three, suggested the predictive validity of the ISR triage teams' risk assessments was comparable to the AUCs produced for the DYRA and SAFVR in manuscript one. Overall, the triage teams' risk categories had a poor ability to discriminate between aggressors who had a recurrence reported to police and aggressors who did not have a recurrence; with AUC statistics that were superior to the DYRA, but inferior to the SAFVR.

In the fields of FV and IPV, SPJ is often used to describe approaches like the ISR's, which are more akin to unstructured clinical judgement or consensus-based decision-making

than SPJ (e.g., DASH; Richards, 2009). Because the ISR's risk assessment process allowed practitioners to select risk factors, rather than completing a prescribed instrument, we documented the factors triage team members recorded during their risk assessment. We were interested to find a) whether those factors were associated with the risk categories triage teams allocated, and b) whether those factors were associated with FV-related outcomes. We found triage teams recorded many various factors, and most were associated with the risk categories, but fewer than half were associated with FV-related outcomes.

Following this work, we pondered whether the factors ISR triage teams recorded added value to their risk assessments by highlighting barriers to responsivity and informing case management in addition to (or rather than) contributing to the assessments' ability to predict FV recurrence. Accordingly, in chapter four, we sought to understand the value of the ISR triage teams' risk assessments for case management though a latent class analysis of the teams' recorded factors. We found the factors could be used to generate three broad—but meaningful— classes of cases with implications for their treatment and management, which indicated the ISR triage teams' assessments also served as case management planning sessions.

How can we improve risk assessment for FV in Aotearoa New Zealand?

Validate and Improve New Zealand Police's Risk Assessment Instruments

In addressing the first overarching research question, we found the DYRA and SAFVR instruments had poor predictive validity and both substantially over-predicted recurrence. These findings also answered the second overarching research question by indicating that the DYRA should undergo further development. The SAFVR instrument was created by systematically identifying administrative information held by New Zealand Police and the Ministry of Justice that was statistically associated with FV-related offending within two years (Bissielo & Knight, 2016), and the SAFVR's predictive validity has been

previously established (Heister, 2018). Indeed, in this research we found the SAFVR was the best-performing instrument. Yet, the SAFVR still over-predicted recurrence, which could suggest it is a useful triage tool for screening out low risk cases but would require adjustment to be used to identify the aggressors and families at greatest risk of harm. On the other hand, the research presented in this thesis was the first to examine the predictive validity of the DYRA. The results showed the DYRA was poorly calibrated with recurrence outcomes, and further work is required to find optimal score thresholds so that the DYRA's risk categories adequately distinguish between lower and higher risk aggressors (Hanson et al., 2017). Promisingly, in response to these findings, New Zealand Police has already commissioned further validation of the DYRA instrument and its items, and is now considering how to improve the instrument (D. Polaschek, personal communication with author, 11 August, 2020).

Structure and Validate the ISR's Risk Assessments

In addition, while addressing the first overarching research question with relation to the ISR risk categories, we also answered the second overarching research question by concluding the ISR's risk assessment protocol would benefit from increased structure and does not currently meet the definition of SPJ. To our knowledge, the predictive validity of the ISR triage teams' risk assessments has not been previously explored in research, and manuscript two provides an example of how to examine similarly unstructured risk assessments performed by practitioners. We found the risk categories produced from the triage teams' assessments had poor predictive validity, and fewer than half of the factors triage teams recorded were associated with recurrence or physical recurrence. The triage teams' risk assessments did not include several aspects integral to the SPJ approach; although the assessments were supported by a guide of empirically based risk factors (Integrated Safety Response, 2018), triage teams could include information outside the guide and were

not required to systematically record the presence of the factors the guide contained. We suggest structure should be added to these assessments by requiring practitioners to score the guide like a SPJ instrument or by implementing another established structured instrument (e.g., LS/CMI; Andrews at al., 2004). This would systematically draw practitioners' attention to each risk factor—which could improve the assessments' predictive validity—and create data that could be used to systematically validate and improve the ISR triage teams' risk assessments (e.g., Spivak et al., 2020).

Examine Predictive Validity for Multiple Outcomes

In this thesis we sought to overcome one of the main limitations of existing risk assessment research; the reliance on single dichotomous outcomes (Heckert & Gondolf, 2005; Jones et al., 2010). FV risk assessment instruments are often validated based on their ability to predict a single outcome (e.g., reoffending), hence it follows that some instruments could potentially only reflect risk for that outcome and not the sub-criminal forms of harm that families and whānau experience like verbal abuse, property damage, or threats (Heckert & Gondolf, 2005). Therefore, we examined instruments' predictive validity for multiple outcomes that we thought more accurately reflected the goals of practitioners trying to decide how to respond to cases (e.g., planning a short-term response, versus longer-term interventions and harm reduction).

In manuscript one, we used three outcomes: recurrence (i.e., another FV-related call for police service) within three days (i.e., during the 72-hour safety plan), recurrence within 12 weeks, and recurrence within 24 weeks. Recurrence captured both "argument only" episodes and offenses, and using different follow-up lengths allowed us to explore how well the DYRA achieved its goal of predicting short-term risk and compare the DYRA to the SAFVR, which had previously been validated against the outcome of FV-related offending within 2 years (Bissielo & Knight, 2016; New Zealand Police, n.d.). Then, in manuscript two

we used the ISR triage teams' factors and risk categories to predict both recurrence and physical recurrence within 24 weeks. We chose these outcomes because according to the ISR risk assessment guide, the risk categories should capture triage teams' level of concern that FV would occur in the future and the level of harm they thought would occur as a result (Integrated Safety Response, 2018).

Finally, in manuscript three, presented in chapter five, we modeled the behavioral patterns of FV aggressors. In this study we estimated latent states from the information reported to police about aggressors' harmful behaviors and used those states as the outcome measure (Jones et al., 2010; Zucchini et al., 2016). We combined five indicators for the presence of harm available in police reports: verbal abuse, physical harm, sexual harm, threats of harm, and property damage, generating a more complete picture of the harm that aggressors commit and that their families and whānau experience compared with single dichotomous indicators like reoffending (Heckert & Gondolf, 2005; Jones et al., 2010; Zhang et al., 2010). We concluded manuscript three by summarizing how communicating the probability of aggressors moving to defined latent states with multiple outcomes, rather than communicating risk for a single dichotomous outcome, could contribute to improved FV risk assessment and practice.

Redevelop Risk Assessment Instruments to Perform Equitably Across Sub-groups of Cases

During the research presented in this thesis, we considered differences in the risk categories' performance for sub-groups of cases created according to aggressor characteristics (e.g., age, ethnicity, and gender) and aggressor-victim relationships. In manuscript one, both the DYRA and SAFVR instruments performed relatively consistently across sub-groups based on aggressor characteristics and aggressor-victim relationships; but both instruments' discriminative ability was poorer for non-intimate partner FV cases and the DYRA performed more poorly for aggressors who were women or Māori. In manuscript two,

we found the predictive validity of the ISR triage teams' risk assessments varied across sub-groups, performing poorly for aggressors who were Māori or women, and for non-intimate partner cases. Taken together, these findings indicate the instruments used to assess risk for FV cases in New Zealand should be redeveloped to ensure they adequately and equitably capture risk among different FV aggressors and cases (Woldgabreal et al., 2020; Zottola et al., 2021).

Implications

The findings of this research have several implications for risk assessment in the FV context. We will discuss those implications in the three following sections. The sections are organized thematically and relate to whether the purpose of risk assessment is prediction or case management, whether the goal of risk assessment is to predict criminal or sub-criminal behavior, and how static and dynamic risk factors can inform decisions about risk and case management in the FV context.

Risk Prediction versus Case Management

First and foremost, the results presented in this thesis challenge how we conceptualize the task of risk assessment. Based on definitions from research, the purpose of risk assessment is to systematically speculate about the likelihood of future criminal behavior (Bonta & Andrews, 2016; Hilton & Ennis, 2020). Accordingly, to evaluate an instruments' validity, researchers will examine statistics reflecting its discrimination (i.e., ability to separate people who reoffended and people who did not) and calibration (i.e., concordance between expected and actual rates of reoffending; Helmus & Babchishin, 2016; Singh, 2013). From this evaluation of an instrument's predictive validity, researchers then make conclusions about that instrument's value to practice (Kropp, 2008). In this view, if an instrument classifies a person as at high risk of reoffending, and that person does reoffend, the instrument was accurate, and is performing well (Singh, 2013).

However, when practitioners are planning to intervene to prevent further harm, it is a good outcome for a person who was classified as being at high risk to not reoffend (Kropp, 2008). That situation may not mean the instrument was inaccurate; it is possible the instrument adequately predicted risk that was then mitigated by intervention, and the risk assessment process may have even helped the practitioner plan that response. Many SPJ risk assessment instruments explicitly inform case management, such as the HCR-20 that draws practitioner attention to the areas of supervision, monitoring, treatment, and victim safety planning (Douglas et al., 2013, 2014). Indeed, in chapter 4 of this thesis, we found some of the factors ISR triage teams recorded had implications for how practitioners should manage those cases. These results support the idea that—in addition to predictive validity—the value of risk assessment instruments created for use in treatment or prevention settings should also be determined based on whether and how they support practitioners to plan their response (Kropp, 2008; Douglas et al., 2014).

The notion that risk assessments are important for informing case management has implications for how the RNR model for correctional treatment is used to inform risk assessment research and practice (Bonta & Andrews, 2016). The risk and need principles are widely considered to have the greatest influence on risk assessment research and practice, but the results presented herein suggest responsivity is also an important component of assessing risk (Eisenberg et al., 2022). Therefore, in addition to measuring factors related to criminal behavior, perhaps instruments developed for assessing risk at the front line and informing prevention actions (e.g., DYRA; Jolliffe Simpson et al., 2021) should additionally measure responsivity-related factors that can inform what actions agencies take to reach people and foster their engagement with services (Douglas et al., 2014).

Furthermore, in the correctional context risk assessments are most often conducted using factors measured at the level of the individual (e.g., the central eight; Bonta &

Andrews, 2016) that communicate the likelihood of that individual engaging in criminal behavior (Polaschek, 2012). But, in this thesis, we found FV practitioners frequently recorded stressors of conflict or barriers to responsivity for victims, or at the level of the relationship, or the entire family/whānau. In theory, the RNR model can apply to a wide range of interventions—including family-based interventions—but its principles focus on individual offenders (Bonta & Andrews, 2016; Polaschek, 2012, 2016). Therefore, researchers should investigate how the RNR model can be extended to accommodate factors that exist more widely than the individual and improve how the model can inform risk assessments performed for social units like families or whānau. In particular, the responsivity principle could be extended to include family-wide factors that act as barriers to responsivity—rather than individual characteristics of the aggressor—and the need principle could broaden to include external stressors that precipitate conflict and perpetuate cumulative harm (Bonta & Andrews, 2016; Polaschek, 2016).

On the other hand, the finding that ISR triage team members frequently recorded factors for victims, relationships, and family/whānau is in line with the ecological approach for understanding the causes of FV. The ecological model puts forward that phenomena like FV cannot be explained by one factor or a single theory; instead, FV occurs due to the interaction of people, stressors, and the environment (Whiting et al., 2020). In manuscript two and chapter four we found ISR triage teams recorded a heterogeneous range of factors, many of which were at the level of the family unit, rather than relating to individuals, which suggests that the ecological approach is embedded in FV practice.

Risk Prediction for Criminal versus Sub-Criminal Behavior

The results of this research have implications for the goal of risk assessment in the FV context; whether it is to predict offending, or whether assessments should also predict sub-criminal behavior. To reiterate, risk assessment is defined as systematic speculation about the

likelihood of future criminal behavior (Bonta & Andrews, 2016; Hilton & Ennis, 2020). However, in the context of risk assessment for FV, many of the behaviors worth predicting—such as verbal abuse, property damage, and threats—do not meet the threshold of criminal offenses. Yet, these forms of harm can be chronic, and negatively impact victims in a cumulative way (Ansara & Hindin, 2011; Dichter et al., 2018; Robinson et al., 2018; Wiener, 2017). FV episodes involving sub-criminal behaviors are also common; indeed, only a third of episodes reported to police in New Zealand involve physical harm (Jolliffe Simpson et al., 2021; New Zealand Police, 2022).

In this research, we examined risk for outcomes that included FV aggressors' sub-criminal behaviors, including recurrence (i.e., another FV-related call for police service) and the types of harm recorded in episodes, regardless of whether those episodes constituted criminal offenses. Resultantly, the findings reflected how well the assessments categorized risk for the outcomes that practitioners have in mind when planning their response (e.g., Will FV occur again? Will the aggressor physically harm their family members? Heckert & Gondolf, 2005; Henning et al., 2021; Kropp, 2008). This approach generated results that were relevant for FV practice, and we concluded that in addition to predicting offending, risk assessment instruments created to inform FV practice should also reflect the likelihood of ongoing harm. The predictive validity of the SAFVR, DYRA, and ISR risk categories for those outcomes could be improved if sub-criminal forms of harm are factored in as those assessments are redeveloped.

Internationally, risk assessment practices are geared towards predicting offending to reflect the requirements of the criminal justice system (e.g., informing parole decisions), and this focus is embedded in the RNR model. The meta-analytic research used to inform the RNR model was concerned with *what works* to reduce recidivism, rather than other things that are *nice to have*, such as wellbeing or agreeableness (Bonta & Andrews, 2016). Indeed,

the risk and need principles explain that risk factors are measurable characteristics that can be used to predict peoples' likelihood of engaging in criminal behavior (Bonta & Andrews, 2016; Polaschek, 2012). On the other hand, the Good Lives Model (GLM) is a strengths-based framework that states treatment targets should promote personal fulfillment, even if they do not have an empirical association with recidivism (Ward et al., 2012; Whitehead et al., 2007). These factors will differ from person to person according to their values and life goals but may include tasks such as obtaining a driver's license, attending university, and improving personal relationships (Whitehead et al., 2007). In summary, the GLM states that supporting wellbeing and personal fulfillment will lead to a reduction in risk, whereas the RNR model—which is supported by empirical evidence—states the opposite (Andrews et al., 2011).

If we accept that in the FV context the goal of risk assessment is to predict FV aggressors' criminal and sub-criminal behavior, then aspects of each of these models could inform that approach. The GLM is about achieving basic human goods, and that focus has been criticized because it may not lead to reductions in offending above the reductions achieved through adherence to the RNR principles (Andrews & Bonta, 2011). In fact, improving personal relationships is a key part of all RNR-informed rehabilitation (Bonta & Andrews, 2016; Polaschek, 2012). However, the GLM's factors that promote personal fulfillment could be used as a starting point in the identification of factors that precipitate conflict and harmful—but sub-criminal—interactions within families. In this research we found many of the factors ISR triage teams identified in their risk assessments related to situational stressors such as accommodation problems and child custody issues that would not traditionally be used to predict offending but would be addressed under a GLM-informed approach. Therefore, further work in this space should integrate the RNR model's focus on what works, but also consider what works to predict and reduce sub-criminal outcomes, and

account for human goods that my indirectly reduce conflict within families and whānau if promoted. Research may reveal that predictors of criminal and sub-criminal outcomes are the same, and promoting human goods does not reduce family conflict, but it is worth exploring these possibilities in the interest of producing research with value to FV practice. Of course, it would not be a seamless transition to integrate the GLM into this research, because like the RNR model, the GLM focuses on individuals; what does personal fulfillment mean in a family?

Static and Dynamic Risk Factors

Finally, the results of this research have implications for how static and dynamic risk factors can inform decisions about risk and case management in the FV context. In this research, we found New Zealand Police used the SAFVR, a static risk assessment, paired with the DYRA, a purportedly dynamic risk assessment, to inform their response; and the ISR's risk assessments combined both static and dynamic factors. Static risk assessment uses risk factors that do not change (e.g., age at first conviction), or change very slowly (e.g., age), that are algorithmically combined and compared with a benchmark from previous data to produce a risk rating. As a result, static risk assessment does not account for the factors currently contributing to a person's offending; therefore, static risk assessment instruments are useful for triage but cannot inform treatment other than by guiding the level of service required (Bonta & Andrews, 2016).

On the other hand, dynamic risk factors are conceptualized as the characteristics currently increasing a person's likelihood of engaging in criminal behavior (Douglas & Skeem, 2005). These factors can change, and do so either slowly (i.e., stable dynamic risk factors, e.g., peer group influences) or quickly (i.e., acute dynamic risk factors, e.g., current levels of substance use). Assessments using dynamic risk factors recognize that the circumstances precipitating FV change with time and context (Jones & Gondolf, 2001;

Kropp, 2008), and can be updated to reflect changes in risk (Davies et al., 2022). Because dynamic risk factors identify the characteristics currently increasing the likelihood of criminal behavior, they inform which characteristics should be targeted in treatment (Bonta & Andrews, 2016; Kropp, 2008). Further research about dynamic risk factors has the greatest potential to improve our predictions of criminal behavior above the benchmark set by static risk assessment (Douglas & Skeem, 2005; Kropp, 2008).

With that said, in the research presented in this thesis, we found the SAFVR—a static instrument—had the best predictive validity overall, and across sub-groups based on aggressor characteristics and aggressor-victim relationships; even for non-IPV cases, which were underserved by DYRA and ISR risk categories. The relatively consistent performance of the SAFVR instrument across sub-groups suggests that underlying criminal propensity explains a degree of involvement in FV, and assessment based on the aggressor's criminal history remains an adequate approach when triaging out cases that do not require further assessment and intervention services. But rather than indicating dynamic risk assessment is not worthwhile, the finding that the DYRA instrument preformed more poorly suggests dynamic risk assessment is currently a missed opportunity within risk assessment for FV in New Zealand. Indeed, some items in the DYRA instrument were static (e.g., "Has X ever …") and could not change once they occurred. There is a clear need to better understand and measure the dynamic risk factors for FV, and align the DYRA and ISR risk assessments with those factors to improve predictions above the benchmark set by the SAFVR and inform the characteristics FV practitioners target to achieve behavior change (Douglas & Skeem, 2005; Kropp, 2008).

Strengths and Limitations of the Current Research

Inclusive View of FV

One of the main strengths of the research presented in this thesis was that it used a broad definition of FV, based on New Zealand legislation (Family Violence Act, 2018). Numerous studies have called for more inclusive research on FV that recognizes the heterogeneity of FV cases and the types of situations to which police routinely respond (e.g., Chan et al., 2021; Dixon & Browne, 2003; Dixon et al., 2007; McEwan et al., 2018; Saxton et al., 2022). The existing body of research on FV is dominated by studies on IPV, meaning that harm committed in the context of other familial relationships is largely overlooked (e.g., elder abuse; Lawson, 2015, p.18). As a result, collective understandings of the mechanisms of FV, and the validity of risk assessments are largely based on the drivers of IPV. Indeed, while conducting the research presented in this thesis, we noticed that some of the DYRA items and ISR triage teams' factors were only relevant for cases where the aggressor and victim were current or former intimate partners, and we consistently found those instruments performed more poorly for non-IPV cases. There is a need for research that uncovers whether non-IPV cases of FV should be assessed using procedures created for IPV cases, or whether a need exists for further calibration or specialized instruments. It is possible that many drivers are the same for other types of FV (especially where they overlap with empirically established predictors of criminal behavior in general; Bonta & Andrews, 2016); however, to explore that possibility we need to conduct research that compares the risk factors for—and mechanisms of—FV among people in different types of relationships.

Use of Administrative Field Data

In the same vein, an additional strength of this research was that it used a community-based sample of FV cases with episodes reported to police. As a result, the sample represented FV that agencies deal with on a routine basis more accurately than much of the

existing research on FV, which focuses on samples of people with more severe behavior (e.g., men referred to non-violence treatment programs; Heckert & Gondolf, 2005; Jones et al., 2010). In doing so, the thesis complemented existing research by describing the less-severe end of FV cases, and generated findings that were more relevant for practitioners on the front line of New Zealand's response to FV. On the other hand, the decision to base this research on police reports and case notes for FV episodes was also a limitation. Despite their efforts to train staff and conduct quality assurance, police and the ISR recorded information is ultimately not generated for research, and we could not quantify the data's reliability. Therefore, this information was likely less reliable than information researchers can generate when, for example, interviewing aggressors and victims themselves. Moreover, by focusing on information reported police we based our analyses on information that was reported and missed the significant proportion of FV episodes that do not come to police attention (which could be as large as 80%; New Zealand Crime and Victims Survey, 2018).

Using police reports and case notes for FV episodes was also a strength because it allowed us to examine the validity of field ratings, factors, and risk categories using data generated during practice, rather than for research (Singh et al., 2013). Hence, the data and findings reflected the judgements practitioners make, rather than the judgements researchers make, which increases our confidence that the results are relevant to practice. But, using the risk categories meant we evaluated the assessments' predictive validity using a scale with decreased variance (e.g., low, moderate, high = three items) compared with research that analysed other risk assessments' predictive validity using scales with greater variance (e.g., raw scores ranging from 0-15; van der Put et al., 2019). This decreased variance may account for the poorer AUCs we observed for these instruments (Helmus & Babchishin, 2016; Zottola et al., 2021).

Moreover, focusing on police reports and case notes for FV episodes was a limitation because it meant we did not include other sources of information that could have enriched the analyses. For example, adding information about aggressors' engagement in other types of offending would have provided a more complete picture of their antisocial behavior patterns. Research suggests some FV aggressors also commit violence outside the home, or engage in general criminal offending, whereas other FV aggressors are exclusively violent within the home (Hilton & Eke, 2016; Holtzworth-Munroe & Stuart, 1994; Piquero et al., 2006). Including information about other offending may have enabled us to examine the extent to which engagement in other types of criminal behavior mapped onto the sub-groups we identified and explore how police and the ISR's risk assessment sand case management plans differed according to aggressors' degree of criminality.

Another source of information that would have enhanced this research is the type and level of interventions that families were offered and chose to engage with, because those interventions could have decoupled families' initial risk categorizations from the outcomes we measured (Dowling & Morgan, 2019). All FV cases in this research were referred to the ISR for further triaging and case management, which included the allocation of intervention services (Integrated Safety Response, 2019). In chapter four, we used the factors that ISR triage teams recorded to identify a sub-group of cases with high needs and barriers to responsivity, and the fact we could identify this group suggests the ISR was attuned to those families' needs. If the intervention services ISR triage teams allocated were effective, those interventions could have decoupled the risk present among cases from the initial risk assessment and because we could not measure interventions, we could not rule out that possibility. Hence, further examination of the interventions that families and whānau receive could establish whether, and how, interventions moderate the predictive validity of the risk factors (Dowling & Morgan, 2019).

Focus on, and Selection of, Aggressors

Another potential limitation of this research was the initial selection of a predominant aggressor. Selecting a predominant aggressor is fraught; when officers attend FV episodes, they do not always know about the FV history of the parties involved. In unclear situations, police select the predominant aggressor based on the parties' physical capability to cause harm, contributing to a gender-bias towards the male partner in heterosexual relationships (Hamel, 2011) and potentially in other situations, too (e.g., same-sex, sibling, and parent-child relationships). Indeed, in this study aggressors were chosen based on the person police listed as 'perpetrator' or 'suspect' in the report for the index episode, and in cases where when police assigned two parties the role of 'mutual participant', we chose the person that police indicated was 'posing risk' in the DYRA. Therefore, it is possible these biases contributed to the incorrect identification of the predominant aggressor in this research, meaning we missed people who were aggressors but were not identified (Hamel, 2011).

In addition, by focusing on the demographic and risk characteristics of aggressors from the index episodes and recording the harm those aggressors committed over two years, we did not capture those aggressors' experiences of victimization, nor the harm committed by other members of their family or whānau. Many people who commit FV also experience victimization during their lives, with some research indicating exposure to FV or FV victimization during childhood is a risk factor for committing FV in later life (Mallory et al., 2016; Ryan & Roman 2021; Stith et al., 2004). In addition, some FV cases are bi-directional; episodes can involve multiple people committing harmful acts, and people can switch between the roles of victim and aggressor over time (Straus, 2015; Straus & Gozjolko, 2016). Because we chose to focus on the aggressor from the sample of index episodes, we could not account for aggressors' experiences of victimization, cases of role switching between episodes, or bi-directional harm within episodes (Straus, 2015; Straus & Gozjolko, 2016).

Suggestions for Future Research

Improve Responsiveness to Māori

Risk assessment instruments have been criticized for reinforcing biases in the criminal justice system; particularly those against Indigenous or minority communities (Woldgabreal et al., 2020; Zottola et al., 2021). Many risk assessment instruments are based on the central eight risk factors (Bonta & Andrews, 2016), which are not necessarily biased on their own but may systematically differ in their rates between groups of people. For example, Indigenous and minority groups may score higher for criminal history due to bias or over-policing, or for antisocial attitudes because they mistrust the criminal justice system as a result (Mears et al., 2016; Tompson et al., 2021; Woldgabreal et al., 2020). In this research, we found New Zealand Police's DYRA risk assessment instrument and the ISR's risk assessments performed more poorly for aggressors who were Māori, despite the ISR's commitment to responsiveness to Māori (Mossman et al., 2017) and New Zealand Police's goal of achieving better outcomes for all New Zealanders by working in partnership with Māori (New Zealand Police, 2020). Māori experience disproportionately high rates of incarceration in New Zealand—around 15% of people in New Zealand identify as Māori, compared with more than half of people in New Zealand's prisons—and are also disproportionately represented in higher-risk prison security classifications (McIntosh, & Workman, 2017). Therefore, risk assessment instruments that perform poorly for Māori may continue and exacerbate the dirproportionaly of Māori in the criminal justice system (Woldgabreal et al., 2020). Clearly, the DYRA and ISR risk assessments should be redeveloped to ensure they equitably and accurately capture risk for aggressors who are Māori (Singh et al., 2013). Moreover, the experiences and values of Māori people should be centered in how the assessments are scored and interpreted (Woldgabreal et al., 2020).

In this research there was some indication that the ISR were attuned to the needs present among whānau experiencing FV, and many of the agencies providing FV services were kaupapa Māori. However, further research should examine when, why, and how the ISR assigns interventions (including kaupapa Māori interventions). Researchers should envisage a theory of change for how those interventions work to reduce FV, and evaluate the extent to which the interventions achieve that aim, in order to systematically identify how agencies can improve responsiveness to Māori and outcomes for all New Zealanders. There is a nation-wide movement to improve responses to Māori in the criminal justice system in New Zealand. Ara Poutama (Department of Corrections) recently released Hōkai Rangi, a new five-year strategy, that seeks to eliminate the overrepresentation of Māori in the criminal justice system and promote wellbeing (Ara Poutama Aotearoa, 2019). Furthermore, in 2020 the government released Te Aorerekura, a new national strategy of successive steps to reduce FV and sexual violence, that is underpinned by tikanga Māori principles (New Zealand Government 2022). Further research will continue to be necessary to understand how the enactment of Te Aorerekura (New Zealand Government 2022) and Hōkai Rangi (Ara Poutama Aotearoa, 2019) will affect responses to FV, including how risk is assessed, cases are managed, and interventions are delivered.

Operationalize Risk for Latent States or Pathways

The categorical labels used to communicate the results from risk assessments (e.g., low, medium, or high) are problematic for many reasons, including that people interpret them differently and that they are applied non-uniformly across instruments (Hanson et al., 2017). Resultantly, high risk could refer to a 30% or 90% likelihood of a person being reconvicted depending on the instrument in question. Moreover, the single dichotomous outcomes for which these labels express risk do not align with the outcomes that practitioners responding

to FV have in mind when planning their response (Heckert & Gondolf, 2005); practitioners may be more interested in the likelihood of ongoing harm than of offending per se.

Indeed, the results from this research suggest there may be merit in conceptualizing FV-related outcomes as latent states or pathways, rather than as single dichotomous outcomes (Heckert & Gondolf, 2005). For example, existing instruments could be validated to concretely express the probability of an aggressor moving to certain latent states or following certain pathways (e.g., desistance, high-harm, or low harm; Jones et al., 2010). For example, an aggressor may have a .7 probability of moving to a given state that contains a .9 probability of verbal abuse and a .4 probability of physical violence. This way of communicating risk would be more uniformly interpreted than risk categories and align with the outcomes practitioners have in mind (Hanson et al., 2017; Heckert & Gondolf, 2005). Further research is necessary to operationalize risk for latent states or pathways, and to validate instruments to predict those outcomes before this approach could be applied in practice.

Modeling Pathways

Future research is necessary to expand the body of knowledge about FV aggressors' behavioral patterns over time, to inform how risk for FV is assessed. A crucial step in that research is to develop a harm index of behaviors that constitute FV, including behaviors that are sub-criminal (Bland & Ariel, 2015). The Revised Conflict Tactics Scale is widely used to measure behaviors within the context of IPV but requires an interview with victims and cannot be completed from the information available in police reports (Straus & Douglas, 2017). Previous research has applied the Cambridge Crime Harm Index (e.g., Barnham et al., 2017; Bland & Ariel, 2015), but this index gives FV episodes without criminal offenses no weight, or a weighting of 0.1, and does not describe the different types of behaviors within episodes. Therefore, in this research we sought to understand the probability of different

behaviors by using five dichotomous indicators of harm to estimate aggressors' latent states behind the information reported to police (Jones et al., 2010). This approach did not weight the different types of harm to reflect severity, and an index that weights the frequency and severity of different types of sub-criminal forms of harm would improve the latent states' reflection of aggressor behavior and family or whānau experiences of harm. Such an index could be developed from a questionnaire to ask people within New Zealand—including people who have experienced FV—about their opinions of the severity of different types of behaviors, and behaviors in combination (Castro et al., 2006).

This research should also incorporate measures of psychological violence—including coercive control—which is a common and harmful component of FV (Ansara & Hindin, 2011; Dichter et al., 2018; Fanslow, Malihi, et al., 2021; Wiener, 2017). The behaviors that constitute coercive control are difficult for police officers to detect, and can often be hidden (Robinson et al., 2018). Indeed, the reports used in this study did not systematically record the presence of coercive control, and further research should include coercive control when operationalizing FV outcomes.

Finally, further research about FV aggressors' behavioral patterns over time should endeavor to identify and describe the small group of cases that account for a disproportionately large amount of FV-related harm (Adams, 2016). The risk categories examined in this research can be described as triage instruments that were useful for screening out low-risk cases but not for identifying cases at the highest risk of ongoing harm. Further research should seek to better understand these aggressors and families or whānau, by collecting larger or higher-risk samples. As a result of this work, instruments could be designed to identify these very high-risk cases after initial triage (e.g., PPIT; Robinson & Clancy, 2020) or alternatively, existing instruments could be modified to use the five-level

framework (I, II, III, IV, V), which could lead to the identification of very high risk cases as a result (Hanson et al., 2017).

Conclusion

Despite considerable advancements in risk assessment research and practice over recent decades, risk assessment instruments have limitations that curb their value to FV practice, and there is limited evidence for the validity of risk assessment instruments used by agencies that respond to FV in Aotearoa New Zealand. Taken together, the results of the research presented in this thesis show that the risk assessment instruments used in New Zealand can predict risk for FV recurrence better than chance, but those assessments should be further developed to improve their value to FV practice. Importantly, further work is needed for those instruments to equitably reflect risk for different groups of people. In addition, to understand the value of risk assessment instruments for practice, researchers should examine both the instruments' predictive validity and their utility for case management.

This thesis highlighted that the FV risk assessment instruments are limited by their reliance on single dichotomous outcomes and use of categorical labels to communicate risk. We suggest that instruments could instead be designed to communicate the probability of aggressors moving to defined latent states that incorporate multiple outcomes. That change could contribute to improved FV risk assessment and practice because it would be uniformly interpreted and would also more closely reflect both the forms of harm that families experience and the outcomes that practitioners have in mind when assessing risk. Overall, this thesis advanced our understanding of the limitations of—and opportunities in—risk assessment research and contributed to the evidence base for risk assessment for FV in New Zealand.

References for Chapters One, Four and Six

Adams, A. (2016). *Social Investment in the Criminal Justice System*. https://www.beehive.govt.nz/speech/social-investment-criminal-justice-system

Andrews, D. A., Bonta, J. L., & Wormith, J. S. (2004). *Level of Service/Case Management Inventory (LS\CMI): An offender assessment system.* Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2011). The Risk-Need-Responsivity (RNR) Model: Does adding the Good Lives Model contribute to effective crime prevention? *Criminal Justice and Behavior*, *38*(7), 735-755. https://doi.org/10.1177/0093854811406356

Ansara, D. L., & Hindin, M. J. (2011). Psychosocial consequences of intimate partner violence for women and men in Canada. *Journal of Interpersonal Violence*, *26*(8), 1628-1645. https://doi.org/10.1177/0886260510370600

Asparouhov, T. & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus, *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 329-341. https://doi.org/10.1080/10705511.2014.915181

Ara Poutama Aotearoa (2019). *Hōkai Rangi: Ara Poutama Aotearoa Strategy 2019-2024.* https://www.corrections.govt.nz/__data/assets/pdf_file/0003/38244/Hokai_Rangi_Strategy.pdf

Arias-Rivera, S., Hidalgo, V., & Lorence, B. (2020). A scoping study on measures of child-to-parent violence. *Aggression and Violent Behavior*, *52*, 1-11. https://doi.org/10.1016/j.avb.2020.101426

Barnham, L., Barnes, G. C., & Sherman, L. W. (2017). Targeting escalation of intimate partner violence: Evidence from 52,000 offenders. *Cambridge Journal of Evidence-Based Policing*, *1*(2-3), 116-142. https://doi.org/10.1007/s41887-017-0008-9

Belsky, J. (1980). Child maltreatment: An ecological integration. *American Psychologist, 35*(4), 320-335. https://doi.org/10.1016/j.cpr.2015.01.002

Bissielo, A., & Knight, G. (2016). *Family Violence Risk Assessment Redevelopment: Static Risk Score*. (New Zealand Police internal report). New Zealand Police.

Bland, M. (2020). Algorithms can predict domestic abuse, but should we let them? In Jahankhani, H., Akhgar, B., Cochrane, P., Dastbaz, M. (Eds), *Policing in the Era of AI and Smart Societies: Advanced Sciences and Technologies for Security Applications* (pp. 139-155). Springer. https://doi.org/10.1007/978-3-030-50613-1_6

Bland, M., & Ariel, B. (2015). Targeting escalation in reported domestic abuse: Evidence from 36,000 callouts. *International Criminal Justice Review*, *25*(1), 30-53. https://doi.org/10.1177/1057567715574382

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (2017). *Manual for version 2 of the sexual violence risk—20: Structured professional judgment guidelines for assessing and managing risk of sexual violence.* Protect International Risk and Safety Services Inc.

Bonta, J., & Andrews, D. A. (2016). *The Psychology of Criminal Conduct* (6th ed.). Routledge. https://doi.org/10.4324/9781315677187

Boxall, H., Morgan, A., & Brown, R. (2020). The prevalence of domestic violence among women during the COVID-19 pandemic. *Australasian Policing*, *12*(3), 38-46. https://doi.org/10.52922/sb04718

Bronfenbrenner, U. (1981). *The ecology of human development: Experiments by nature and design*. Harvard University Press. https://doi.org/10.2307/j.ctv26071r6

Carlson, B. E. (1984). Causes and maintenance of domestic violence: An ecological analysis. *Social Service Review*, *58*(4), 569-587. https://doi.org/10.1086/644239

Castro, R., García, L., Ruíz, A., & Peek-Asa, C. (2006). Developing an index to measure violence against women for comparative studies between Mexico and the United States. *Journal of Family Violence*, *21*(1), 95-104. https://doi.org/10.1007/s10896-005-9005-z

Cavanaugh, C. E., Messing, J. T., Petras, H., Fowler, B., Flair, L. L., Kub, J., Agnew, J., Fitzgerald, S., Bolyard, R., & Campbell, J. C. (2012). Patterns of violence against women: A latent class analysis. *Psychological Trauma: Theory, Research, Practice, and Policy*, *4*(2), 169-176. https://doi.org/10.1037/a0023314

Chan, K. L., Chen, Q., & Chen, M. (2021). Prevalence and correlates of the co-occurrence of family violence: A meta-analysis on family polyvictimization. *Trauma, Violence and Abuse*, *22*(2), 289-305. https://doi.org/10.1177/1524838019841601

Cohen, T. H., Lowenkamp, C. T., Bechtel, K., & Flores, A. W. (2020). Risk assessment overrides: Shuffling the risk deck without any improvements in prediction. *Criminal Justice and Behavior*, *47*(12), 1609-1629. https://doi.org/10.1177/0093854820953449

Contini, M., & Wilson, B. (2019). *Canadian multi-agency risk assessment committee (MARAC) model program*. University of Guelph. https://www.cesinstitute.ca/canadian-multi-agency-risk-assessment-committee-marac-model-program

Cooper, C., Selwood, A., & Livingston, G. (2008). The prevalence of elder abuse and neglect: A systematic review. *Age and Ageing*, *37*(5), 151-160. https://doi.org/10.1093/ageing/afm194

Davies, S. T., Lloyd, C. D., & Polaschek, D. L. L. (2022). Does reassessment enhance the prediction of imminent criminal recidivism? Replicating Lloyd et al. (2020) with high-risk parolees. *Assessment*, *29*(5), 962-980. https://doi.org/10.1177/1073191121993216

De Bortoli, L., Ogloff, J., Coles, J., & Dolan, M. (2017). Towards best practice: combining evidence-based research, structured assessment and professional judgement. *Child and Family Social Work*, *22*(2), 660-669. https://doi.org/10.1111/cfs.12280

Desmarais, S. L., Reeves, K. A., Nicholls, T. L., Telford, R. P., & Fiebert, M. S. (2012). Prevalence of physical violence in intimate relationships, part 2: Rates of male and female perpetration. *Partner Abuse*, *3*(2), 170-198. https://doi.org/10.1891/1946-6560.3.2.170

Devries, K. M., Mak, J. Y. T., García-Moreno, C., Petzold, M., Child, J. C., Falder, G., Lim, S., Bacchus, L. J., Engell, R. E., Rosenfeld, L., Pallitto, C., Vos, T., Abrahams, N., & Watts, C. H. (2013). The global prevalence of intimate partner violence against women. *Science*, *340*(6140), 1527-1528. https://doi.org/10.1126/science.1240937

Dichter, M. E., Thomas, K. A., Crits-Christoph, P., Ogden, S. N., & Rhodes, K. V. (2018). Coercive control in intimate partner violence: Relationship with women's experience of violence, use of violence, and danger. *Psychology of Violence*, *8*(5), 596-604. https://doi.org/10.1037/vio0000158

Dixon, L., & Browne, K. (2003). The heterogeneity of spouse abuse: A review. *Aggression and Violent Behavior*, *8*(1), 107-130. https://doi.org/10.1016/S1359-1789(02)00104-0

Dixon, L., Hamilton-Giachritsis, C., Browne, K., & Ostapuik, E. (2007). The co-occurrence of child and intimate partner maltreatment in the family: Characteristics of the violent perpetrators. *Journal of Family Violence*, *22*(8), 675-689. https://doi.org/10.1007/s10896-007-9115-x

Dobbs, T., & Eruera, M. (2014). *Kaupapa Māori wellbeing framework: The basis for whānau violence prevention and intervention*. New Zealand Family Violence Clearinghouse. https://nzfvc.org.nz/issues-papers-6

Domestic Protection Act (1982). http://www.nzlii.org/nz/legis/hist_act/dpa19821982n120253

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20 Assessing Risk for Violence: User Guide*. Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., Guy, L. S., & Wilson, C. M. (2014). Historical-Clinical-Risk Management-20, Version 3 (HCR-20V3): Development and overview. *International Journal of Forensic Mental Health*, *13*(2), 93-108. https://doi.org/10.1080/14999013.2014.906519

Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology Public Policy and Law*, *11*(3), 347-383. https://doi.org/10.1037/1076-8971.11.3.347

Dowling, C., & Morgan, A. (2019). *Predicting repeat domestic violence: Improving police risk assessment*. Australian Institure of Criminology. https://www.aic.gov.au/publications/tandi/tandi581

Dutton, D. (1995). *The domestic assault of women: Psychological and criminal justice perspectives*. University of British Columbia Press.

Dutton, D. (2006). *Rethinking Domestic Violence*. University of British Columbia Press.

Eisenberg, M. J., van Horn, J. E., van der Put, C. E., Stams, G. J. J. M., & Hendriks, J. (2022). Protective factors as uni- or bipolar factors and their incremental validity and accuracy in predicting general recidivism. *International Journal of Law and Psychiatry*, *81*, 101772. https://doi.org/10.1016/j.ijlp.2021.101772

Every-Palmer, S., Jenkins, M., Gendall, P., Hoek, J., Beaglehole, B., Bell, C., Williman, J., Rapsey, C., & Stanley, J. (2020). Psychological distress, anxiety, family violence, suicidality, and wellbeing in New Zealand during the COVID-19 lockdown: A cross-sectional study. *PLOS ONE*, *15*(11), e0241658. https://doi.org/10.1371/journal.pone.0241658

Family Violence Act (2018). http://www.legislation.govt.nz/act/public/2018/0046/latest/LMS112966.html

Fanslow, J., Hashemi, L., Malihi, Z., Gulliver, P., & McIntosh, T. (2021). Change in prevalence rates of physical and sexual intimate partner violence against women: data from two cross- sectional studies in New Zealand, 2003 and 2019. *BMJ Open*, *11*, e044907. https://doi.org/10.1136/bmjopen-2020-044907

Fanslow, J., Malihi, Z., Hashemi, L., Gulliver, P., & McIntosh, T. (2021). Change in prevalence of psychological and economic abuse, and controlling behaviours against women by an intimate partner in two cross-sectional studies in New Zealand, 2003 and 2019. *BMJ Open*, *11*, e044910. https://doi.org/10.1136/bmjopen-2020-044910

Foran, H. M., & O'Leary, K. D. (2008). Alcohol and intimate partner violence: A meta-analytic review. *Clinical Psychology Review*, *28*(7), 1222-1234. https://doi.org/10.1016/j.cpr.2008.05.001

Garrington, C., & Boer, D. P. (2020). Structured professional judgement in violence risk assessment. In J. S. Wormith, L. A. Craig, & T. E. Hogue (Eds.), *The Wiley Handbook of What Works in Violence Risk Management* (1st ed., pp. 145-162). John Wiley & Sons Ltd. https://doi.org/10.1002/9781119315933.ch7

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. https://doi.org/10.1037/1076-8971.2.2.293

Hague, G., & Bridge, S. (2008). Inching forward on domestic violence: the 'co- ordinated community response' and putting it in practice in Cheshire. *Journal of Gender Studies*, *17*(3), 185-199. https://doi.org/10.1080/09589230802204134

Hamel, J. (2011). In dubious battle: The politics of mandatory arrest and dominant aggressor laws. *Partner Abuse*, *2*(2), 224-245. https://doi.org/10.1891/1946-6560.2.2.224

Hanson, R. K., Bourgon, G., McGrath, R. J., Kroner, D., D'Amora, D. A., Thomas, S. S., & Tavarez, L. P. (2017). *A five-level risk and needs system: Maximizing assessment results in corrections through the development of a common language*. National Reentry Resource Center. https://saratso.org/pdf/A_Five_Level_Risk_and_Needs_System_Report.pdf

Hanson, R. K., Helmus, L., & Bourgon, G. (2007). *The validity of risk assessments for intimate partner violence: A meta-analysis*. Public Safety Canada http://www.publicsafety.gc.ca/res/cor/rep/_fl/vra_ipv_200707_e.pdf

Hart, S. D., Douglas, K. S., & Guy, L. S. (2016). The structured professional judgement approach to violence risk assessment: Origins, nature, and advances. In D. P. Boer, A. R. Beech, T. Ward, L. A. Craig, M. Rettenberger, L. E. Marshall, & W. L. Marshall (Eds.), *The Wiley Handbook on the Theories, Assessment, and Treatment of Sexual Offending* (pp. 643–666). Wiley Blackwell. https://doi.org/10.1002/9781118574003.wattso030

Heckert, D. A., & Gondolf, E. W. (2005). Do multiple outcomes and conditional factors improve prediction of batterer reassault? *Violence and Victims*, *20*(1), 3-24. https://doi.org/10.1891/vivi.2005.20.1.3

Heister, J. (2018). *SAFVR Reanalysis of Morris and Mossman (2015)*. (New Zealand Police internal report). New Zealand Police.

Helmus, L. M., & Babchishin, K. M. (2016). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior*, *44*(1), 8–25. https://doi.org/10.1177/0093854816678898

Henning, K., Campbell, C. M., Stewart, G., & Johnson, J. (2021). Prioritizing police investigations of intimate partner violence using actuarial risk assessment. *Journal of Police and Criminal Psychology, 36*(4), 667-678. https://doi.org/10.1007/s11896-021-09466-7

Hilton, N. Z., & Eke, A. W. (2016). Non-specialization of criminal careers among intimate partner violence offenders. *Criminal Justice and Behavior, 43*(10), 1347–1363. https://doi.org/10.1177/0093854816637886

Hilton, N. Z., & Eke, A. W. (2017). Assessing risk of intimate partner violence. In J. C. Campbell & J. T. Messing (Eds.), *Assessing Dangerousness: Domestic Violence Offenders and Child Abusers*. Springer. https://doi.org/10.1891/9780826133274.0006

Hilton, N. Z., & Ennis, L. (2020). Intimate partner violence risk assessment and management: An RNR approach to threat assessment. In J. S. Wormith, L. A. Craig, & T. E. Hogue (Eds.), *The Wiley Handbook of What Works in Violence Risk Management* (1st ed.). John Wiley & Sons Ltd. https://doi.org/10.1002/9781119315933.ch8

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist, 34*(3), 400-409. https://doi.org/10.1177/0011000005285877

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2010). *Risk Assessment for Domestically Violent Men: Tools for Criminal Justice, Offender Intervention, and Victim Services*. American Psychological Association. https://doi.org/10.1037/12066-000

Holtzworth-Munroe, A., & Stuart, G. L. (1994). Typologies of male batterers: Three subtypes and the differences among them. *Psychological Bulletin, 116*(3), 476–497. https://doi.org/10.1037/0033-2909.116.3.476

Holtzworth-Munroe, A. (2005). Male versus female intimate partner violence: Putting controversial findings into context. *Journal of Marriage and Family*, *67*(5), 1120-1125. https://doi.org/10.1111/j.1741-3737.2005.00203.x

Hoyle, C. (2008). Will she be safe? A critical analysis of risk assessment in domestic violence cases. *Children and Youth Services Review*, *30*(3), 323-337. https://doi.org/10.1016/j.childyouth.2007.10.009

Integrated Safety Response. (2018). *Risk Assessment Quick Guide*. Author.

Integrated Safety Response. (2019). *12 week review of cases referred to the Integrated Safety Response (ISR) pilot*. https://www.police.govt.nz/sites/default/files/publications/isr-12-week-case-review-report.pdf

Jolliffe Simpson, A. D., Joshi, C., & Polaschek, D. L. L. (2021). Predictive validity of the DYRA and SAFVR: New Zealand Police's family violence risk assessment instruments. *Criminal Justice and Behavior*, *48*(10), 1487-1508. https://doi.org/10.1177/0093854821997525

Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2023). Unpacking multiagency structured professional judgement risk assessment for family violence. *Journal of Interpersonal Violence.* Advance online publication. https://doi.org/10.1177/08862605221147069

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. https://doi.org/10.1023/A:1012268725273

Jones, A. S., Heckert, D. A., Gondolf, E. D., Zhang, Q., & Ip, E. H. (2010). Complex behavioral patterns and trajectories of domestic violence offenders. *Violence and Victims*, *25*(1), 3-17. https://doi.org/10.1891/0886-6708.25.1.3

Jones, L., Bellis, M. A., Wood, S., Hughes, K., McCoy, E., Eckley, L., Bates, G., Mikton, C., Shakespeare, T., & Officer, A. (2012). Prevalence and risk of violence against children with disabilities: A systematic review and meta-analysis of observational studies. *Lancet*, *380*(9845), 899-907. https://doi.org/10.1016/S0140-6736(12)60692-8

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515-526. https://doi.org/10.1037/a0016755

Krishnan, V., & Morrison, K. B. (1995). An ecological model of child maltreatment in a Canadian province. *Child Abuse & Neglect*, *19*(1), 101-113. https://doi.org/10.1016/0145-2134(94)00101-Y

Kropp, P. R. (2004). Some questions regarding spousal assault risk assessment. *Violence Against Women*, *10*(6), 676-697. https://doi.org/10.1177/1077801204265019

Kropp, P. R. (2008). Intimate partner violence risk assessment and management. *Violence and Victims*, *23*(2), 202-220. https://doi.org/10.1891/0886-6708.23.2.202

Kropp, P. R., Hart, S. D., & Belfrage, H. (2005). *Brief spousal assault form for the evaluation of risk (B-SAFER).* User manual. ProActive ReSolutions.

Lanza, S. T., & Cooper, B. R. (2016). Latent class analysis for developmental research. *Child Development Perspectives*, *10*(1), 59-64. https://doi.org/10.1111/cdep.12163

Linzer, D. A. & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, *42*(10), 1-29. http://www.jstatsoft.org/v42/i10/.

Little, L., & Kantor, G. K. (2002). Using ecological theory to understand intimate partner violence and child maltreatment. *Journal of Community Health Nursing*, *19*(3), 133-145. https://doi.org/10.1207/s15327655jchn1903_02

Mallory, A. B., Dharnidharka, P., Deitz, S. L., Barros-Gomes, P., Cafferky, B., Stith, S. M., & Van, K. (2016). A meta-analysis of cross cultural risk markers for intimate partner

violence. *Aggression and Violent Behavior*, *31*, 116-126. https://doi.org/10.1016/j.avb.2016.08.004

McEwan, T. E., Shea, D. E., & Ogloff, J. R. P. (2018). An acturial instrument for police triage of Australian family violence reports. *Criminal Justice and Behavior, 46*(4), 590–607. https://doi.org/10.1177/0093854818806031

McIntosh, T., & Workman, K. (2017). Māori and Prison. In A. Deckert, & R. Sarre (Eds.) *The Palgrave Handbook of Australian and New Zealand Criminology, Crime and Justice.* Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-55747-2_48

Mears, D. P., Cochran, J. C., & Lindsey, A. M. (2016). Offending and racial and ethnic disparities in criminal justice: A conceptual framework for guiding theory and research and informing policy. *Journal of Contemporary Criminal Justice, 32*(1), 78–103. https://doi.org/10.1177/1043986215607252

Messing, J. T., Campbell, J. C., & Snider, C. (2017). Validation and adaptation of the danger assessment-5: A brief intimate partner violence risk assessment. *Journal of Advanced Nursing*, *73*(12), 3220-3230. https://doi.org/10.1111/jan.13459

Messing, J. T., & Thaller, J. (2013). The average predictive validity of intimate partner violence risk assessment instruments. *Journal of Interpersonal Violence*, *28*(7), 1537-1558. https://doi.org/10.1177/0886260512468250

Moody, G., Cannings-John, R., Hood, K., Kemp, A., & Robling, M. (2018). Establishing the international prevalence of self-reported child maltreatment: a systematic review by maltreatment type and gender. *BMC Public Health*, *18*(1), 1-15. https://doi.org/10.1186/s12889-018-6044-y

Mossman, E. (2014). *Evaluation of the use of ODARA in New Zealand*. (New Zealand Police internal report). New Zealand Police.

Mossman, E., Paulin, J., & Wehipeihana, N. (2017). *Evaluation of the family violence Integrated Safety Response pilot*. SUPERU. https://thehub.sia.govt.nz/assets/documents/ISR_pilot_evaluation_FINAL.pdf

Moulds, L., Day, A., Mayshak, R., Mildred, H., & Miller, P. (2019). Adolescent violence towards parents: Prevalence and characteristics using Australian police data. *Australian & New Zealand Journal of Criminology*, *52*(2), 231-249. https://doi.org/10.1177/0004865818781206

Nagin, D. S. (2005). Group-based modeling of development. *Harvard University Press.* https://doi.org/10.4159/9780674041318

New Zealand Crime and Victims Survey (2018). *Topical report: Offences against New Zealand adults by family members*. Ministry of Justice. https://www.justice.govt.nz/assets/Documents/Publications/9ZU3Q-NZCVS-topical-report-Offences-by-family-members-Cycle-1-2018.pdf

New Zealand Family Violence Clearinghouse (2017). *Data Summaries 2017: Snapshot*. https://nzfvc.org.nz

New Zealand Family Violence Death Review Committee (2021a). *Child abuse and neglect deaths in Aotearoa New Zealand: Fact sheets*. https://www.hqsc.govt.nz/assets/FVDRC/Publications/FVDRC_2021_CAN_English_web2.pdf

New Zealand Family Violence Death Review Committee (2021b). *Intimate partner violence deaths in Aotearoa New Zealand: Fact sheets*. https://www.hqsc.govt.nz/assets/FVDRC/Publications/FVDRC_2021_IPV_English_web.pdf

New Zealand Government (2022). *Te Aorerekura The Enduring Spirit of Affection: National Strategy to Eliminate Family Violence and Sexual Violence.*

https://violencefree.govt.nz/assets/National-strategy/Finals-translations-alt-formats/Te-Aorerekura-National-Strategy-final.pdf

New Zealand Police (2020). *Our Business.* https://www.police.govt.nz/about-us/publication/our-business

New Zealand Police (2022). *Daily Occurrences of Crime and Family Violence Investigations*. https://www.police.govt.nz/about-us/statistics-and-publications/data-and-statistics/daily-occurrences-crime

New Zealand Police (n.d.-a). *Family harm new risk measures handout.* https://www.police.govt.nz/sites/default/files/publications/family-harm-new-risk-measures-handout-web.pdf

New Zealand Police (n.d.-b). *Police Safety Orders*. https://www.police.govt.nz/advice-services/family-violence/police-safety-orders

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science, 4*(4), 440–461. https://doi.org/10.1037/tps0000176

Piquero, A. R., Brame, R., Fagan, J., & Moffitt, T. E. (2006). Assessing the offending activity of criminal domestic violence suspects: Offense specialization, escalation, and de-escalation evidence from the spouse assault replication program. *Public Health Reports*, *121*(4), 409-418. https://doi.org/10.1177/003335490612100409

Polaschek, D. L. L. (2012). An appraisal of the risk-need-responsivity (RNR) model of offender rehabilitation and its application in correctional treatment. *Legal and*

Criminological Psychology, *17*(1), 1–17. https://doi.org/10.1111/j.2044-8333.2011.02038.x

Polaschek, D. L. L. (2016). *Responding to perpetrators of family violence*. New Zealand Family Violence Clearinghouse. https://nzfvc.org.nz/sites/nzfvc.org.nz/files/NZFVC-issues-paper-11-responding-perpetrators.pdf

Prins, S. J., & Reich, A. (2018). Can we avoid reductionism in risk reduction? *Theoretical Criminology*, *22*(2), 258-278. https://doi.org/10.1177/1362480617707948

R Core Team. (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. http://www.R-project.org/

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior*, *29*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7

Richards, L. (2009). *Domestic Abuse, Stalking and Harassment and Honour Based Violence (DASH, 2009) Risk Identification and Assessment and Management Model*. https://www.dashriskchecklist.co.uk/wp-content/uploads/2016/09/DASH-2009.pdf

Robbins, R., McLaughlin, H., Banks, C., Bellamy, C., & Thackray, D. (2014). Domestic violence and multi- agency risk assessment conferences (MARACs): a scoping review. *The Journal of Adult Protection*, *16*(6), 389-398. https://doi.org/10.1108/JAP-03-2014-0012

Robinson, A. L., & Clancy, A. (2020). Systematically identifying and prioritising domestic abuse perpetrators for targeted intervention. *Criminology & Criminal Justice*, *21*(5), 687-704. https://doi.org/10.1177/1748895820914380

Robinson, A. L., & Howarth, E. (2012). Judging risk. *Journal of Interpersonal Violence*, *27*(8), 1489-1518. https://doi.org/10.1177/0886260511425792

Robinson, A. L., Myhill, A., & Wire, J. (2018). Practitioner (mis) understandings of coercive control in England and Wales. *Criminology & Criminal Justice, 18*(1), 29-49. https://doi.org/10.1177/1748895817728381

Robinson, A. L., & Tregidga, J. (2007). The perceptions of high-risk victims of domestic violence to a coordinated community response in Cardiff, Wales. *Violence Against Women, 13*(11), 1130-1148. https://doi.org/10.1177/1077801207307797

Ryan, J., & Roman, N. (2021). Understanding family Violence: A family-centred perspective. In N. Roman (Ed.), *Family Violence: Prevalence, Risk Factors and Perspectives*. Nova Science Publishers. https://doi.org/10.52305/bnif2138

Saxton, M. D., Jaffe, P. G., Dawson, M., Straatman, A.-L., & Olszowy, L. (2022). Complexities of the police response to intimate partner violence: Police officers' perspectives on the challenges of keeping families safe. *Journal of Interpersonal Violence*, *37*(5-6), 2557-2580. https://doi.org/10.1177/0886260520934428

Schiamberg, L. B., & Gans, D. (1999). An ecological framework for contextual risk factors in elder abuse by adult children. *Journal of Elder Abuse & Neglect*, *11*(1), 79-103. https://doi.org/10.1300/J084v11n01_05

Schiamberg, L. B., & Gans, D. (2000). Elder abuse by adult children: An applied ecological framework for understanding contextual risk factors and the intergenerational character of quality of life. *The International Journal of Aging and Human Development*, *50*(4), 329-359. https://doi.org/10.2190/dxax-8tj9-rg5k-mpu5

Schreiber, J. B. (2017). Latent class analysis: An example for reporting results. *Research in Social and Administrative Pharmacy*, *13*(6), 1196–1201. https://doi.org/10.1016/j.sapharm.2016.11.011

Shorrock, S., McManus, M. A., & Kirby, S. (2020). Profile of repeat victimization within multi-agency referrals. *International Review of Victimology*, *26*(3), 332-343. https://doi.org/10.1177/0269758020902890

Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences and the Law*, *31*(1), 8-22. https://doi.org/10.1002/bsl.2052

Sloper, P. (2004). Facilitators and barriers for co-ordinated multi-agency services. *Child: Care, Health & Development*, *30*(6), 571-580. https://doi.org/10.1111/j.1365-2214.2004.00468.x

Smith Slep, A. M., Foran, H. M., & Heyman, R. E. (2014). An ecological model of intimate partner violence perpetration at different levels of severity. *Journal of Family Psychology*, *28*(4), 470-482. https://doi.org/10.1037/a0037316

Smith Slep, A. M., & O'Leary, S. G. (2001). Examining partner and child abuse: Are we ready for a more integrated approach to family violence? *Clinical Child and Family Psychology Review*, *4*(2), 87-107. https://doi.org/10.1023/A:1011319213874

Spivak, B., McEwan, T., Luebbers, S., & Ogloff, J. (2020). Implementing evidence-based practice in policing family violence: The reliability, validity and feasibility of a risk assessment instrument for prioritising police response. *Policing and Society*, *31*(4), 483-502. https://doi.org/10.1080/10439463.2020.1757668

Stith, S. M., Green, N. M., Smith, D. B., & Ward, D. B. (2008). Marital satisfaction and marital discord as risk markers for intimate partner violence: A meta-analytic review. *Journal of Family Violence*, *23*(3), 149-160. https://doi.org/10.1007/s10896-007-9137-4

Stith, S. M., Liu, T., Davies, L. C., Boykin, E. L., Alder, M. C., Harris, J. M., Som, A., McPherson, M., & Dees, J. E. M. E. G. (2009). Risk factors in child maltreatment: A

meta-analytic review of the literature. *Aggression and Violent Behavior*, *14*(1), 13-29. https://doi.org/10.1016/j.avb.2006.03.006

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-analytic review. *Aggression and Violent Behavior*, *10*(1), 65-98. https://doi.org/10.1016/j.avb.2003.09.001

Stoltenborgh, M., Bakermans-Kranenburg, M. J., Alink, L. R., & van IJzendoorn, M. H. (2015). The prevalence of child maltreatment across the globe: Review of a series of meta-analyses. *Child Abuse Review*, *24*(1), 37-50. https://doi.org/10.1002/car.2353

Straus, M. A., & Douglas, E. M. (2017). Eight new developments, uses, and clarifications of the Conflict Tactics Scale. *Journal of Family Issues*, *38*(14), 1953-1973. https://doi.org/10.1177/0192513x17729720

Svalin, K., & Levander, S. (2019). The predictive validity of intimate partner violence risk assessment conducted by practitioners in different settings: A review of the literature. *Journal of Police and Criminal Psychology*, 35(2), 115–130. https://doi.org/10.1007/s11896-019-09343-4

Swartout, K. M., Cook, S. L., & White, J. W. (2012). Trajectories of intimate partner violence victimization. *Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health*, *13*(3), 272-277. https://doi.org/10.5811/westjem.2012.3.11788

Timshel, I., Montgomery, E., & Dalgaard, N. T. (2017). A systematic review of risk and protective factors associated with family related violence in refugee families. *Child Abuse & Neglect, 70*, 315-330. https://doi.org/10.1016/j.chiabu.2017.06.023

Tompson, L., Davies, S. & Polaschek, D. L. L. (2021). *Systematic evidence map of disparities in police outcomes: Final report.* University of Waikato. https://researchcommons.waikato.ac.nz/handle/10289/14860

Tucker, C. J., Finkelhor, D., Shattuck, A. M., & Turner, H. (2013). Prevalence and correlates of sibling victimization types. *Child Abuse & Neglect*, *37*(4), 213-223. https://doi.org/10.1016/j.chiabu.2013.01.006

Turner, E., Medina, J., & Brown, G. (2019). Dashing hopes? The predictive accuracy of domestic abuse risk assessment by police. *The British Journal of Criminology*, *59*(5), 1013-1034. https://doi.org/10.1093/bjc/azy074

van der Put, C. E., Gubbels, J., & Assink, M. (2019). Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools. *Aggression and Violent Behavior*, *47*, 100–116. https://doi.org/10.1016/j.avb.2019.03.008

Walker, K., Bowen, E., & Brown, S. (2013). Desistance from intimate partner violence: A critical review. *Aggression and Violent Behavior*, *18*(2), 271-280. https://doi.org/10.1016/j.avb.2012.11.019

Walklate, S., & Mythen, G. (2011). Beyond risk theory: Experiential knowledge and 'knowing otherwise'. *Criminology & Criminal Justice*, *11*(2), 99-113. https://doi.org/10.1177/1748895811398456

Ward, T., Yates, P. M., & Willis, G. M. (2012). The Good Lives Model and the Risk Need Responsivity Model: A critical response to Andrews, Bonta, and Wormith (2011). *Criminal Justice and Behavior*, *39*(1), 94-110. https://doi.org/10.1177/0093854811426085

Weeks, L. E., & Leblanc, K. (2011). An ecological synthesis of research on older women's experiences of intimate partner violence. *Journal of Women & Aging*, *23*(4), 283-304. https://doi.org/10.1080/08952841.2011.611043

Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent class analysis: A guide to best practice. *Journal of Black Psychology, 46*(4), 287–311. https://doi.org/10.1177/0095798420930932

Whitehead, P. R., Ward, T., & Collie, R. M. (2007). Time for a change: Applying the Good Lives Model of rehabilitation to a high-risk violent offender *International Journal of Offender Therapy and Comparative Criminology*, *51*(5), 578-598. https://doi.org/10.1177/0306624x06296236

Whiting, J. B., Merchant, L. V., Bradford, A. B., & Smith, D. B. (2020). The ecology of family violence. In K. S. Wampler, M. Rastogi, & R. Singh (Eds.), *The Handbook of Systemic Family Therapy* (Vol. 4). John Wiley & Sons Ltd. https://doi.org/10.1002/9781119788409.ch7

Wiener, C. (2017). Seeing what is 'invisible in plain sight': Policing coercive control. *The Howard Journal of Crime and Justice*, *56*(4), 500-515. https://doi.org/10.1111/hojo.12227

Williams, K. R., & Grant, S. R. (2006). Empirically examining the risk of intimate partner violence: The revised Domestic Violence Screening Instrument (DVSI-R). *Public Health Reports*, *121*(4), 400-408. https://doi.org/10.1177/003335490612100408

Woldgabreal, Y., Day, A., & Tamatea, A. (2020). Do risk assessments play a role in the enduring 'color line'? *Advancing Corrections Journal*, *10*, 18-28. http://hdl.handle.net/11343/252731

Wortley, R. (2011). *Psychological Criminology: An Integrative Approach.* Willan. https://doi.org/10.4324/9780203806098

Zhang, Q., Jones, A. S., Rijmen, F., & Ip, E. H. (2010). Multivariate discrete hidden Markov models for domain-based measurements and assessment of risk factors in child

development. *Journal of Computational and Graphical Statistics*, *19*(3), 746-765. https://doi.org/10.1198/jcgs.2010.09015

Zottola, S. A., Desmarais, S. L., Lowder, E. M., & Duhart Clarke, S. E. (2021). Evaluating fairness of algorithmic risk assessment instruments: The problem with forcing dichotomies. *Criminal Justice and Behavior*, 009385482110405. https://doi.org/10.1177/00938548211040544

Zucchini, W., MacDonald, I. L., & Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R* (2nd ed.). CRC Press. https://doi.org/10.1201/b20790

Appendix A: Co-authorship Forms

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Mataurangs Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3: Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2020). Predictive validity of the DYRA and SAFVR: New Zealand police's family violence risk assessment measures. *Criminal Justice and Behavior, X,* X-X.

Nature of contribution by PhD candidate	Collected and analysed data, wrote the paper.
Extent of contribution by PhD candidate (%)	70%

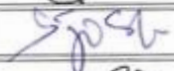
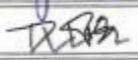
CO-AUTHORS

Name	Nature of Contribution
Dr. Chaitanya Joshi	Offered advice throughout the project on data analysis, and presentation.
Prof. Devon Polaschek	Offered advice throughout the project on content, writing, data analysis, and presentation.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Apriel Jolliffe Simpson		02/06/2020
Dr. Chaitanya Joshi		02/06/2020
Prof. Devon Polaschek		02/06/2020

July 2015

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Waharoa Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3: Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (Manuscript submitted for publication). Unpacking multi-agency structured professional judgement risk assessment for family violence.

Nature of contribution by PhD candidate	Collected and analysed data, wrote the paper.
Extent of contribution by PhD candidate (%)	70%

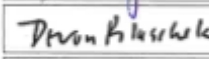
CO-AUTHORS

Name	Nature of Contribution
Dr. Chaitanya Joshi	Offered advice throughout the project on data analysis and written presentation.
Prof. Devon Polaschek	Offered advice throughout the project on data analysis and written presentation.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Apriel Jolliffe Simpson		02/07/2022
Dr. Chaitanya Joshi		02/07/2022
Prof. Devon Polaschek		02/07/2022

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 5: Jolliffe Simpson, A. D., Joshi, C. & Polaschek, D. L. L. (2022). Modeling behavioral patterns of family violence aggressors in Aotearoa New Zealand [Manuscript submitted for publication].

Nature of contribution by PhD candidate	Collected and analysed data, wrote the paper.
Extent of contribution by PhD candidate (%)	70%

CO-AUTHORS

Name	Nature of Contribution
Dr. Chaitanya Joshi	Offered advice througout the project on modeling techniques, analysis and written presentation.
Prof. Devon Polaschek	Offered advice througout the project on analysis and written presentation.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Apriel Jolliffe Simpson		06/08/2022
Dr. Chaitanya Joshi		8/8/22
Prof. Devon Polaschek	Devon Polaschek	06/08/2022

Appendix B: Supplemental Materials for Manuscript One

Table S1

AUC for DYRA and SAFVR Risk Categories Predicting FV Recurrence Within 3 days Across Different Relationship and Aggressor Characteristics

	n (% recurrence)[a]	AUC[b] 95% CI	*SE*	*p*	SN[b]	SP[b]	PPV[b]	NPV[b]
DYRA								
Type of FV								
IPV (all)	1249 (5.6)	.52 [46-.58]	.03	.573	.81	.28	.06	.96
IPV + children	611 (5.7)	.52 [.43-.61]	.05	.668	.86	.25	.06	.97
IPV - children	638 (5.5)	.52 [.47-.61]	.05	.726	.77	.32	.06	.96
Other FV (all)	568 (3.7)	.55 [.43-.67]	.06	.436	.71	.39	.04	.97
Other FV + children	307 (3.6)	.62 [.48-.76]	.07	.183	.91	.38	.05	.99
Other FV - children	261 (3.8)	.47 [.28-.66]	.10	.712	.50	.41	.03	.95
Aggressor characteristics								
Male	1368 (5.6)	.51 [.45-.57]	.03	.785	.78	.30	.06	.96
Female	446 (3.4)	.62 [.49-.75]	.06	.114	.87	.38	.05	.99
Under 18	139 (3.6)	.58 [.33-.82]	.12	.572	.80	.33	.04	.98
Over 18	1678 (5.1)	.53 [.47-.59]	.03	.340	.79	.32	.06	.97
Māori	735 (5.6)	.53 [.45-.62]	.04	.493	.73	.35	.06	.96
European	860 (4.8)	.54 [.46-.62]	.04	.351	.88	.27	.06	.98
SAFVR								
Type of FV								
IPV (all)	1249 (5.6)	.53 [.46-.60]	.04	.443	.71	.30	.06	.95
IPV + children	611 (5.7)	.50 [.41-.60]	.05	.946	.71	.27	.06	.94
IPV - children	638 (5.5)	.55 [.45-.65]	.05	.326	.71	.33	.06	.95
Other FV (all)	568 (3.7)	.55 [.41-.70]	.07	.414	.62	.49	.04	.97
Other FV + children	307 (3.6)	.49 [.29-.68]	.10	.880	.45	.57	.04	.97
Other FV - children	261 (3.8)	.62 [.44-.82]	.10	.166	.80	.40	.05	.98
Aggressor characteristics								
Male	1368 (5.6)	.51 [.68-.45]	.04	.683	.71	.28	.05	.94
Female	446 (3.4)	.62 [.47-.77]	.08	.109	.60	.60	.05	.98
Under 18	139 (3.6)	.46 [.17-.75]	.15	.756	.20	.90	.07	.97
Over 18	1678 (5.1)	.54 [.48-.60]	.03	.208	.72	.32	.05	.95
Māori	735 (5.6)	.57 [.49-.66]	.05	.115	.85	.22	.06	.96
European	860 (4.8)	.53 [.43-.62]	.05	.589	.63	.39	.05	.95

Note SN = sensitivity; SP = specificity

[a]The number of aggressors in the sub-sample and the percentage of aggressors with a recurrence in the sub-sample.

[b]Interpretive criteria for AUCs from Rice & Harris (2005): .56-.63 = small, .64-.70 = moderate, .71-1.0 = large

Table S2

AUC for DYRA and SAFVR Risk Categories Predicting FV Recurrence Within 12 weeks Across Different Relationship and Aggressor Characteristics

	n (% recurrence)[a]	AUC[b] [95% CI]	*SE*	*p*	SN[b]	SP[b]	PPV[b]	NPV[b]
DYRA								
Type of FV								
IPV (all)	1249 (38.1)	.54 [.51-.58]	.02	.011	.76	.30	.40	.68
IPV + children	611 (37.0)	.54 [.49-.58]	.02	.153	.80	.27	.39	.70
IPV - children	638 (39.2)	.55 [.50-.60]	.02	.033	.73	.34	.41	.66
Other FV (all)	568 (27.5)	.55 [.49-.60]	.03	.088	.67	.41	.30	.77
Other FV + children	307 (28.0)	.56 [.48-.63]	.04	.133	.70	.39	.31	.77
Other FV - children	261 (26.8)	.53 [.46-.61]	.04	.401	.64	.43	.29	.77
Aggressor characteristics								
Male	1368 (36.9)	.54 [.51-.57]	.02	.008	.76	.33	.40	.70
Female	446 (28.0)	.55 [.49-.61]	.03	.100	.67	.38	.30	.75
Under 18	139 (28.8)	.60 [.50-.70]	.05	.063	.80	.37	.34	.82
Over 18	1678 (35.3)	.54 [.52-.57]	.02	.002	.74	.34	.38	.70
Māori	735 (41.1)	.54 [.50-.58]	.02	.071	.70	.37	.44	.64
European	860 (33.3)	.56 [.52-.60]	.02	.005	.79	.29	.36	.74
SAFVR								
Type of FV								
IPV (all)	1249 (38.1)	.61 [.57-.64]	.02	< .001	.80	.36	.43	.74
IPV + children	611 (37.0)	.59 [.54-.63]	.02	< .001	.81	.32	.41	.74
IPV - children	638 (39.2)	.62 [.58-.67]	.02	< .001	.79	.39	.46	.74
Other FV (all)	568 (27.5)	.58 [.53-.63]	.03	.002	.58	.51	.31	.76
Other FV + children	307 (28.0)	.56 [.49-.63]	.04	.115	.44	.57	.29	.73
Other FV - children	261 (26.8)	.62 [.54-.69]	.04	.002	.74	.44	.33	.82
Aggressor characteristics								
Male	1368 (36.9)	.59 [.56-.62]	.02	< .001	.79	.32	.41	.73
Female	446 (28.0)	.64 [.59-.70]	.03	< .001	.54	.65	.38	.79
Under 18	139 (28.8)	.63 [.53-.73]	.05	.011	.13	.90	.33	.72
Over 18	1678 (35.3)	.61 [.58-.63]	.01	< .001	.79	.37	.40	.76
Māori	735 (41.1)	.58 [.53-.62]	.02	< .001	.85	.26	.45	.72
European	860 (33.3)	.59 [.55-.63]	.02	< .001	.69	.42	.37	.73

Note SN = sensitivity; SP = specificity

[a]The number of aggressors in the sub-sample and the percentage of aggressors with a recurrence in the sub-sample.

[b]Interpretive criteria for AUCs from Rice & Harris (2005): .56-.63 = small, .64-.70 = moderate, .71-1.0 = large

Appendix C: Supplemental Materials for Manuscript Two

Table S1

ISR Risk Assessment Guide Items versus the Factors Coded from the ISR Triage Teams' Risk Assessments

ISR guide item	*Factor*	ICC	ICC label
Victim risks			
Pregnancy or recent birth	Pregnancy or recent birth	.91	Excellent
Mental health issues	Mental health issues	.71	Moderate
Alcohol and/or drug abuse	Alcohol use	.89	Good
	Drug use	.81	Good
Suicide/self-harm	Suicide/self-harm	.71	Moderate
Isolation	Isolation [c]	.73	Moderate
Deprivation	Housing problems [c]	.74	Moderate
Non-engagement	Non-engagement [c]	.55	Moderate
History of victimization/cumulative harm	FV/criminal history	.81	Good
Fear for own safety	Fear	.98	Excellent
Relationship risks			
Recent separation/recent protection order	Separation	.97	Excellent
	Protection order	[e]	[e]
Escalation in frequency and/or severity of violence	Escalation	.87	Good
	Frequent episodes	.94	Excellent
	Duration of episodes	.95	Excellent
Financial difficulties	Financial stress	.94	Excellent
Previous violence in relationships	[a]		
Gang	Gang affiliations	1	Perfect
Children/young people risks			
Stepchildren/children from prior relationships	Non-biological children	.97	Excellent
Vulnerability	Vulnerable children	.76	Good
History of victimization/cumulative harm	History of victimization/cumulative harm	.50	Moderate
Challenging behavior displayed by children	Challenging behavior	.11	Poor

Table S1 Continued

ISR guide item	*Factor*	ICC	ICC label
Aggressor risks			
History of violent behavior/FV	FV/criminal history	.90	Excellent
Attempted to kill the victim	[b]		
Harmed/threatened to harm the victim	Electronic/social media harm	1	Perfect
	Fraud/theft	0	Poor
	Harassment	[e]	[e]
	Injuries	.98	Excellent
	Intimidating behavior	.55	Moderate
	Items thrown	.91	Excellent
	Physical harm	.94	Excellent
	Property damage	.97	Excellent
	Psychological abuse	.89	Good
	Threats	.81	Good
	Verbal abuse	.96	Excellent
Forced sex	Sexual assault	[e]	[e]
Strangulation, suffocation, drowning, choking	Non-fatal strangulation	.65	Moderate
Breach of protection order	Breach of protection order	1	Perfect
Weapons/access to weapons	Weapons	.94	Excellent
Unemployment/loss of employment	Unemployment [c]	.93	Excellent
Stalked the victim	Stalking	[e]	[e]
Coercive and controlling behaviors, obsessive, jealous	Controlling behavior	.86	Good
Harmed/threatened to harm, intimidated children, family members, or pets	Harmed/threatened to harm, intimidated children, family members, or pets	.62	Moderate
Suicide/self-harm	Suicide/self-harm	.86	Good
Alcohol and/or drug abuse	Alcohol use	.93	Excellent
	Drug use	.78	Good
Mental health issues	Mental health issues	.85	Good
Previous police involvement relevant to family harm	[a]		
Practice considerations			
Escalation/change in risk	[a]		
Vulnerability	Vulnerable elderly person	.86	Good
Previous patterns of behaviors	[a]		
Further information required/verification	Lack of information	.96	Excellent
Key risk factors to target	[d]		
Impact on children	[d]		
Complex needs	[a]		

Table S1 Continued

ISR guide item	*Factor*	ICC	ICC label
Engaging family/whānau/key supports	Cultural barriers	.83	Good
	Breakdown in family dynamics	.77	Good
	Complex family dynamics	.54	Moderate
	Lack of support	.82	Good
	Collusive family	[e]	[e]
Protective factors/strengths to build	[b]		
Cumulative harm	[a]		
Barriers to change	Normalization	.95	Excellent
	Denial of FV	1	Perfect
	Unreported FV	1	Perfect
	Intergenerational FV	1	Perfect
Other	Aggressor not spoken to	1	Perfect
	Breaching/non-complying with conditions/orders	.97	Excellent
	Cohabiting	.82	Good
	Conflicting stories	1	Perfect
	Custody issues	.82	Good
	Delay in police attendance	1	Perfect
	Denial of end of relationship	.89	Good
	Episode occurred in public	.95	Excellent
	Infidelity	.75	Good
	Left address/not located [c]	.69	Moderate
	No formal orders	.95	Excellent
	On/off relationship	.76	Good
	Property dispute	.75	Good
	Refusing to leave [c]	1	Perfect
	Situational stress	.76	Good
	Uncooperative with police [c]	.84	Good
	Victim not spoken to	.80	Good

Note. FV = Family Violence; ISR = Integrated Safety Response; ICC = Intraclass Correlation Coefficients. Intraclass correlation coefficients were interpreted using guidelines from Koo and Li (2015): < .50 = poor, .50-.74 = moderate, .75-.89 = good, .90-1.0 = excellent.

[a] This item overlapped with other items, and we did not code a separate variable for it

[b] This item was not present in the data

[c] This variable was coded at the level of the family unit rather than for the aggressor or victim because we could not tell which person the variable related to

[d] This item was too vague for us to code

[e] This variable was absent in all cases coded for inter rater reliability

The variable '*any mention of children'* was excluded from analyses because it was not independent from other variables

Table S2

ISR Factor Definitions

	Definition
Aggressor	
Alcohol use	Current alcohol use or historic alcohol problems for the aggressor.
Breaching/non-complying with conditions/orders	The aggressor breached or is not complying with bail, sentence conditions, a police safety order, or other order from Police, Ara Poutama (Department of Corrections), or Ministry of Justice (excluding protection orders).
Drug use	Current illicit drug use or illicit historic drug problems for the aggressor.
FV/criminal history	A history of FV as the aggressor or victim, or a criminal history, for the aggressor.
Mental health issues	Current or historic mental health issues in general, or specific mental health conditions, or intellectual disability, or traumatic brain injury, for the aggressor.
Suicide/self-harm	Current or historic suicide or self-harm, including threatened or attempted suicide, and suicidal ideation, for the aggressor.
Victim	
Alcohol use	Current alcohol use or historic alcohol problems for the victim.
Drug use	Current illicit drug use or illicit historic drug problems for the victim.
Fear	Victim fear for the safety of themselves or others.
FV/criminal history	A history of FV as the aggressor or victim, or a criminal history, for the victim.
Mental health issues	Current or historic mental health issues in general, or specific mental health conditions, or intellectual disability, or traumatic brain injury, for the victim.
Suicide/self-harm	Current or historic suicide or self-harm, including threatened or attempted suicide, and suicidal ideation, for the victim.
Relationship	
Cohabiting	The aggressor and victim lived together, or one party was bailed to/was under sentence conditions to live at the other party's address.
Custody issues	Dispute about custody arrangements, parenting, or family court proceedings.
Denial of end of relationship	The intimate relationship between the aggressor and victim had ended but one party was denying or not accepting the relationship's end.
Financial stress	Dispute about money or financial stress.
Gang affiliations	The aggressor or victim had gang affiliations.

Table S2 Continued

	Definition
Housing problems	There were problems with the physical space the aggressor and/or victim lives in, e.g., living in emergency housing, overcrowding, unsanitary conditions, homelessness, or pending eviction.
Infidelity	Current or historic, suspected or confirmed, infidelity by the aggressor or victim.
Isolation	The victim or family unit are socially or geographically isolated, or do not have a phone, or have limited/no access to transportation.
On/off relationship	The intimate relationship between the aggressor and victim was on/off, unstable, or the relationship status was unclear.
Property dispute	Dispute about property, including separating relationship property after the end of a relationship.
Protection order	There is a protection order between the aggressor and victim. Mutually exclusive with the *breach of protection order* variable.
Separation	The intimate relationship between the aggressor and victim has ended, recently or historically, or one party wants to separate.
Situational stress	Dispute triggered by situational stressors such as illness, grief, or other life events, excluding financial stress.
Unemployment	The aggressor or victim is unemployed or has lost their job.
Child/family	
Breakdown in family dynamics	Dispute triggered by a breakdown in family dynamics.
Collusive family	Other family members actively colluded with the aggressor (e.g., covered up or obscured information, pressured children not to disclose FV).
Complex family dynamics	Complex family dynamics involving extended family.
History of victimization/cumulative harm	FV has previously affected the child(ren), e.g., previous involvement by Oranga Tamariki (child protection services).
Pregnancy or recent birth	The victim or aggressor is pregnant, was recently pregnant, or recently gave birth.
Non-biological children	There were children in the family that were not the biological child of one of the parents in the family unit, e.g., children from previous relationship.
Vulnerable children	There were vulnerable children, defined as children under 5 years of age, or children with mental health issues or disabilities.
Vulnerable elderly person	An elderly person was harmed or taken advantage of by a family member.

Table S2 Continued

	Definition
Harm	
Controlling behavior	The aggressor used controlling behavior during the index episode or had a history of using controlling behavior towards the victim.
Electronic/social media harm	The aggressor harmed the victim through social media, text message, or other electronic means.
Harassment	The aggressor harassed the victim.
Harmed/threatened to harm, intimidated children, family members, or pets	The aggressor harmed or threatened to harm the victims' children, family members or pets.
Injuries	One or more parties received injuries in the index episode.
Intimidating behavior	The aggressor used intimidating behavior not better captured by another harm category during the index episode.
Items thrown	The aggressor threw items during the index episode.
Non-fatal strangulation	The aggressor non-fatally strangled the victim during the index episode or in a previous episode.
Physical harm	The aggressor physically harmed the victim.
Property damage	The aggressor damaged property belonging to the victim or themselves during the index episode.
Psychological abuse	The aggressor used psychological abuse during the index episode, or recently psychologically abused the victim.
Sexual assault	The aggressor sexually assaulted the victim.
Stalking	The aggressor stalked the victim.
Threats	The aggressor made threats during the index episode.
Verbal abuse	The aggressor used verbal abuse during the index episode.
Weapons	The aggressor used weapons in the index episode or has access to weapons.
Episode	
Aggressor not spoken to	Police did not speak to the aggressor when responding to the index episode.
Breach of protection order	The aggressor breached a protection order during the index episode.
Conflicting stories	The aggressor and victim gave contrasting explanations for the index episode.
Delay in police attendance	There was a delay between police being called and attending the episode.
Episode occurred in public	The index episode occurred in public or in a vehicle.
Left address/not located	The aggressor or victim left the address and were not located by police responding to the index episode.
Refusing to leave	The aggressor or victim refused to leave during the index episode.

Table S2 Continued

	Definition
Uncooperative with police	The aggressor or victim did not cooperate with police during the index episode.
Victim not spoken to	Police did not speak to the victim when responding to the index episode.
Practice considerations	
Cultural barriers	The aggressor or victim speak a language not commonly spoken in New Zealand, or cultural issues prevent them reporting FV episodes.
Denial of FV	The aggressor or victim deny the index episode occurred.
Duration of episodes	The index episode was drawn out or occurred over a long period of time (i.e., ISR used words like prolonged or sustained).
Escalation	There was escalating behavior in the index episode or in recent episodes.
Frequent episodes	FV occurred frequently or there were multiple calls for police service in a relatively short period of time.
Intergenerational FV	The FV involved family members from more than two generations.
Lack of information	There were gaps in information, statements were not forthcoming, or there was otherwise not enough information available about the circumstances of the FV episode.
Lack of support	The victim or family unit had limited or no support in their social network.
No formal orders	There were no formal orders in place between the aggressor and victim to prevent FV, e.g., protection order or parenting order.
Non-engagement	The victim or aggressor declined referrals or did not engage with support services or had a history of not engaging with support services.
Normalization	FV was a normalized experience for the victim or family unit.
Unreported FV	FV episodes involving the aggressor and victim were not reported in the past.

Note. FV = Family Violence; ISR = Integrated Safety Response.

Appendix D: Supplemental Materials for Chapter Four

Table S1

Conditional Probability of ISR Triage Teams' Recorded Factors By Latent Class

Factor	*f* (%)	*1* (*n* = 149)	*2* (*n* = 208)	*3* (*n* = 485)
Aggressor				
Alcohol use	229 (27.2)	.22	.40	.23
Breaching/non-complying with conditions/orders	78 (9.3)	.14	.17	.04
Drug use	151 (17.9)	.27	.24	.12
FV/criminal history	231 (27.4)	.49	.51	.09
Mental health issues	134 (15.9)	.31	.09	.14
Suicide/self-harm	90 (10.7)	.27	.11	.05
Victim				
Alcohol use	99 (11.8)	.00	.24	.10
Drug use	39 (4.6)	.01	.10	.03
Fear	97 (11.5)	.35	.11	.03
FV/criminal history	212 (25.2)	.47	.49	.07
Mental health issues	64 (7.6)	.10	.09	.06
Suicide/self-harm	24 (2.9)	.06	.04	.01
Relationship				
Cohabiting	44 (5.2)	.01	.08	.05
Custody issues	38 (4.5)	.04	.03	.06
Denial of end of relationship	29 (3.4)	.11	.00	.03
Financial stress	92 (10.9)	.15	.09	.10
Gang affiliations	40 (4.8)	.08	.10	.01
Housing problems	70 (8.3)	.07	.15	.06
Infidelity	30 (3.6)	.08	.03	.02
Isolation	64 (7.6)	.17	.10	.03
On/off relationship	39 (4.6)	.07	.07	.03
Property dispute	28 (3.3)	.01	.02	.05
Protection order	13 (1.5)	.06	.02	.00
Separation	157 (18.6)	.24	.16	.18
Situational stress	163 (19.4)	.22	.16	.20
Unemployment	42 (5.0)	.10	.06	.03
Child/family				
Breakdown in family dynamics	29 (3.4)	.02	.06	.03
Collusive family	6 (0.7)	.01	.01	.00
Complex family dynamics	21 (2.5)	.00	.06	.02
History of victimization/cumulative harm	35 (4.2)	.05	.13	.00
Pregnancy or recent birth	40 (4.8)	.08	.10	.01
Non-biological children	79 (9.4)	.09	.18	.05
Vulnerable children	100 (11.9)	.17	.16	.08
Vulnerable elderly person	11 (1.3)	.05	.00	.01

Table S1 continued

Factor	*f* (%)	*1* (*n* = 149)	*2* (*n* = 208)	*3* (*n* = 485)
Harm				
Controlling behavior	176 (20.9)	.47	.22	.11
Electronic/social media harm	25 (3.0)	.10	.00	.02
Harassment	9 (1.1)	.01	.00	.02
Harmed/threatened to harm, intimidated children, family members, or pets	28 (3.3)	.03	.08	.01
Injuries	86 (10.2)	.05	.30	.03
Intimidating behavior	123 (14.6)	.27	.20	.08
Items thrown	42 (5.0)	.08	.08	.03
Non-fatal strangulation	43 (5.1)	.09	.10	.02
Physical harm	263 (31.2)	.29	.68	.15
Property damage	132 (15.7)	.15	.20	.14
Psychological abuse	33 (3.9)	.15	.02	.01
Sexual assault	5 (0.6)	.01	.00	.01
Stalking	31 (3.7)	.12	.01	.02
Threats	106 (12.6)	.36	.09	.06
Verbal abuse	364 (43.2)	.26	.27	.56
Weapons	73 (8.7)	.17	.16	.03
Episode				
Aggressor not spoken to	106 (12.6)	.13	.19	.10
Breach of protection order	28 (3.3)	.07	.01	.03
Conflicting stories	19 (2.3)	.01	.07	.01
Delay in police attendance	30 (3.6)	.04	.03	.03
Episode occurred in public	61 (7.2)	.07	.11	.06
Left address/not located	57 (6.8)	.10	.09	.05
Refusing to leave	40 (4.8)	.03	.04	.06
Uncooperative with police	88 (10.5)	.08	.18	.08
Victim not spoken to	26 (3.1)	.03	.04	.03
Practice considerations				
Cultural barriers	9 (1.1)	.02	.01	.01
Denial of FV	18 (2.1)	.00	.05	.01
Duration of episodes	59 (7.0)	.12	.13	.02
Escalation	94 (11.2)	.26	.20	.02
Frequent episodes	82 (9.7)	.10	.11	.09
Intergenerational FV	22 (2.6)	.02	.08	.01
Lack of information	184 (21.9)	.07	.53	.13
Lack of support	23 (2.7)	.06	.04	.01
No formal orders	39 (4.6)	.06	.05	.04
Non-engagement	70 (8.3)	.05	.16	.06
Normalization	65 (7.7)	.04	.19	.04
Unreported FV	141 (16.7)	.31	.27	.07

Note. f = frequency. FV = Family Violence.

Appendix E: Supplemental Materials for Manuscript Three

Table S1

Transition Probabilities for the Three State Model, with Standard Errors in Brackets

		T2	
T1	State 1	State 2	State 3
State 1	.84 (.08)	.06 (.09)	.11 (.02)
State 2	.38 (.28)	.60 (.29)	.01 (.09)
State 3	.28 (.07)	.42 (.07)	.31 (.03)
		T3	
T2	State 1	State 2	State 3
State 1	.99 (.03)	.00 (.00)	.01 (.03)
State 2	.03 (.14)	.59 (.12)	.38 (.07)
State 3	.26 (.07)	.32 (.07)	.42 (.04)
		T4	
T3	State 1	State 2	State 3
State 1	.95 (.03)	.00 (.02)	.05 (.01)
State 2	.12 (.14)	.57 (.12)	.29 (.06)
State 3	.27 (.04)	.25 (.04)	.48 (.04)
		T5	
T4	State 1	State 2	State 3
State 1	.97 (.01)	.00 (.00)	.03 (.01)
State 2	.00 (.05)	.54 (.07)	.46 (.07)
State 3	.25 (.05)	.34 (.05)	.41 (.04)
		T6	
T5	State 1	State 2	State 3
State 1	.95 (.02)	.00 (.01)	.04 (.01)
State 2	.00 (.00)	.57 (.08)	.43 (.08)
State 3	.33 (.07)	.27 (.07)	.40 (.04)
		T7	
T6	State 1	State 2	State 3
State 1	.97 (.03)	.01 (.03)	.03 (.01)
State 2	.11 (.21)	.48 (.17)	.41 (.10)
State 3	.41 (.06)	.20 (.07)	.40 (.04)
		T8	
T7	State 1	State 2	State 3
State 1	.96 (.02)	.00 (.00)	.04 (.02)
State 2	.00 (.05)	.34 (.23)	.66 (.23)
State 3	.55 (.04)	.00 (.01)	.45 (.04)

Figure S1

The Proportion of Aggressors With Further Episodes Reported to Police and the Proportion of Aggressors With No Further Episodes Reported to Police at Each of the Eight Three-Month Time Intervals