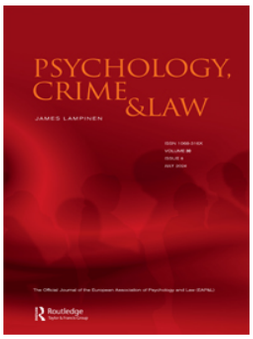

Psychology, Crime & Law

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gpcl20

Assessing risk among women who perpetrate intimate partner abuse

Julia Nazarewicz, Michael D. Trood, Troy E. McEwan, Susanne Strand, Stefan Luebbers & Benjamin L. Spivak

To cite this article: Julia Nazarewicz, Michael D. Trood, Troy E. McEwan, Susanne Strand, Stefan Luebbers & Benjamin L. Spivak (27 Jun 2024): Assessing risk among women who perpetrate intimate partner abuse, Psychology, Crime & Law, DOI: 10.1080/1068316X.2024.2369239

To link to this article: https://doi.org/10.1080/1068316X.2024.2369239

PSYCHOLOGY, CRIME & LAW
https://doi.org/10.1080/1068316X.2024.2369239

Routledge
Taylor & Francis Group

Assessing risk among women who perpetrate intimate partner abuse

Julia Nazarewicz[a]†, Michael D. Trood[a]†, Troy E. McEwan[a,b], Susanne Strand[a,c], Stefan Luebbers[a,d] and Benjamin L. Spivak[a]

[a]Centre for Forensic Behavioural Science, Swinburne University of Technology and the Victorian Institute for Forensic Mental Health (Forensicare), Melbourne, Australia; [b]Centre for Education and Research in Forensic Psychology, University of Kent, Canterbury, UK; [c]School of Behavioural, Social and Legal Sciences, Örebro University, Örebro, Sweden; [d]Youth Forensic Specialist Service, Alfred Health, Melbourne, Australia

ABSTRACT

This paper presents a prospective evaluation of the predictive validity of three risk assessment instruments in a sample of Australian women identified by police as intimate partner abuse (IPA) perpetrators. Using a subsample from Spivak et al. (2020), 410 female IPA perpetrators were screened using the Victoria Police Screening Assessment for Family Violence Risk (VP-SAFvR) and evaluated alongside two samples of 60 and 229 female IPA perpetrators assessed using the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER) and a modified version of the Lethality Screen respectively. Of the three instruments, the VP-SAFvR possessed indicators of effective discrimination (i.e. sensitivity, specificity, area under the curve) and predictive validity (i.e. positive predictive value, negative predictive value) on general IPA recidivism and its intended outcome of family or intimate partner abuse. The B-SAFER risk judgement similarly predicted its intended outcome of physical IPA recidivism, with notable indicators of discrimination and predictive validity. The results of the Modified Lethality Screen were conversely mixed on measures of discrimination and prediction for its intended outcome of severe IPA. The current findings suggest that these instruments function consistently for women and men who are identified by police as perpetrating family or intimate partner abuse.

ARTICLE HISTORY
Received 10 October 2022
Accepted 7 June 2024

KEYWORDS
Intimate partner abuse; actuarial assessment; structured professional judgement; violence risk assessment; female perpetrators

Over the past 20 years police have been increasingly expected to effectively assess and manage risks associated with intimate partner abuse (IPA; the physical, sexual and/or psychological abuse of an individual by their former or current intimate partner; Medina-Ariza et al., 2016). Attempting to improve the efficacy of police assessment and management, many police organisations employ risk assessment instruments that guide judgements about a person's risk of further IPA victimisation or perpetration (see

†These authors contributed equally to this work and should be considered equal in order of authorship

Supplemental data for this article can be accessed online at https://doi.org/10.1080/1068316X.2024.2369239.

Graham et al., 2021). Although multiple existing instruments have been shown to provide valid assessments of the risk of future violence in cases of male-to-female IPA (see van der Put et al., 2019), it remains unclear whether existing IPA risk assessments developed for use by police can accurately assess risk when the identified perpetrator is a woman. While there is little doubt that most injury and death caused by IPA is attributable to men (see Straus, 2011), improving understanding and management of the significant subgroup of abuse perpetrated by women remains an important research and public safety goal.

Women identified as perpetrators of physical and non-physical intimate partner abuse

More women than ever before are being identified as perpetrators of IPA by police. In Australia, for instance, the number of women recorded as family violence (FV; the physical, sexual and/or psychological abuse of an individual by a relative or legal guardian) perpetrators by police in the state of Victoria increased by 27.97% (while male perpetration rates rose by only 16.48%) between June 2018 and June 2022 (Crime Statistics Agency, 2022). While the rise in female IPA perpetration has been attributed to the misidentification of women as perpetrators when they are actually victims defending themselves (see Gerstenberger & Williams, 2013), there is considerable evidence that some women do instigate or predominantly commit abuse in their relationships (Babcock et al., 2003; Babcock et al., 2016; Henning et al., 2006; Hines & Douglas, 2010, 2019; Straus, 2011).

This evidence suggests that there are important gender differences in abusive behaviour, and that women engage in different forms of abuse (e.g. physical, emotional, sexual) at different rates to men. For example, despite apparent gender symmetry in physical abuse perpetration (Ehrensaft et al., 2004; Fiebert, 2014; Straus, 2004), men cause more physical injury to their partners than women. This is perhaps attributable to differences in their average physical size and strength (Cantos et al., 1994; Straus, 2011). In contrast, women may engage in psychological abuse and controlling behaviour to an equal or greater extent than men, though the methods of perpetrating this abuse may vary (Bates & Graham-Kevan, 2016; Carney & Barner, 2012; Hines & Douglas, 2010).

Importantly, sex differences in physical and non-physical IPA perpetration potentially limit the abilities of existing instruments to assess the risk of IPA among women. Most instruments were developed to assess the risk of discrete episodes of re-assault among men (see Hanson et al., 2007). As highlighted by Stark (2012), physical IPA is generally one behaviour of a much more complex pattern of psychological control and abuse. Evidence supporting this argument indicates that victim-reported coercive and controlling behaviours frequently co-occur with physical, sexual, and psychological forms of victimisation experienced by both sexes (e.g. Ansara & Hindin, 2010; Johnson et al., 2014; Tanha et al., 2010). Further, while most risk factors for physical IPV in male perpetrators are also applicable to females (see Spencer et al., 2016), it is unclear whether the same can be said for psychological and controlling forms of abuse.

The validity of police administered IPA risk assessment instruments for women

Relatively few instruments or guidelines have attempted to assess IPA risk in women, which is reflected in the exclusion of women from most validation studies until the

2010s. In their review of 15 years of research on the Spousal Assault Risk Assessment Guide (SARA; Kropp et al., 1995), Helmus and Bourgon (2011) found only one study that included women in their sample (Wong & Hisashima, 2008). Similarly, a systematic review of 39 studies that investigated IPA risk assessment instruments identified only three articles that included female perpetrators (Nicholls et al., 2013).

The validation studies of several instruments designed to assess the risk of FV (inclusive of IPA) have included female perpetrators and demonstrated promising results. For instance, the Domestic Violence Screening Inventory Revised (DVSI-R; Williams & Grant, 2006) is equally effective in predicting women's and mens FV recidivism (Williams, 2012; Williams & Grant, 2006). Similarly, McEwan et al. (2019) found that the Victoria Police Screening Assessment for Family Violence Risk (the VP-SAFvR; McEwan et al., 2019) – an instrument developed for use by frontline police officers – predicted future abuse of any family member by a woman with a moderate effect size (*AUC* = .65-.67) in both the development sample and a subsequent field trial (Spivak et al., 2020). However, to date the VP-SAFvR has not been validated to predict female IPA recidivism specifically, and while there is some evidence the DVSI-R's predictive validity also applies to female IPV (see Gerstenberger et al., 2019), it is not intended to be administered by police officers (Williams & Grant, 2006).

A small number of IPA-specific risk assessment instruments have also demonstrated some utility in predicting female perpetration but face similar questions of ecological validity in the context of front-line policing. The Danger Assessment (DA; Campbell, 1986; Campbell et al., 2009), an instrument designed to assess women's risk of severe or lethal victimisation by male partners, was found to effectively predict self-reported re-victimisation among female same-sex couples using 8 of the original 20 items in addition to 10 items specific to this relationship type (Glass et al., 2008). For heterosexual couples, the Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2004) has shown some promise in effectively predicting female IPA against male partners when items are not modified for gender differences (Hilton et al., 2014). Nonetheless, the assessments in both Glass et al. (2008) and Hilton et al.'s (2014) studies were undertaken by researchers using self-selected and incarcerated samples respectively so their results cannot be assumed to transfer to a policing context.

Other instruments developed explicitly for use by front-line officers and purported to be suitable for use with women who perpetrate IPA have not yet been evaluated in this context. The Lethality Screen (Messing et al., 2014) – a shortened version of the DA (Campbell, 1986; Campbell et al., 2009) – has only been investigated in female victim/male perpetrator relationships (Messing et al., 2014; Messing et al., 2017). Preliminary research conducted by the instrument's authors showed it was highly sensitive (92-93%) to near-lethal and severe violence in male perpetrator/female victim dyads, as well as IPA that was defined more broadly (e.g. any incident of verbal or psychological abuse; Messing et al., 2014; Messing et al., 2017). However, the same research showed low specificity for all forms of IPA, reflecting a high false-positive rate. Similarly, the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER; Kropp et al., 2005, 2010) was developed to allow police to assess the ongoing risk of IPA by both men and women but has no published validation studies among female IPA perpetrators.

Aims of IPA/FV risk assessment

While sharing the same broad goals of effective identification of higher risk IPA/FV cases, the more specific purpose and structure of IPA/FV risk assessment tools varies depending on the intended context of use (Gondolf, 2012, p. 171). As outlined by Medina-Ariza et al. (2016), instruments differ in how they define the harm or violence that is being assessed, whether they have an overt focus on helping to allocate resources most effectively, and whether they are embedded within a specific broader response or are intended to be used across different systems. They also vary in ways that affect their feasiblity of use in different contexts (Graham et al., 2021). Instruments vary in the number of risk factors included, how much specialist training they require, and whether they assess risk actuarially (i.e. using a predictive algorithm) or via structured professional judgement (Medina-Ariza et al., 2016; van der Put et al., 2019).

These are important differences in aims, rationale, and feasibility that have real impacts on how different IPA/FV risk assessment tools are used. For instance, the VP-SAFvR was created to help police triage FV cases, including IPA, based on the likelihood of future police reports of any FV within the same family unit, even where physical violence is not apparent (McEwan et al., 2019). It is intended to be used as part of a broader system where those 'screened in' receive subsequent comprehensive assessment for risk of severe harm (Spivak et al., 2020). Conversely, the Lethality Screen was created to assist frontline police officers in estimating the risk of future severe IPA at point of first contact (Messing et al., 2014) and is frequently used within a broader Lethality Assessment Program that facilitates a risk-informed collaborative police and social service intervention (Messing et al., 2015). The B-SAFER is different again, providing a nuanced and detailed assessment of physical violence risk and is intended to more closely guide risk management through assessment of dynamic risk factors using a structured professional judgement approach (Kropp et al., 2005, 2010). Hence, while assessing the risk of IPA is within the purview of all three instruments, and they are all intended to be scored by frontline police officers, they have important differences in the nature of the risk being assessed, their intended use, and their rationale.

The present study

Despite police departments increasingly adopting risk assessments there is a dearth of research investigating their utility among female IPA perpetrators who appear to commit different types of IPA (e.g. psychological abuse) at different rates and levels of severity to males. This study addresses these gaps by assessing the predictive validity of three police administered risk assessment instruments: an actuarial tool developed for a high volume outcome; an actuarial tool intended for a low base-rate/high severity outcome, and; a structured professional judgement (SPJ) instrument intended to assess the risk of physical violence. Each instrument will be used to prospectively assess the risk of IPA recidivism where women were identified as the perpetrators of IPA towards a current or former partner.

The research uses a subsample of women screened with the VP-SAFvR from Spivak et al. (2020), as well as two previously unpublished samples in which the B-SAFER was administered, or a modified version of the Lethality Screen was scored based on available

data. Furthermore, each instrument is validated against both a universal measure of IPA recidivism and an outcome reflecting the specific intended purpose of each instrument. Given this investigation is one of only a handful to feature women identified as perpetrators of IPA, the relationship between specific risk factors and IPA recidivism is also examined.

Method

The research was conducted across two police divisions in metropolitan Melbourne, the capital of the state of Victoria, Australia (population approx. 4.9 million people; Australian Bureau of Statistics, 2022). In Victoria, a family violence incident (FVI) is recorded whenever police attend an incident between two family members (including intimate partners) that involves physical, sexual, emotional, coercively controlling, or psychological abuse, regardless of whether the incident results in charges (Victoria Police, 2014). Additional information on Victoria Police's family violence recording procedures is available elsewhere (see McEwan et al., 2019; Spivak et al., 2020).

Samples

VP-SAFvR sample

The VP-SAFvR sample consisted of 410 women identified as female IPA perpetrators[1] extracted from a larger study by Spivak et al. (2020) of 3,963 unique dyads linked to an FVI identified between September 2016 and June 2017 and followed for an average of 14.88 months (SD = 2.70). Cases were excluded if they involved a male perpetrator (n = 3,100) or the gender of the perpetrator was unknown (n = 4), did not involve current or former intimate partners (n = 446), the relationship status of the dyad was unknown (n = 2), or where there was an irreconcilable difference in dyad characteristics (n = 1).

B-SAFER sample

As part of a larger field experiment of police IPA risk assessment and intervention, the B-SAFER was administered by specialist trained police (supported by a forensic psychologist) to individuals only after they had been determined to be at an elevated risk of FV on the VP-SAFvR. The B-SAFER sample comprised 60 assessments from cases involving women perpetrating IPA against their current or former intimate partners. A total of 863 intimate partner dyads were rated on the B-SAFER. Of these, 786 were excluded for involving male perpetrators and female victims, three for involving male same-sex partners, and another 14 were removed due to missing data (over 20% of risk factors in each case). The B-SAFER sample was followed for an average of 13.40 months (SD = 2.85).

Modified Lethality Screen sample

The Modified Lethality Screen sample consisted of 229 female IPA perpetrators, derived from the same sample as the VP-SAFvR. Cases were excluded if they involved a male perpetrator (n = 3,100) or the gender of the perpetrator was unknown (n = 4), did not involve current or former intimate partners (n = 446), if the VP-SAFvR was not completed in full (making scoring of equivalent Lethality Screen items impossible; n = 149), where over 20% of data was missing (n = 32), the relationship status of the dyad was unknown (n

= 2), or due to unreliable data ($n = 1$). The Modified Lethality Screen sample was followed for an average of 14.74 months (SD = 2.75). The demographic characteristics of each sample are presented in Table 1.

Measures

Victoria Police Screening Assessment for Family Violence Risk (VP-SAFvR; McEwan et al., 2019).

The VP-SAFvR is an actuarial screening tool developed to identify which FV cases are at increased risk of any future police reported family violence. It is intended as the first step in a tiered system of risk assessment and management tailored to the victim's safety, with those 'screened in' going on to receive further specialist FV police risk assessment and management (beyond standard responses to crime and referral to family violence support agencies, which happens in all cases). Hence, the VP-SAFvR comprises 14 questions assessing both the victim and perpetrator's personal circumstances, criminal history, and relationship risk factors. Questions receive a 'not scored', 'no' (scored 0), or 'yes' (scored 1 or 2) response. Numerical scores are summed to obtain a total score, with a threshold score of 4 being repeatedly demonstrated as maximising correct classification of future FV (McEwan et al., 2019; Sheed et al., 2023; Spivak et al., 2020). Items 10–14 are factors related to criminal history and are scored automatically from records from the Victoria Police Law Enforcement Assistance Program (LEAP), the database that tracks all prior police contacts.

Increases in VP-SAFvR total score are associated with incremental increases in the 12-month incidence of future police-reported family violence, with a moderate ability to discriminate between cases with and without this outcome (*AUC* = .66; McEwan et al., 2019; Spivak et al., 2020). A similar effect size was reported for intimate partner abuse (*AUC* = .64-.65; McEwan et al., 2019; Spivak et al., 2020) and family violence involving female perpetrators (*AUC* = .67; Spivak et al., 2020). In practice, the VP-SAFvR is administered in conjunction with additional risk factors that police officers can use to inform whether

Table 1. Demographic Characteristics of Female IPA Perpetrators in VP-SAFvR, B-SAFER and Modified Lethality Screen Samples.

	VP-SAFvR Sample ($N = 410$)		B-SAFER Sample ($N = 60$)		Modified Lethality Screen Sample ($N = 229$)	
Age (years)	*M* (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range
Perpetrator	34.84 (9.92)▲	14–70	33.60 (8.68)	20–67	35.32 (9.35)▼	16–69
Victim	37.96 (11.43)▲	14–82	37.55 (10.55)	20–67	38.47 (10.91)▼	17–70
Relationship Type	*n* (%)		*n* (%)		*n* (%)	
Married	125 (30.5)		6 (10)		61 (26.6)	
De-Facto◆	93 (22.7)		13 (21.7)		52 (22.7)	
Former Intimate	124 (30.3)		35 (58)		80 (34.9)	
Dating	46 (11.2)		5 (8.3)		21 (9.2)	
Same-Sex	22 (5.4)		1 (1.7)		15 (6.6)	

Note: ▲:$N = 406$; ▼:N(perpetrators) = 227, N(victims) = 226; ◆a de-facto relationship is a type of common law relationship legally recognised in Australia that is defined as two people who are not married, but who live together or have lived together as a couple in a genuine domestic arrangement in which their lives are shared (Victoria Legal Aid, 2022). De-facto couples have the same legal rights and obligations as married couples in Australian law; at the time the data was collected, Victoria Police recorded same-sex couples without a specific relationship status, meaning the specific relationship status of these couples at the time of the FVI is unknown, although they could not legally be married in Australia at that time.

they wish to override a VP-SAFvR total score of less than 4 to prompt further specialist police risk assessment.

Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER; Kropp et al., 2005, 2010)

The B-SAFER is a 15-item SPJ guideline designed to assist in the assessment and management of perpetrators of IPA based on the risk of future spousal assault (defined by the authors as actual, attempted, or threatened physical harm perpetrated by a man or woman against someone with whom he or she has, or has had, an intimate, sexual relationship). Risk items are divided equally across three domains (intimate partner violence, psychosocial adjustment, and victim vulnerability), and are rated relative to the most recent four weeks (current) or any time before that (past) on a 3-point scale (no/absent, possibly/partially present, and yes/present). For research purposes, it is common to assign a numerical score of 0, 1, or 2 to these ratings to calculate a total score (Storey & Strand, 2013). The tool allows professional judgement to be used in determining how these factors are weighted and combined, and whether additional factors require consideration (Storey & Strand, 2013). Police make an overall judgement about level of case prioritisation (*low/standard*, *moderate/elevated* or *high/urgent* risk) as well as risk judgements for acute violence risk, long-term violence risk, and lethal violence risk. Risk management strategies based on the assessment are also recommended and documented. The psychometric properties of the B-SAFER have been published (Au et al., 2008; Serie et al., 2017; Thijssen & de Ruiter, 2011), including its ability to discriminate between male recidivists and non-recidivists with a moderate to strong effect (*AUC* = .69-.76; Au et al., 2008; Belfrage & Strand, 2008, 2012; Loinaz, 2014; Storey et al., 2014).

Lethality Screen (Messing et al., 2014)

The Lethality Screen is an 11-question screening tool, designed for use within the context of a wider criminal justice and social service response to IPA, known as the Lethality Assessment Program. The Lethality Screen is intended to be asked of the victim, with questions rated 'yes' (factor present), 'no' (factor absent), or 'not answered'. A victim is subsequently rated as being in *high danger* of future severe or lethal partner violence if they respond affirmatively to one or more of the first three Lethality Screen questions (the items most strongly associated with risk for homicide or near-homicide on the DA; Campbell et al., 2009), or if they respond 'yes' to four or more of the remaining eight questions. While a *high danger* rating has been observed as highly sensitive to future IPA and severe or near-lethal IPA, it possessed poor specificity, translating to high rates of false positives and an associated weak positive predictive power (Messing et al., 2017). Previous research has shown that women killed by male partners have higher rates of each Lethality Screen risk factor compared to case control samples of women abused but not killed (Campbell et al., 2007; Glass et al., 2008; Koziol-McLain et al., 2006; Wilson & Daly, 1993).

In the present study, the Lethality Screen items were scored retrospectively using information from the risk assessment conducted by Victoria Police (both the VP-SAFvR and additional risk factors; see above and Spivak et al., 2020). This was due to the police operational environment preventing simultaneous scoring and use of two separate risk assessment tools which could produce different results. Table 2 presents the corresponding

Table 2. VP-SAFvR Items modified to simulate the Lethality Screen.

Lethality Screen Items	Relevant VP-SAFvR and Additional Risk Factors
1. Has he/she ever used a weapon against you or threatened you with a weapon?	Has your partner threatened or assaulted you with a firearm or weapon?
2. Has he/she threatened to kill you or your children?	Has your partner threatened to harm or kill you? Has your partner ever harmed or threatened to harm the children?
3. Do you think he/she might try to kill you?	What is your level of fear of further violence?
4. Does he/she have a gun or can he/she get one easily?	Do you or your partner have access to firearms?
5. Has he/she ever tried to choke you?	Has your partner ever strangled or suffocated you, or attempted to do this?
6. Is he/she violently or constantly jealous or does he/she control most of your daily activities?	Does the respondent act in ways that are jealous or controlling of the AFM?
7. Have you left him/her or separated after living together or being married?	Has there been a recent separation or is separation imminent?
8. Is he/she unemployed?	Is the respondent unemployed?
9. Has he/she ever tried to kill himself/herself?	Has the respondent ever threatened or attempted suicide?
10. Do you have a child that he/she knows is not his/hers?	Are there children from a previous relationship living in the house?
11. Does he/she follow or spy on you or leave threatening messages?	Has the respondent been following/approaching or repeatedly contacting/harassing you when you have told them not to?

Note: AFM: affected family member (i.e. identified victim)

items from the Lethality Screen and risk factors assessed by Victoria Police. Although most items reflect identical or almost identical risk factors, Item 3 was unable to be entirely captured by risk factors scored by Victoria Police. The item was coded as present if victims responded that they were 'very fearful' of further violence. Analyses of the Modified Lethality Screen's predictive validity were therefore undertaken with and without this item to assess its influence. Lastly, classifications of *not high danger* and *high danger* on the Modified Lethality Screen are used in the present study to differentiate the sample's risk of engaging in IPA recidivism.

Outcomes used to validate each instrument

All three instruments were assessed on a common outcome to allow for a common performance benchmark. Given the dual priority of victim safety and perpetrator accountability of the police force administering the assessments in this study (see Victoria Police, 2022), the following outcome was used:

> General IPA recidivism: Whether the index perpetrator was identified as the primary aggressor in one or more subsequent FVIs in any form (i.e. coercive control or other non-physical abuse, physical, or severe FVIs) against the index victim during the follow-up period.

Reflecting the different rationale of each instrument and the availability of data, several different outcome variables were created to reflect the constructs that the instruments are intended to assess. First, the VP-SAFvR was additionally assessed against a general family or intimate partner abuse (FIPA) outcome where they could be either the perpetrator or victim, defined as:

> FIPA: Whether the index perpetrator was identified as the primary aggressor *or* victim in one or more subsequent FVIs with the index victim or a related child. This outcome was not mutually exclusive of general IPA recidivism.

The Modified Lethality Screen was additionally assessed against a severe subtype of IPA recidivism, defined as:

> Severe IPA recidivism: Whether the index perpetrator was identified as the primary aggressor in one or more subsequent FVIs against the index victim that resulted in a charge involving physical harm, sexual harm, stalking, or threats. The presence of charges was taken to indicate a heightened level of seriousness. This outcome was not mutually exclusive of general IPA recidivism.

Given the B-SAFER's specific purpose, physical IPA was coded in addition to general IPA recidivism. Due to data availability, physical IPA was measured between the index perpetrator and index victim *or* a different intimate partner:

> Physical IPA recidivism: Whether the index perpetrator was identified as the primary aggressor in one or more subsequent FVIs against the index victim or a different intimate partner which led to at least one charge for physical violence or involved police-reported physical violence (full list of charges of physical violence available on request).

Procedure

Ethical approval was obtained from Swinburne University Human Research Ethics Committee (#2016/114) and Victoria Police Human Research Ethics Committee (#224/16) to conduct this research. All data analysed were supplied to the research team without direct interaction with participants and were devoid of identifying information.

Between July 2016 and June 2017, a new risk assessment and management process was trialled across two Victoria Police divisions. The VP-SAFvR was completed for each FVI that came to the attention of police in these divisions, with elevated scores necessitating an escalation of the case to the divisional family violence team (FVT), which is a dedicated unit with specialist training and roles related to FV. Members of the FVT then completed the B-SAFER for IPA cases to determine whether additional risk assessment and management were warranted. Police officers were aware that they were completing both the VP-SAFvR and B-SAFER. Police were able to liaise with an on-site clinical or forensic psychologist to discuss B-SAFER assessments and management recommendations. Follow-up data were extracted by Victoria Police analysts in February 2018 for the B-SAFER and in May 2018 for the VP-SAFvR, blind to risk assessment items or scores.

Data analyses

Data analysis was conducted using IBM SPSS Statistics Versions 25–27 (IBM Corp., 2015) and R Studio (R Core Team, 2021). Forty-eight percent of B-SAFER cases had missing data on less than 20% of items, and item scores were imputed using the Expectation Maximisation (EM) method. EM is an iterative approach commonly used to find maximum-likelihood estimates for model parameters when data is complete or has missing data points (Gupta & Chen, 2010). The VP-SAFvR and Modified Lethality Screen had no missing items deleted or imputed.

Total scores for the VP-SAFvR and B-SAFER were calculated by summing the numerical scores of all items (including both past and current iterations for the B-SAFER). Eta (η) correlations assessed the strength of the relationship between B-SAFER total scores and the

risk judgement/case prioritisation assigned to each case. H^2 values greater than .26 are considered large (Lakens, 2013). As the Lethality Screen was not developed to provide a total score, its predictive and discriminant ability was assessed solely on risk classification.

The ability of each instruments' intended basis of assessment (e.g. score on the VP-SAFvR, case prioritisation judgement on the B-SAFER, and classification on the Modified Lethality Screen) to discriminate between recidivists and non-recidivists was assessed using ROC curves plotting sensitivity and 1-specificity (Douglas et al., 2005). Predictive validity was determined by examining positive and negative predictive values (PPVs and NPVs) at various thresholds. The ROC analysis provides a measure of a test's ability to discriminate between groups (i.e. recidivist, non-recidivist) by producing a curve shown on a graph that plots the proportion of true positives against false positives at each possible score (Douglas et al., 2005). The area under the ROC curve (the AUC) indicates the probability that a randomly selected recidivist would have a higher risk score on a given assessment than a randomly selected non-recidivist (Rice & Harris, 1995, 2005). AUC values range from 0 to 1, with .50 indicating discrimination no better than chance, and 1 indicative of perfect (100%) positive discrimination (Douglas et al., 2005). Guidelines for the interpretation of AUC values suggest that .56 represents a small discriminant effect, .64 a moderate effect, and .71 a large effect (Rice & Harris, 2005).

Further analyses were undertaken to examine the relationship between specific risk factors and IPA outcomes. Where instrument items were binary, chi-square tests were used to determine the association of each item with IPA outcomes across all three instruments. The Gamma (γ) statistic (Goodman & Kruskal, 1954) (ranging from – 1 to +1) was used for all individual B-SAFER items (past and current) with ordinal three-level item responses. A Gamma value of positive 1 indicates a perfect relationship while a value of 0 indicates no relationship. Risk ratios were used to provide an indication of the strength of the association between risk items and outcomes. Levels indicating any presence of an item were aggregated to calculate risk ratios among three-level items. The familywise error rate was controlled using the Benjamini and Hochberg (1995) procedure. The adjustment was made within each instrument and outcome so were calculated from a combination of chi-square and Gamma statistic's probability values among the B-SAFER analyses. It was not possible to examine interrater reliability in the three tools as concurrent scoring was not possible in the field trial environment in which data was collected.

Results

VP-SAFvR

Ninety-three out of the 410 female IPA perpetrators (22.68%) scored on the VP-SAFvR perpetrated general IPA during the follow-up period. Of these, the majority had a further FVI against the same victim once (52.69%), 30.11% had two subsequent FVIs, and 17.20% returned to police attention for three or more episodes of general IPA (total ranging from 1-17).

One hundred and forty-four female IPA perpetrators scored on the VP-SAFvR were identified in one or more FIPA events during the follow-up period. The FIPA outcome comprised the $n = 93$ (64.58%) general IPA recidivists, $n = 49$ (34.03%) who were identified

as a victim and not a perpetrator in a subsequent FVI, and the nature of the remaining two (1.39%) FVIs is unknown.

There was a significant difference in mean VP-SAFvR total scores between women who returned to police attention for general IPA recidivism ($M = 5.67$, $SD = 2.38$) and women who did not ($M = 4.36$, $SD = 2.61$; $t(408) = 4.31$[.71, 1.90], $p < .001$). Total VP-SAFvR scores ranged from 0 to 13 ($M = 4.66$, $Mdn = 5.00$, $SD = 2.62$) and were not normally distributed (Shapiro–Wilk test of normality, $p < .001$, skew = .28, $SE = .12$).

Discriminant and predictive validity

Table 3 depicts the predictive and discriminant abilities of the VP-SAFvR. The classification accuracy of the VP-SAFvR was assessed relative to the recommended threshold score of 4. Results showed that for both outcomes sensitivity was high, indicating that, in cases scoring four or above on the VP-SAFvR, 82% and 76% of women went on to perpetrate general IPA or were involved in an FIPA incident respectively. While specificity was moderate for both general IPA (.42%) and FIPA (.44%) outcomes, NPV values were high; around 89% of women scoring below the threshold did not perpetrate general IPA during the follow-up period, while 77% were not involved in an FIPA incident. The PPV values for both general IPA recidivism (.29) and FIPA (.42) were both higher than each outcome's base rate within the sample (22.68% and 35.12% respectively).

VP-SAFvR total scores were effective at discriminating between female perpetrators who did and did not engage in general IPA recidivism with a moderate and significant effect size ($AUC = .65$, [.59, .71], $p < .001$). The VP-SAFvR total scores also effectively discriminated between those who were or were not involved in FIPA and the effect size was small to moderate but significant ($AUC = .60$, [.55, .66], $p < .001$).

The association of individual VP-SAFvR risk items with general IPA recidivism and FIPA are presented in Supplemental Tables 1 and 2. Two items – the length of time FV had been occurring in the relationship (greater than one month) and a history of any FVI between the same index parties was significantly associated with both general IPA ($\chi^2 = 10.35$, adjusted $p < .01$, risk ratio = 1.93; $\chi^2 = 15.99$, adjusted $p < .01$, risk ratio = 2.21) and FIPA ($\chi^2 = 7.97$, adjusted $p < .05$, risk ratio = 1.52; $\chi^2 = 10.52$, adjusted $p < .05$, risk

Table 3. Discriminant and Predictive Validity of the VP-SAFvR, B-SAFER, and Modified Lethality Screen.

Instrument (*N*) and outcome	Recidivism (%)	AUC [95% CI]	Sensitivity	Specificity	PPV	NPV
VP-SAFvR ($N = 410$)						
General IPA	22.68	.65 [.59, .71]***	.82	.42	.29	.89
FIPA	35.12	.60 [.55, .66]***	.76	.44	.42	.77
B-SAFER (Total score; $N = 60$)						
General IPA	40.00	.54 [.40, .69]	–	–	–	–
Physical IPA	21.70	.65 [.48, .82]	–	–	–	–
B-SAFER (Risk judgement; $N = 60$)						
General IPA	40.00	.49 [.35, .63]	.50	.50	.40	.60
Physical IPA	21.70	.70 [.54, .85]*	.77	.57	.33	.90
Modified Lethality screen ($N = 229$)						
General IPA	28.40	.51 [.44, .57]	.32	.69	.29	.72
Severe IPA	2.60	.68 [.46, .90]	.67	.70	.06	.99

Note: VP-SAFvR Sensitivity, Specificity, PPV, and NPV were calculated using the recommended threshold score of 4; B-SAFER (Risk judgement) Sensitivity, Specificity, PPV, and NPV were calculated using a threshold of Low versus Moderate/High risk classification; Modified Lethality screen AUC, Sensitivity, Specificity, PPV, and NPV were calculated in relation to High Danger (vs Not High Danger) classification; FIPA: family or intimate partner abuse; IPA: intimate partner abuse; PPV: positive predictive value; NPV: negative predictive value; * $p < .05$; ** $p < .01$; *** $p < .001$

ratio = 1.57). In both instances, the risk ratio was larger for general IPA recidivism compared with that of the FIPA outcome. For example, dyads with a history of any FVI were 2.21 times as likely to have general IPA recidivism during the follow-up period relative to those without such a history, whereas individuals possessing the same risk factor were only 1.57 times as likely to have an FIPA event compared to those without a history of any FVIs. No other VP-SAFvR items were significantly associated with the recidivism outcomes. The VP-SAFvR assessment includes an override procedure whereby responding officers can overrule a score of less than 4 if they believe the incident warrants specialist review. In total, 5 (1.22%) cases were overridden (median score of 2). No further analyses were undertaken on these cases due to the small sample size.

B-SAFER

Thirteen of the 60 female perpetrators (21.70%) assessed using the B-SAFER perpetrated physical IPA recidivism during the follow-up period. Twenty-four (40%) were identified as a perpetrator of general IPA recidivism. The majority of general IPA recidivists perpetrated non-physical FVIs.[2]

Thirty women (50%) were assessed by police as being in the *Low* case prioritisation category, 21 (35%) in the *Moderate* group, and nine (15%) in the *High* category. Kendall's test of rank correlation indicates there was a moderate association between the police risk judgements and B-SAFER total scores ($\tau_b = .29$, $z = 2.75$, $p < .01$), indicating that 19% of the variance in risk judgements made by police can be explained by variation in B-SAFER total scores (see Gilpin, 1993). The distribution of B-SAFER risk judgements against the B-SAFER total scores are depicted graphically in Supplemental Figure 2. The Shapiro–Wilk test of normality indicated B-SAFER total scores were approximately normally distributed ($p > .05$, raw total *mean* = 28.43, *median* = 29, *SD* = 8.25, *skew* = -.06, *SE* = .31).

Discriminant and predictive validity

As shown in Table 3 neither the B-SAFER total scores nor police risk judgements were effective at discriminating between women who did and did not perpetrate general IPA recidivism. However, when assessed against its intended outcome, physical IPA, the B-SAFER total scores had a moderate but non-significant discriminant ability while police judgements accurately categorised recidivists and non-recidivists 70% of the time with a large and significant effect size (*AUC* = .70 [.54, .85], $p < .05$).

Overall, police risk judgements correctly identified 77% of the women who went on to have physical IPA recidivism as being at moderate or high risk. The NPV was also high, with 90% of women classified as low risk not going on to perpetrate physical IPA in the follow-up period. However, specificity was low at 57%, which translates to a false-positive rate of 43% (i.e. 43% of women judged high risk did not recidivate). The PPV of 33% for physical IPA was notably higher than the 21.70% base rate of physical IPA recidivism for the sample. In other words, the rate at which the B-SAFER judgements correctly categorised moderate/high risk of future physical IPA was nearly 1.5 times the physical IPA reoffending rate.

Measures of sensitivity and specificity for the B-SAFER risk judgements on measures of general IPA recidivism showed that women who were classified as moderate or high risk

who did or did not go on to recidivate were correctly identified approximately 50% of the time (*sensitivity* = .50, *specificity* = .50). The PPV for B-SAFER risk judgements on the same outcome indicates a classification of moderate or high risk was equal to the general IPA recidivism base rate (recidivism rate = 40%, *PPV* = .40).

With regards to the B-SAFER's past and current risk factors, the B-SAFER current 'Health problems' item was significantly associated with general IPA recidivism ($\gamma^2 = 0.61$, adjusted $p < .05$, *risk ratio* = 2.63). One item – current 'Relationship problems' – was significantly associated with physical IPA recidivism, ($\gamma^2 = 1.00$, adjusted $p < .05$). Notably, the gamma coefficient indicated a perfect relationship – all women who returned to police attention for physical IPA scored 'yes' on this item. However, it was not possible to calculate a risk ratio for this item given the absence of any false negatives. No other individual B-SAFER items were significantly associated with the recidivism outcomes; the results are presented in Supplemental Tables 3 and 4.

Modified Lethality Screen

Sixty-five out of the 229 female IPA perpetrators (28.40%) scored on the Modified Lethality Screen perpetrated general IPA recidivism in the follow-up period. Of these, six women (2.6% of the entire sample) engaged in severe IPA recidivism. No homicides were recorded in the sample during the follow-up period. Considering the low base rate of severe IPA recidivism, results pertaining to this outcome should be interpreted with caution.

Female IPA perpetrators were classified as being of *high danger* of seriously re-victimising their partner in 72 cases (31.4%), while the remaining 157 women (68.6%) were coded as being *not high danger*. Of those screened as *high danger*, 65 (90.3%) scored 'yes' on one or more of the first three questions so were automatically rated *high danger*.

Discriminant and predictive validity

The discriminant and predictive validity of the Modified Lethality Screen for general and severe IPA recidivism is shown in Table 3. The specificity value of the Modified Lethality Screen on the general IPA recidivism outcome was .69, indicating 69% of women who did not subsequently perpetrate general IPA were classified as *not high danger*. The NPV value of .72 reflects that 72% of the women classified as *not high danger* did not engage in general IPA recidivism during the follow-up period. The sensitivity (.32) and PPV (.29) values were relatively lower and respectively indicate the Modified Lethality Screen incorrectly categorised 68% of those who subsequently perpetrated general IPA as *not high danger* while at the same time correctly predicting 29% of female IPA offenders were high risk of general IPA, which was approximately equivalent to base rate of general IPA in the Modified Lethality Screen sample (28.40%).

The performance of the Modified Lethality Screen was markedly stronger on its intended outcome, severe IPA. The sensitivity (.67), specificity (.70), and NPV (.99) values together indicate the Modified Lethality Screen had improved performance categorising female IPA perpetrators as *high danger* when they committed severe IPA recidivism and *not high danger* when they did not compared to the instrument's performance on the general IPA outcome. Notably, the Modified Lethality Screen produced a PPV of .06 (i.e. 6%); while very low, this is substantially improved prediction over the base rate of severe IPA recidivism (2.6%).

The Modified Lethality Screen risk rating was not effective at discriminating between women who did and did not perpetrate general IPA recidivism (*AUC* = .51 [.44, .57], *p* > .05). A moderate but non-significant effect was observed for severe IPA recidivism (*AUC* = .68 [.46, .90], *p* > .05), likely due to the infrequency the outcome. Re-running the predictive analyses without Item 3 (*Do you think he/she might try to kill you?*) saw no improvement in the Modified Lethality Screen risk judgements at predicting general or severe IPA recidivism. Measures of discriminant and predictive ability were approximately the same when considering risk judgements without the inclusion of Item 3 for both the general IPA recidivism and severe IPA recidivism outcomes (available on request). No individual Modified Lethality Screen items were significantly associated with either recidivism outcome; the results are presented in Supplemental Tables 5 and 6.

Discussion

This study assessed the predictive validity of three police-employed risk assessment tools – the VP-SAFvR, the B-SAFER, and the Lethality Screen (modified version) – in a sample of women who had contact with Australian police officers for perpetrating incidents of IPA. The B-SAFER and VP-SAFvR demonstrated good predictive and discriminant validity in female IPA perpetrators for their associated outcomes (physical IPA, general IPA recidivism and FIPA, respectively). The Modified Lethality Screen did not effectively predict future general IPA among these women, and while it did improve identification of those with severe IPA recidivism substantially over the base rate, 33% of women with severe IPA outcomes were assessed as *not high danger*.

VP-SAFvR results

The VP-SAFvR had reasonable utility in discriminating between women with and without general IPA and FIPA outcomes. These findings are consistent with those of a recent validation study of the VP-SAFvR which observed moderate effect sizes for FIPA irrespective of sex and among female perpetrators of FV (Spivak et al., 2020). Although the sample used in the current study was derived from the broader FV sample reported by Spivak et al. (2020), its use among the specific sub-group of female perpetrators of IPA had not yet been investigated. The tool was found to have a small to moderate but significant effect size for this group on measures of FIPA, providing only limited support for the authors' contention that the VP-SAFvR can be used to assess the risk of future FV broadly for both men and women.

The patterns of high sensitivity, lower specificity, low PPV, and high NPV observed among the sample of perpetrators assessed with the VP-SAFvR parallel the classification accuracy identified by Spivak et al. (2020) for subsequent FIPA among males and females combined, as well as FV recidivism among female perpetrators only. The performance of the VP-SAFvR, when considered alongside the results of Spivak et al.'s (2020) study illustrates that instruments designed for application with broad legislated definitions of family violence can produce adequate predictive and discriminant utility among specific presentations and sub-populations.

Given both the apparent performance stability of the VP-SAFvR and its widespread use in Victoria, a useful next step would be to establish reoffending norms for scores on the

instrument. In line with Hanson et al. (2017a; 2017b), this might take the form of several standardised risk categories which reference recidivism rates in the community to estimate the rate per 100 individuals who are expected to reoffend. In practical terms, this approach rapidly communicates an individual's risk of recidivism to frontline officers who can then intervene accordingly. The creation of standardised risk categories is beyond the scope of the current study and remains a task for future research.

B-SAFER results

The current results indicate that the B-SAFER risk judgement may be effective in predicting physical IPA recidivism in women, supporting its creators' contention that the tool can be used to assess risk of ongoing physical IPA in perpetrators of all genders. These findings align with those found in male samples showing a link between the B-SAFER total scores, risk judgements, and physical measures of IPA recidivism (Storey et al., 2014; Svalin et al., 2018).

The B-SAFER total score nor risk judgement performed better than chance at discriminating between female perpetrators who did or did not perpetrate general IPA recidivism against the same intimate partner. These findings contrast with previous research in male samples suggesting that the B-SAFER is predictive of both psychological and physical IPA recidivism (de Ruiter et al., 2008; Soeiro & Almeida, 2010; as cited in Strand & Storey, 2019), including police contacts for IPA assault, unlawful threats, harassment/stalking, violations of court orders, or sexual assault (Loinaz, 2014; Storey et al., 2014; Svalin et al., 2018). The B-SAFER's performance for general IPA recidivism appears to be driven by the fact that the majority of general IPA recidivists in the B-SAFER sample engaged in non-physical IPA, therefore the poor results on the general IPA recidivism outcome likely represent the instrument's inability to identify non-physical IPA recidivism in the current female sample. At least one previous study has found that female perpetrators tend to have lower scores on individual B-SAFER items compared with males (Storey & Strand, 2013). It may be that differences in B-SAFER item scores between males and females affect the prediction of non-physical but not physical recidivism. It is worth reiterating here that the general IPA recidivism outcome only included IPA that occurred between the index perpetrator and victim so does not perfectly capture the intent of the BSAFER.

Modified Lethality Screen results

Finally, the Modified Lethality Screen classification was found to be a relatively unreliable predictor of both general IPA and severe IPA recidivism in female IPA perpetrators. This finding lies in contrast to those of the development and validation research with male perpetrators conducted by the instrument's authors (Messing et al., 2014; Messing et al., 2017). For severe IPA recidivism, which approximated the target outcome for the Lethality Screen, the sensitivity and positive predictive values were substantially lower than those observed during the instrument's validation for severe violence (93% and 21% respectively, Messing et al., 2017). The low positive predictive power is partly due to the very low base rate of severe IPA recidivism in the sample (2.6%) and is actually a marked improvement over chance prediction. Conversely, the specificity and negative predictive power were markedly higher than those of Messing et al. (2017) for the

same outcomes (21% and 93% respectively; again the high NPP is because most women did not engage in severe IPA recidivism). In other words, the Modified Lethality Screen correctly classified a greater proportion of women as *not high danger* but misclassified a higher proportion of women as *high danger*, relative to the instrument's validation study of male perpetrators committing severe violence (Messing et al., 2017).

These findings should be interpreted cautiously given both the low base rate of severe IPA recidivism and the use of a modified version of the tool. The definition of severe IPA in the present study as FVIs involving criminal charges likely meant some severe FVIs, such as those involving insufficient evidence to proceed with charges or where the incident escaped police detection, were missed. In the validation study of the Lethality Screen, Messing et al. (2017) spoke directly to female victims and observed higher rates of near lethal (11.48%) and severe violence (18.80%). It is likely that rates of near lethal and severe violence would be lower among male victims of female perpetrated IPA, but men may also be less likely to report such violence to police. Regardless, the presence of charges provided the best assessment of severe IPA from the data available in this study and the absence of more sensitive measures of severity (e.g. emergency department presentations of victims at the time of the incident) remains a limitation. Furthermore, although the VP-SAFvR was partly designed to be asked of victims, the use of modified items does not reflect the true context for which the Lethality Screen was developed, and this represents a significant limitation of this element of this study.

Variability in IPA recidivism statistics in this study

The variability in recidivism rates reported in these analyses is partly due to sampling differences. Specifically, the B-SAFER sample represented a restricted group that required prior categorisation at a certain risk level using the VP-SAFvR before the guideline was used. Hence, it is reasonable to assume individuals with a higher probability of recidivism were selected to the exclusion of those with a lower probability. Further, the use of samples coming to the attention of specialised FVTs (to which referral was only made if the case was deemed to be at increased risk and/or in need of additional consideration; Victoria Police, 2014) may explain why the rates of recidivism captured across all three of the study samples were higher than the female recidivism rates noted in prior research (e.g. Henning et al., 2009; Ménard et al., 2009; Renauer & Henning, 2005).

Additionally, the B-SAFER and VP-SAFvR samples included IPA recidivism against the same victim only (i.e. general IPA recidivism) and two separate outcomes with broader definitions: the former was also assessed against physical IPA recidivism, which could include the same or different victims and the latter featured an FIPA outcome where the index perpetrator could be either the primary aggressor or victim in an FVI with the index victim or a related child. Expanding the definition of IPA captures more instances of the target behaviour and may even speak to an individual's greater propensity for violent behaviour generally, itself a risk marker for recidivism (Babcock et al., 2003). At the same time, expanding the definition of IPA recidivism to include other victims will increase the base-rate of offending, thereby overestimating the risk posed to the index victim. This again draws the focus to the intended outcomes of risk assessment, underscoring that consideration needs to be given to the outcomes used to validate or re-validate an instrument.

Despite high recidivism rates being found for IPA in general, low rates were observed for physical IPA and severe violence, based on the presence of charges. This is particularly problematic for assessments made using the Modified Lethality Screen, which was designed to predict severe/lethal violence. Although a positive result for the people involved, the absence of indicators of severe IPA at follow-up makes it difficult to validly identify risk markers from a research perspective (Storey & Hart, 2014). Low base rates could also reflect that most women simply do not engage in serious or lethal violence to the extent that men do (Straus, 2011), which would suggest that the Lethality Screen may not be an appropriate classification tool for women identified as IPA perpetrators in the general population.

Implications

The results of this research highlight the importance of the intended applications and outcomes of tools designed to assess the risk of future IPA. Of the instruments evaluated, the strongest results were found among those that were developed on representative samples or specifically for the prediction of more common forms of IPA. On one hand, the results of the VP-SAFvR illustrate the utility of developing risk assessment instruments to match a broad definition of IPA that constitutes both physical violence and psychological abuse. On the other, the B-SAFER's results demonstrate how a more careful and targeted approach to risk assessment can be used within a tiered policing intervention to assess high-risk female perpetrators' risk of physical violence with reasonable accuracy.

The present study therefore underscores how these tools could serve different functions within a broader response to IPA. An actuarial instrument like the VP-SAFvR can be applied by frontline officers to assess the risk of future IPA on a large scale, rapidly categorising those who are low-risk for future incidents of physical *or* psychological abuse within the same dyad or family unit whilst referring high-risk cases for further assessment and service linkage. Therefore, it compliments a dual strategy of victim safety and perpetrator accountability in the context of a high-volume outcome. Conversely, the B-SAFER in the current study required additional resources (i.e. a specialist team and support from a clinician) so would be impractical for police officers to administer for every FVI call-out. Nevertheless, when applied to perpetrators already assessed as being at an elevated risk for IPA, the B-SAFER can provide a sound assessment of the risk of physical IPA to any victim. Hence, on the current results the B-SAFER may be more appropriate within a perpetrator accountability framework where there is an elevated risk of violence.

The poor predictive performance of the Modified Lethality Screen – a modified version of a tool developed on male samples for the prediction of severe/lethal violence – illustrates the pitfalls of predicting low base rate outcomes and suggests the need for further research to determine whether markers of severe IPA are consistent across men and women. This last point has been considered by other authors, who have highlighted how risk factors for IPA recidivism may differ between men and women (Henning et al., 2009), not only in content but the extent to which they are relevant to risk (De Vogel & Nicholls, 2016). It also reinforces the importance of validating and modifying risk assessment instruments for use across distinct populations.

On the other hand, this research highlights the utility of certain global factors for predicting IPA recidivism. That is, the VP-SAFvR and B-SAFER were designed for use with both

sexes and have demonstrated here such tools can effectively predict differing forms of IPA involving women. The discriminant abilities of the VP-SAFvR when predicting general IPA recidivism (*AUC* = .65, [.59, .71]) or FIPA (*AUC* = .60, [.55, .66]) and the B-SAFER risk judgement prediction of physical IPA (*AUC* = .70, [.54, .85]) are consistent with those of male perpetrators assessed with the VP-SAFvR on measures of FIPA in both the parent sample (*AUC* = .66; Spivak et al., 2020) and developmental study (*AUC* = .66, [.65, .67]; McEwan et al., 2019). Both the base rate of FIPA recidivism and calibration indices amongst male perpetrators in McEwan et al. (2019) Spivak et al.'s (2020) studies (base rate = 30.9%, *PPV* = .39, *NPV* = .81; base rate = 24.1%, *PPV* = .30, *NPV* = .87 respectively) are comparable to those of the VP-SAFvR and B-SAFER in the abovementioned samples of the present study, suggesting a consistent level of calibration across male and female samples on measures of FIPA and physical IPA.

The results also echo those of another general tool, the DVSI-R, when predicting female to male IPV (*AUC* = .63, [.60, .66]) in a study by Gerstenberger et al. (2019). However, the DVSI-R possessed poor measures of discrimination when used to predict IPV among same-sex female couples (*AUC* = .56, [.47, .65]) despite same-sex females having significantly greater odds of re-arrest. Hence, the predictive utility of IPA risk factors might only generalise between men and women when the latter are abusive in heterosexual relationships. Furthermore, several non-gendered factors, such as items indicative of prior police contact for FV or health problems, were shown to meaningfully predict differing IPA outcomes within both the VP-SAFvR and B-SAFER samples. Nonetheless, the predictive ability of individual factors was not consistent across outcomes nor instruments and thereby directs further research to investigate whether and when specific factors predict future IPA, and to determine whether these differ across sexes.

The high rates of false positives observed among the severe IPA outcomes underline the importance of aligning police departments' aims for risk assessment and the instrument's rationale. The nature of low base rate outcomes means that instruments such as the Lethality Screen will inevitably misidentify a disproportionate number of individuals as high risk who will not go on to perpetrate severe and near lethal FV, even with further fine-tuning of their predictive and discriminant abilities (see Rosen, 1954; Trood et al., 2023). Indeed, the Lethality Screen was developed to ensure that people thought to be at increased risk of lethal/near lethal FV recidivism are connected to an advocate for safety planning and linkage to services, accepting that a large proportion of individuals who will not go on to perpetrate near lethal recidivism will also receive the intervention (Messing et al., 2017). It may not be appropriate for policing responses to perceived risk of lethal/near lethal violence to be directed with such imprecision, not only because of resource constraints but also because of the highly punitive and restrictive nature of many police interventions intended to prevent severe harm. The VP-SAFvR and B-SAFER results in the present study suggest that it may be possible to implement a more graduated police response to perpetrators with greater precision if a more general IPA outcome is assessed. This is consistent with the focussed deterrence strategy described by Sechrist and Weil (2018), though using a more complex method of categorising offenders than just police history of IPA. Of course, such a response would need to be in conjunction with social and health services where appropriate (Spivak et al., 2020).

In addition to using tools that consider the risks associated with IPA for both male and female perpetrators, the current study also supports the use of SPJ when assessing female

IPA perpetration. Although the assessment of risk and identification of treatment needs among male and female offenders may be subject to bias (Coontz et al., 1994; Skeem et al., 2005), the results of the current study suggest police discretion in assessing the risk of female IPA perpetration may be particularly useful. Considering that the B-SAFER total scores only explained about 19% of the variance in the B-SAFER risk judgements, its likely the SPJ approach allowed police more flexibility in how they applied the risk factors and what risk judgements they made. This appears to be supported by existing research indicating that police consider additional factors alongside established risk markers when assessing women's risk of IPA recidivism (Storey & Strand, 2012, 2013, 2017), suggesting opportunities to incorporate professional judgement may be particularly important in cases involving female perpetrators (see De Vogel & Nicholls, 2016). Interestingly, the current study found the B-SAFER case prioritisation judgement to be particularly useful in the prediction of future physical IPA. This might suggest that other factors external to the summative component of the tool may have contributed to police officers' risk judgements or alternatively that they are giving stronger weight to certain items. Limited evidence from the broader literature indicates that risk judgements derived from SPJ instruments tend to outperform total scores from the same instruments in the prediction of general violence (see Guy et al., 2015). The present findings appear to support this trend and suggest professional judgements informed by scored instruments may be especially useful in the assessment of physical IPA in high-risk women.

We are hesitant to suggest any practical implications associated with the current results for the Lethality Screen, given the method by which it was administered in this study. However, we note that in the Australian context, the quite specific purpose of the Lethality Screen may mean that it is not suitable as a risk assessment instrument. Australian jurisdictions do not focus on physical violence when legally defining IPA, and in many jurisdictions police and legal responses relate to all types of FV, regardless of whether it involves physical abuse or an intimate partner.

Limitations and other considerations

Our results are limited by a low base rate severe IPA recidivism. This concern has been identified in forensic research on female populations more broadly (e.g. Helmus & Bourgon, 2011; Nicholls et al., 2013), and future research would ideally include measures with higher fidelity to recidivism base rates (e.g. multiple sources of administrative data and victim reports). This is an important point considering the socio-political context of IPA and female perpetration, and the factors that might influence identified recidivism rates. For instance, many IPA victims – especially men – do not report their victimisation experiences to the police (Archer, 2000; Cho & Wilke, 2010; Felson & Paré, 2005; MacQueen & Norris, 2016), often because of social stigma and gender stereotypes (Walker et al., 2018), which threatens the accuracy of female IPA perpetration and recidivism measurement when using formal data.

In a similar vein, women may not be as readily identified as perpetrators of IPA as men. Social constructions of gender and relationships perhaps skew the perceptions of police officers recording FVI cases, resulting in an ongoing male-perpetrator/female-victim bias (Henning et al., 2009; Walker et al., 2018) and the inaccurate recording of victims and perpetrators in ambiguous FVIs. During the collection of the VP-SAFvR and B-SAFER data, for example, police members described this being an issue when the perpetrator and victim

roles were unclear, although it was not possible to control for or measure to what extent this occurred during the research period. Given there is evidence a significant proportion of couples also engage in bidirectional violence (i.e. where both parties are the perpetrator/victim, or have a history of swapping victim and perpetrator roles; Langhinrichsen-Rohling et al., 2012), police recording bias may be particularly problematic and result in low numbers of women accurately identified as perpetrators. Ergo, we recommend future research on frontline risk assessments additionally collect information that might reveal misidentification. For example, the instruments themselves could include a rating of the officer's confidence they have identified the primary aggressor which may provide insights into when and how the misidentification of a perpetrator occurred.

The results of this study could equally be interpreted in the context of women being misidentified as perpetrators of abuse when they are the victims (Women's Legal Service Victoria, 2015). While few studies have examined rates of misidentification at FVIs, the available evidence very tentatively suggests that up to 10% of female FV victims in Victoria are misidentified as respondents (i.e. aggressors) in police applications for Family Violence Intervention Orders (Ulbrick & Jago, 2018). If a large proportion of the samples examined here were misidentified as the perpetrator at index, then we might expect a drop in both the predictive and discriminant abilities of the instruments examined. Yet the VP-SAFvR and B-SAFER risk judgements in the current research produced reasonably strong measures of discriminant and predictive ability on measures of general and physical IPA recidivism among women respectively. Further, the VP-SAFvR's discriminant and predictive ability parallels the same instrument's performance among the broader population where most perpetrators are male (Spivak et al., 2023). Hence, the VP-SAFvR and B-SAFER appear to have withstood any possible drop in performance associated with misidentification of women as perpetrators and might only improve alongside a reduction in this phenomenon. Overall, the conclusions of this research are preliminary and require replication in larger samples of women who perpetrate IPA using a combination of police data and other measures.

Lastly, the present study was limited by its comparison of three different instruments on three different samples. Due to data availability, the Modified Lethality Screen could not be applied to every member of the VP-SAFvR sample while the B-SAFER was administered by police in a subset assessed as being at a higher recidivism risk. Ideally, all three instruments would be compared on the same sample of cases to remove the influence of sampling variation. This would also allow for a sharper contrast of the performance of included instruments against their differing aims. This limitation is a consequence of the field environment of this research, which also comes with considerable strengths, not least indicating the feasibility of use in practice of the VP-SAFvR and B-SAFER.

Conclusion

While women are increasingly coming to police attention for IPA perpetration, it is unclear whether existing risk assessment instruments are valid for this group. This study evaluated three risk assessment instruments currently used by police internationally in a sample of Australian women identified as IPA perpetrators. The Modified Lethality Screen had mixed performance on measures of discrimination and calibration. However, this result may have been at least partly due to a low base rate of severe IPA recidivism among the Modified Lethality Screen sample. Consistent with their developmental rationale, the

VP-SAFvR and the SPJ component of the B-SAFER possessed predictive and discriminant results for their respective outcomes of general IPA recidivism and further FIPA, and physical IPA. The results presented here suggest female-perpetrated IPA recidivism can be assessed with reasonable accuracy when they are validated on representative samples and applied within their intended developmental context. Complicating this is the need to ensure that the risk assessment tool chosen is not only valid, but relevant to their legislative landscape, policies, and service systems for IPA victims and perpetrators.

Notes

1. For the sake of brevity, all participants in the current research are referred to as 'female IPA perpetrators', defined as women identified by Victoria Police as perpetrating abuse during an FVI involving a current or former intimate partner.
2. The exact number of physical or non-physical general IPA recidivists could lead to the identification of individual cases so has been omitted.

Acknowledgements

The authors would like to acknowledge the help received from members of Victoria Police's Family Violence Command; assisting Victoria Police members from North West Metro Region; Victoria Police Corporate Statistics staff; and staff from Forensicare in conducting this research. The first author was supported by an Australian Government Research Training Program Scholarship in conducting this research and preparing this publication. The research was funded by grants from the Macedon Ranges and North Western Melbourne Medicare Local, honouring its legacy of commitment to the long-term health of the North western Melbourne and Macedon Ranges community. Correspondence may be addressed to Michael D. Trood, Centre for Forensic Behavioural Science, Swinburne University of Technology, 582 Heidelberg Road, Alphington, Victoria, 3078, Australia. Email: mtrood@swin.edu.au

Data transparency statement

The data reported in this manuscript relating to the VP-SAFvR forms part of a larger dataset collected and published by Spivak et al. (2020). The research questions addressed here are substantially different from those examined Spivak et al. (2020).

Disclosure statement

Troy McEwan is a co-author of the VPSAFvR but receives no financial benefit from its publication or use. No other potential conflicts of interest were reported by the authors.

Funding

This work was supported by an Australian Government Research Training Program Scholarship and was funded by grants from the Macedon Ranges and North Western Melbourne Medicare Local.

Geolocation

This research uses data collected by Victoria Police in Melbourne, Victoria, Australia (37.8136° S, 144.9631° E).

Data availability statement

The data associated with this research contains personal and sensitive information so is not openly accessible.

References

Ansara, D. L., & Hindin, M. J. (2010). Exploring gender differences in the patterns of intimate partner violence in Canada: A latent class approach. *Journal of Epidemiology & Community Health*, *64*(10), 849–854. https://doi.org/10.1136/jech.2009.095208

Archer, J. (2000). Sex differences in aggression between heterosexual partners: A meta-analytic review. *Psychological Bulletin*, *126*(5), 651–680. https://doi.org/10.1037/0033-2909.126.5.651

Au, A., Cheung, G., Kropp, R., Yuk-chung, C., Lam, G. L. T., & Sung, P. (2008). A preliminary validation of the brief spousal assault form for the evaluation of risk (B-SAFER) in Hong Kong. *Journal of Family Violence*, *23*(8), 727–735. https://doi.org/10.1007/s10896-008-9198-z

Australian Bureau of Statistics. (2022, July 26). *Regional Population: Statistics about the population for Australia's capital cities and regions*. Retrieved September 15, 2022, from https://www.abs.gov.au/statistics/people/population/regional-population/latest-release#victoria

Babcock, J. C., Armenti, N., Cannon, C., Lauve-Moon, K., Buttell, F., Ferreira, R., Cantos, A., Hamel, J., Kelly, D., Jordan, C., Lehmann, P., Leisring, P. A., Murphy, C., O'Leary, K. D., Bannon, S., Salis, K. L., & Solano, I. (2016). Domestic violence perpetrator programs: A proposal for evidence-based standards in the United States. *Partner Abuse*, *7*(4), 355–460. https://doi.org/10.1891/1946-6560.7.4.355

Babcock, J. C., Miller, S. A., & Siard, C. (2003). Toward a typology of abusive women: Differences between partner-only and generally violent women in the use of violence. *Psychology of Women Quarterly*, *27*(2), 153–161. https://doi.org/10.1111/1471-6402.00095

Bates, E. A., & Graham-Kevan, N. (2016). Is the presence of control related to help-seeking behavior? A test of Johnson's assumptions regarding sex differences and the role of control in intimate partner violence. *Partner Abuse*, *7*(1), 3–25. https://doi.org/10.1891/1946-6560.7.1.3

Belfrage, H., & Strand, S. (2008). Structured spousal violence risk assessment: Combining risk factors and victim vulnerability factors. *International Journal of Forensic Mental Health*, *7*(1), 39–46. https://doi.org/10.1080/14999013.2008.9914402

Belfrage, H., & Strand, S. (2012). Measuring the outcome of structured spousal violence risk assessments using the B-SAFER: Risk in relation to recidivism and intervention. *Behavioral Sciences & the Law*, *30*(4), 420–430. https://doi.org/10.1002/bsl.2019

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Campbell, J. C. (1986). Nursing assessment for risk of homicide with battered women. *Advances in Nursing Science*, *8*(4), 36–51. https://doi.org/10.1097/00012272-198607000-00006

Campbell, J. C., Glass, N., Sharps, P. W., Laughon, K., & Bloom, T. (2007). Intimate partner homicide: Review and implications of research and policy. *Trauma, Violence, & Abuse*, *8*(3), 246–269. https://doi.org/10.1177/1524838007303505

Campbell, J. C., Webster, D. W., & Glass, N. (2009). The Danger Assessment: Validation of a lethality risk assessment instrument for intimate partner femicide. *Journal of Interpersonal Violence*, *24*(4), 653–674. https://doi.org/10.1177/0886260508317180

Cantos, A. L., Neidig, P. H., & O'Leary, K. D. (1994). Injuries of women and men in a treatment program for domestic violence. *Journal of Family Violence*, *9*(2), 113–124. https://doi.org/10.1007/BF01531958

Carney, M., & Barner, J. (2012). Prevalence of partner abuse: Rates of emotional abuse and control. *Partner Abuse*, *3*(3), 286–335. https://doi.org/10.1891/1946-6560.3.3.286

Cho, H., & Wilke, D. J. (2010). Gender differences in the nature of the intimate partner violence and effects of perpetrator arrest on revictimization. *Journal of Family Violence*, *25*(4), 393–400. https://doi.org/10.1007/s10896-009-9300-1

Coontz, P. D., Lidz, C. W., & Mulvey, E. P. (1994). Gender and the assessment of dangerousness in the psychiatric emergency room. *International Journal of Law and Psychiatry*, *17*(4), 369–376. https://doi.org/10.1016/0160-2527(94)90013-2

Crime Statistics Agency. (2022, September). *Family incidents*. Victoria State Government. https://www.crimestatistics.vic.gov.au/crime-statistics/latest-victorian-crime-data/family-incidents-2

de Ruiter, C., de Jong, E., Reus, M., & Thijssen, J. (2008). *Risk assessment in perpetrators of relational violence: A comparison of RISc and B-SAFER*. Netherlands Institute of Mental Health and Addiction.

De Vogel, V., & Nicholls, T. L. (2016). Gender matters: An introduction to the special issues on women and girls. *International Journal of Forensic Mental Health*, *15*(1), 1–25. https://doi.org/10.1080/14999013.2016.1141439

Douglas, K. S., Guy, L. S., Reeves, K. A., & Weir, J. (2005). *HCR-20 Violence risk assessment scheme: Overview and annotated bibliography*. https://repository.escholarship.umassmed.edu/handle/20.500.14038/45237

Ehrensaft, M. K., Moffitt, T. E., & Caspi, A. (2004). Clinically abusive relationships in an unselected birth cohort: Men's and women's participation and developmental antecedents. *Journal of Abnormal Psychology*, *113*(2), 258–270. https://doi.org/10.1037/0021-843X.113.2.258

Felson, R. B., & Paré, P.-P. (2005). The reporting of domestic violence and sexual assault by nonstrangers to the police. *Journal of Marriage and Family*, *67*(3), 597–610. https://doi.org/10.1111/j.1741-3737.2005.00156.x

Fiebert, M. S. (2014). References examining assaults by women on their spouses or male partners: An updated annotated bibliography. *Sexuality and Culture: An Interdisciplinary Quarterly*, *18*(2), 405–467. https://doi.org/10.1007/s12119-013-9194-1

Gerstenberger, C., Stansfield, R., & Williams, K. R. (2019). Intimate partner violence in same-sex relationships: An analysis of risk and rearrest. *Criminal Justice and Behavior*, *46*(11), 1515–1527. https://doi.org/10.1177/0093854819871984

Gerstenberger, C. B., & Williams, K. R. (2013). Gender and intimate partner violence: Does dual arrest reveal gender symmetry or asymmetry? *Journal of Interpersonal Violence*, *28*(8), 1561–1578. https://doi.org/10.1177/0886260512468325

Gilpin, A. R. (1993). Table for conversion of Kendall's Tau to Spearman's Rho within the context of measures of magnitude of effect for meta-analysis. *Educational and Psychological Measurement*, *53*(1), 87–92. https://doi.org/10.1177/0013164493053001007

Glass, N., Perrin, N., Hanson, G., Bloom, T., Gardner, E., & Campbell, J. C. (2008). Risk for reassault in abusive female same-sex relationships. *American Journal of Public Health*, *98*(6), 1021–1027. https://doi.org/10.2105/AJPH.2007.117770

Gondolf, E. W. (2012). *The future of batterer programs: Reassessing evidence-based practice*. Northeastern University Press.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*(268), 732–764. https://doi.org/10.1080/01621459.1954.10501231

Graham, L. M., Sahay, K. M., Rizo, C. F., Messing, J. T., & Macy, R. J. (2021). The validity and reliability of available intimate partner homicide and reassault risk assessment tools: A systematic review. *Trauma, Violence, & Abuse*, *22*(1), 18–40. https://doi.org/10.1177/1524838018821952

Gupta, M., & Chen, Y. (2010). Theory and use of the EM algorithm. *Foundations and Trends® in Signal Processing*, *4*(3), 223–296. https://doi.org/10.1561/2000000034

Guy, L. S., Douglas, K. S., & Hart, S. D. (2015). Risk assessment and communication. In B. L. Cutler, & P. A. Zapf (Eds.), *APA handbook of forensic psychology. Vol. 1, Individual and situational influences in criminal and civil contexts* (pp. 35–86). American Psychological Association.

Hanson, R. K., Babchishin, K. M., Helmus, L. M., Thornton, D., & Phenix, A. (2017a). Communicating the results of criterion referenced prediction measures: Risk categories for the Static-99R and Static-2002R sexual offender risk assessment tools. *Psychological Assessment*, *29*(5), 582–597. https://doi.org/10.1037/pas0000371

Hanson, R. K., Bourgon, G., McGrath, R. J., Kroner, D., D'Amora, D. A., Thomas, S. S., & Tavarez, L. P. (2017b). *A five-level risk and needs system: Maximizing assessment results in corrections through the development of a common language*. The Council of State Governments.

Hanson, R. K., Helmus, L., & Bourgon, G. (2007). *The validity of risk assessments for intimate partner violence: A meta-analysis 2007-07*. Public Safety Canada. https://www.securitepublique.gc.ca/cnt/rsrcs/pblctns/ntmt-prtnr-vlnce/ntmt-prtnr-vlnce-eng.pdf

Helmus, L., & Bourgon, G. (2011). Taking stock of 15 years of research on the Spousal Assault Risk Assessment Guide (SARA): A critical review. *International Journal of Forensic Mental Health*, *10* (1), 64–75. https://doi.org/10.1080/14999013.2010.551709

Henning, K., Martinsson, R., & Holdford, R. (2009). Gender differences in risk factors for intimate partner violence recidivism. *Journal of Aggression, Maltreatment & Trauma*, *18*(6), 623–645. https://doi.org/10.1080/10926770903103248

Henning, K., Renauer, B., & Holdford, R. (2006). Victim or offender? Heterogeneity among women arrested for intimate partner violence. *Journal of Family Violence*, *21*(6), 351–368. https://doi.org/10.1007/s10896-006-9032-4

Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., & Lines, K. J. (2004). A brief 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

Hilton, N. Z., Popham, S., Lang, C., & Harris, G. T. (2014). Preliminary validation of the ODARA for female intimate partner violence offenders. *Partner Abuse*, *5*(2), 189–203. https://doi.org/10.1891/1946-6560.5.2.189

Hines, D. A., & Douglas, E. M. (2010). Intimate terrorism by women towards men: Does it exist? *Journal of Aggression, Conflict and Peace Research*, *2*(3), 36–56. https://doi.org/10.5042/jacpr.2010.0335

Hines, D. A., & Douglas, E. M. (2019). An empirical test of Johnson's typology of intimate partner violence in two samples of men. *Partner Abuse*, *10*(2), 180–198. https://doi.org/10.1891/1946-6560.10.2.180

IBM Corp. (2015). *IBM SPSS statistics for windows* (Versions 25.0: 27.0). [Software]. IBM Corp.

Johnson, M. P., Leone, J. M., & Xu, Y. (2014). Intimate terrorism and situational couple violence in general surveys: Ex-spouses required. *Violence Against Women*, *20*(2), 186–207. https://doi.org/10.1177/1077801214521324

Koziol-McLain, J., Webster, D., McFarlane, J., Block, C. R., Ulrich, Y., Glass, N., & Campbell, J. C. (2006). Risk factors for femicide-suicide in abusive relationships: Results from a multisite case control study. *Violence and Victims*, *21*(1), 3–21. https://doi.org/10.1891/0886-6708.21.1.3

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

Kropp, P. R., Hart, S. D., & Belfrage, H. (2010). *Brief spousal assault form for the evaluation of risk (B-SAFER), version 2: User manual*. ProActive ReSolutions.

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1995). *Manual for the spousal assault risk assessment guide* (2nd ed.). The British Columbia Institute Against Family Violence.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863–863. https://doi.org/10.3389/fpsyg.2013.00863

Langhinrichsen-Rohling, J., Misra, T. A., Selwyn, C., & Rohling, M. L. (2012). Rates of bidirectional versus unidirectional intimate partner violence across samples, sexual orientations, and race/ethnicities: A comprehensive review. *Partner Abuse*, *3*(2), 199–230. https://doi.org/10.1891/1946-6560.3.2.199

Loinaz, I. (2014). Typologies, risk and recidivism in partner-violent men with the B-SAFER: A pilot study. *Psychology, Crime & Law*, *20*(2), 183–198. https://doi.org/10.1080/1068316X.2013.770854

MacQueen, S., & Norris, P. A. (2016). Police awareness and involvement in cases of domestic and partner abuse. *Policing and Society*, *26*(1), 55–76. https://doi.org/10.1080/10439463.2014.922084

McEwan, T. E., Shea, D. E., & Ogloff, J. R. P. (2019). The development of the VP-SAFvR: An actuarial instrument for police triage of Australian family violence reports. *Criminal Justice and Behavior*, *46*(4), 590–607. https://doi.org/10.1177/0093854818806031

Medina-Ariza, J. J., Robinson, A., & Myhill, A. (2016). Cheaper, faster, better: Expectations and achievements in police risk assessment of domestic abuse. *Policing: A Journal of Policy and Practice*, *10*(4), 341–350. https://doi.org/10.1093/police/paw023

Ménard, K. S., Anderson, A. L., & Godboldt, S. M. (2009). Gender differences in intimate partner recidivism: A 5-year follow-up. *Criminal Justice and Behavior*, *36*(1), 61–76. https://doi.org/10.1177/0093854808325905

Messing, J. T., Campbell, J., Sullivan Wilson, J., Brown, S., & Patchell, B. (2017). The Lethality Screen: The predictive validity of an intimate partner violence risk assessment for use by first responders. *Journal of Interpersonal Violence*, *32*(2), 205–226. https://doi.org/10.1177/0886260515585540

Messing, J. T., Campbell, J., Sullivan Wilson, J., Brown, S., Patchell, B., & Shall, C. (2014). *Police departments' use of the Lethality Assessment Program: A quasi-experimental evaluation*. U. S. Department of Justice. https://www.ncjrs.gov/pdffiles1/nij/grants/247456.pdf.

Messing, J. T., Campbell, J., Webster, D. W., Brown, S., Patchell, B., & Wilson, J. W. (2015). The Oklahoma Lethality Assessment study: A quasi-experimental evaluation of the Lethality Assessment Program. *Social Service Review*, *89*(3), 499–530. https://doi.org/10.1086/683194

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

R Core Team. (2021). R: A language and environment for statistical computing (Version R 4.2.0). [Software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Renauer, B., & Henning, K. (2005). Investigating intersections between gender and intimate partner violence recidivism. *Journal of Offender Rehabilitation*, *41*(4), 99–124. https://doi.org/10.1300/J076v41n04_05

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, *63*(5), 737–748. https://doi.org/10.1037/0022-006X.63.5.737

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

Rosen, A. (1954). Detection of suicidal patients: An example of some limitations in the prediction of infrequent events. *Journal of Consulting Psychology*, *18*(6), 397–403. https://doi.org/10.1037/h0058579

Sechrist, S. M., & Weil, J. D. (2018). Assessing the impact of a focused deterrence strategy to combat intimate partner domestic violence. *Violence Against Women*, *24*(3), 243–265. https://doi.org/10.1177/1077801216687877

Serie, C. M. B., van Tilburg, C. A., van Dum, A., & de Ruiter, C. (2017). Spousal assaulters in outpatient mental health care: The relevance of structured risk assessment. *Journal of Interpersonal Violence*, *32*(11), 1658–1677. https://doi.org/10.1177/0886260515589932

Sheed, A., McEwan, T., Papalia, N., Spivak, B., & Simmons, M. (2023). Assessing risk of family violence by young people: Identifying recidivism base rates and the validity of the VP-SAFvR for youth. *Criminal Justice and Behavior*, *50*(8), 1079–1101. https://doi.org/10.1177/00938548231170799

Skeem, J., Schubert, C., Stowman, S., Beeson, S., & Mulvey, E. (2005). Gender and risk assessment accuracy: Underestimating women's violence potential. *Law and Human Behavior*, *29*(2), 173–186. https://doi.org/10.1007/s10979-005-3401-z

Soeiro, C., & Almeida, I. (2010). *Spousal assault risk assessment and police intervention: Application to the Portuguese population* [Conference presentation]. 10th Annual Meeting of the International Association of Forensic Mental health Services, Vancouver, British Columbia, Canada.

Spencer, C., Cafferky, B., & Stith, S. M. (2016). Gender differences in risk markers for perpetration of physical partner violence: Results from a meta-analytic review. *Journal of Family Violence*, *31*(8), 981–984. https://doi.org/10.1007/s10896-016-9860-9

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

Spivak, B., Trood, M. D., & McEwan, T. E. (2023). *Validation of an actuarial instrument to predict repeat intimate partner violence for frontline police: Calibration, discrimination and classification performance of the VP-SAFvR* [Manuscript submitted for publication]. Centre for Forensic Behavioural Sciences, Swinburne University and Victorian Institute of Forensic Mental Health.

Stark, E. (2012). Looking beyond domestic violence: Policing coercive control. *Journal of Police Crisis Negotiations*, *12*(2), 199–217. https://doi.org/10.1080/15332586.2012.725016

Storey, J. E., & Hart, S. D. (2014). An examination of the Danger Assessment as a victim-based risk assessment instrument for lethal intimate partner violence. *Journal of Threat Assessment and Management*, *1*(1), 199–217. https://doi.org/10.1037/tam0000002

Storey, J. E., Kropp, P. R., Hart, S. D., Belfrage, H., & Strand, S. (2014). Assessment and management of risk for intimate partner violence by police officers using the Brief Spousal Assault Form for the Evaluation of Risk. *Criminal Justice and Behavior*, *41*(2), 256–271. https://doi.org/10.1177/0093854813503960

Storey, J. E., & Strand, S. (2012). The characteristics and violence risk management of women arrested by the police for intimate partner violence. *European Journal of Criminology*, *9*(6), 636–651. https://doi.org/10.1177/1477370812453403

Storey, J. E., & Strand, S. (2013). Assessing violence risk among female IPV perpetrators: An examination of the B-SAFER. *Journal of Aggression, Maltreatment & Trauma*, *22*(9), 964–980. https://doi.org/10.1080/10926771.2013.835015

Storey, J. E., & Strand, S. (2017). The influence of victim vulnerability and gender on police officers' assessment of intimate partner violence risk. *Journal of Family Violence*, *32*(1), 125–134. https://doi.org/10.1007/s10896-016-9905-0

Strand, S. J. M., & Storey, J. E. (2019). Intimate partner violence in urban, rural, and remote areas: An investigation of offense severity and risk factors. *Violence Against Women*, *25*(2), 188–207. https://doi.org/10.1177/1077801218766611

Straus, M. A. (2004). Prevalence of violence against dating partners by male and female university students worldwide. *Violence Against Women*, *10*(7), 790–811. https://doi.org/10.1177/1077801204265552

Straus, M. A. (2011). Gender symmetry and mutuality in perpetration of clinical-level partner violence: Empirical evidence and implications for prevention and treatment. *Aggression and Violent Behavior*, *16*(4), 279–288. https://doi.org/10.1016/j.avb.2011.04.010

Svalin, K., Mellgren, C., Levander, M. T., & Levander, S. (2018). Police employees' violence risk assessments: The predictive validity of the B-SAFER and the significance of protective actions. *International Journal of Law and Psychiatry*, *56*, 71–79. https://doi.org/10.1016/j.ijlp.2017.09.001

Tanha, M., Beck, C. J. A., Figueredo, A. J., & Raghavan, C. (2010). Sex differences in intimate partner violence and the use of coercive control as a motivational factor for intimate partner violence. *Journal of Interpersonal Violence*, *25*(10), 1836–1854. https://doi.org/10.1177/0886260509354501

Thijssen, J., & de Ruiter, C. (2011). Identifying subtypes of spousal assaulters using the B-SAFER. *Journal of Interpersonal Violence*, *26*(7), 1307–1321. https://doi.org/10.1177/0886260510369129

Trood, M. D., Spivak, B., Ogloff, J. R. P., & McEwan, T. E. (2023). *The limits of predicting near lethal and lethal family and intimate partner violence* [Manuscript in preparation]. Centre for Forensic Behavioural Sciences, Swinburne University and Victorian Institute of Forensic Mental Health.

Ulbrick, M., & Jago, M. (2018). *Policy paper 1 "Officer she's psychotic and I need protection": Police misidentification of the 'primary aggressor' in family violence incidents in Victoria*. Women's Legal Service Victoria. https://www.womenslegal.org.au/files/file/WLSV%20Policy%20Brief%201%20MisID%20July%202018.pdf

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

Victoria Legal Aid. (2022, August 3). *De-facto and same sex relationships*. Victoria State Government. https://www.legalaid.vic.gov.au/de-facto-relationships

Victoria Police. (2014). *Victoria Police code of practice for the investigation of family violence*. Victoria Police.

Victoria Police. (2022). The code of practice for the investigation of family violence – Edition 4 version 2. Victoria Police. https://www.police.vic.gov.au/sites/default/files/2022-11/Code-of-practice-for-the-investigation-of-family-violence-Edition-4-Version-2.pdf

Walker, J., Ashby, J., Gredecki, N., & Tarpey, E. (2018). Gender representations of female perpetrators of intimate partner violence. *Journal of Aggression, Conflict and Peace Research*, *10*(3), 170–180. https://doi.org/10.1108/JACPR-02-2017-0273

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

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

Wilson, M., & Daly, M. (1993). Spousal homicide risk and estrangement. *Violence and Victims*, *8*(1), 3–16. https://doi.org/10.1891/0886-6708.8.1.3

Women's Legal Service Victoria. (2015). *Improving the family violence legal system: Royal commission into family violence*. https://womenslegal.org.au/files/file/AMENDED%20AMENDED%20FINAL%20MAIN%20SUBMISSION.05.08.2015.pdf

Wong, T., & Hisashima, J. (2008). *Domestic violence exploratory study on the DVSI and SARA, State of Hawaii, 2003*-2007. Interagency Council on Intermediate Sanctions. http://icis.hawaii.gov/wp-content/uploads/2013/07/SARA-DVSI-Exploratory-Study-Oct-2008.pdf