No. 581 October 2019

Abstract | This study examines how accurately the Family Violence Risk Assessment Tool (FVRAT) predicts repeat domestic violence. The FVRAT is a 37-item tool used by police in the ACT to inform their responses to domestic violence.

This study examines a sample of 350 unique cases of violence involving current or former intimate partners between March and December 2017 in which police used the FVRAT. Repeat domestic violence was measured based on whether a subsequent report of domestic violence was made to police within six months.

The FVRAT is not a strong predictor of repeat domestic violence. However, an empirically refined version of the FVRAT consisting of 10 individually predictive items much more accurately predicts repeat domestic violence.

Predicting repeat domestic violence: Improving police risk assessment

Christopher Dowling and Anthony Morgan

Frontline police spend a substantial portion of their time responding to domestic violence (DV). It is estimated that police attend more than 260,000 DV incidents annually across Australia (Blumer 2016). Police spend between 30 minutes and two hours at the scene of each incident, and an additional one to two hours taking some kind of action in response (eg arrest, referral to services; Crime and Misconduct Commission 2005; Victorian Auditor-General 2009).

The estimated risk of further violence figures prominently in police decisions around how to respond to DV (Kane 2000, 1999). Police increasingly use DV risk assessment tools to standardise, improve and document these predictions. DV risk assessment tools comprise checklists of items reflecting incident, offender, victim and relationship characteristics that are thought to indicate a higher likelihood of repeat DV (Capaldi et al. 2012). Properly designed and empirically validated risk assessment tools can better channel limited police time and resources into the DV cases that most require it, enhance the short-term safety of victims, aid in the investigation of DV incidents by prompting police to collect information on a wider range of factors, and inform the activities of prosecutors and others (Backhouse & Toivonen 2018).

Several tools have been developed for police in Australia, including the Risk Assessment Screening Tool in Tasmania (Mason & Julian 2009), the DV Safety Action Tool in New South Wales (Ringland 2018), and the L17 form in Victoria (Millsteed & Coghlan 2016). International tools, including the Ontario Domestic Assault Risk Assessment (ODARA; Lauria et al. 2017) and the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER; McEwan, Bateson & Strand 2017), have also been examined with Australian samples, and there have been recent national efforts to consolidate and promote consistency in DV risk assessment across Australian jurisdictions (Backhouse & Toivonen 2018).

A substantial body of international research has examined the predictive validity of DV risk assessment tools. This research commonly measures predictive validity using the Area Under the Receiver Operating Characteristic (AUROC) curve score. The AUROC score represents the probability that a randomly selected case with repeat DV will receive a higher risk rating on a prediction tool than a randomly selected case where repeat DV did not occur (Graham et al. 2019; Messing & Thaller 2013). The ODARA has demonstrated the best predictive validity based on this measure, although many tools have returned at least moderate AUROC scores of between 0.60 and 0.75. These tools also rate comparatively well on other measures of predictive validity, including the accuracy with which they correctly classify cases where repeat DV will (sensitivity) and will not (specificity) occur, and the proportions of correctly classified cases across each risk rating (positive (PPV) and negative (NPV) predictive values). However, DV risk assessment tools have generally been shown to be better predictors of when repeat DV will not occur than when it will occur, as indicated by higher specificity and NPV scores.

Australian research has found that most of the risk items included on Tasmania's Risk Assessment Screening Tool and Victoria's L17 tool significantly predict repeat DV (Mason & Julian 2009; Millsteed & Coghlan 2016). Additionally, recent analysis of the ODARA (Lauria et al. 2017) and B-SAFER (McEwan, Bateson & Strand 2017) in Victoria returned AUROC scores of between 0.68 (physical abuse) and 0.72 (non-physical abuse) for the former, and 0.63 for the latter. However, a recent analysis of the DV Safety Action Tool using a sample of over 20,000 male and female DV victims in New South Wales, and examining physical and non-physical instances of repeat DV, found AUROCs of between 0.54 and 0.58, indicating this tool performs only slightly better than chance in predicting repeat DV (Ringland 2018).

Critically, few studies to date have controlled for police actions and other short-term responses (eg arrest, protection and other court orders, or health, legal and service responses) in validating these DV risk assessment tools. This is important given the evidence demonstrating that police responses, including apprehending and charging offenders, influence the likelihood of repeat offending (Dowling et al. 2018a; Mazerolle et al. 2018). These responses are implemented differently, and with different intensity, across cases based on their estimated levels of risk, with higher risk cases generally receiving more intensive responses (Belfrage et al. 2012; Belfrage & Strand 2012; Hilton, Harris & Rice 2007; Storey et al. 2014; Trujillo & Ross 2008). Unsurprisingly then, the number and intensity of these responses appears to alter baseline risk ratings throughout a validation study's follow-up period and influence the likelihood of repeat DV across cases with different risk ratings (Belfrage et al. 2012; Belfrage & Strand 2012; Hilton, Harris & Rice 2007; Storey et al. 2014). As such, failing to account for these responses can confound any analysis of a DV risk assessment tool's predictive validity.

Aim and method

In March 2017 Australian Capital Territory Policing began using the Family Violence Risk Assessment Tool (FVRAT). The FVRAT is a 37-item tool that officers are required to complete when responding to a reported incident of violence perpetrated by a current or former intimate partner. Items are weighted based on their professionally-judged importance to the estimation of risk, and reflect the severity of the most recent incident, characteristics of the relationship between the offender and victim, the couple's history of violence, and the offender's criminal history and mental health issues. These items are summed to produce a score reflecting the likelihood of repeat DV: low (0–13), medium (14–27) and high (28+). Officers can revise these ratings based on their own analysis of the information provided. Implementation of the FVRAT is part of ACT Policing's broader effort to meet the objectives of the ACT's inter-agency Family Violence Intervention Program, specifically the identification and protection of (particularly) at-risk victims. ACT Policing mandates that officers prioritise cases based on their risk rating, and proactively implement immediate measures to ensure the safety of victims in higher risk cases.

This study examined the predictive validity of the FVRAT. Its aim was to determine whether the FVRAT's risk ratings accurately predict repeat DV. It employed a relatively short follow-up period, focusing on the FVRAT's predictive validity within six months of a DV report. This is because the intent of the tool was to inform the immediate response of police to DV, and because recent Australian research shows the risk of repeat DV is highest shortly after a DV incident (Morgan, Boxall & Brown 2018). In light of prior research demonstrating the impact of police responses on DV (Dowling et al. 2018a), this study also examined whether controlling for these responses improves the FVRAT's ability to accurately predict repeat DV.

Sample and data collection

The sample consists of 350 reported incidents of violence involving current or former intimate partners in the Australian Capital Territory between March and December 2017. Nine cases were excluded from the sample because of missing data. Reported incidents are unique in that no offender or victim is represented across more than one report in the sample. Where an offender or victim came to the attention of police for DV more than once within the study period, their earliest reported incident was included in the sample.

The first wave of data collection involved manually coding information on DV cases from hard-copy FVRATs and case logs downloaded from the Police Real-Time Management Information System (PROMIS). This information included demographic details of each offender and victim, risk ratings for each item of the FVRAT and an overall score, and details of police and other short-term responses to DV reports in the following six months. Importantly, while officers using the FVRAT are free to revise actuarial risk ratings based on their professional judgement, it became apparent throughout this wave of data collection that this was not occurring.

The second wave of data collection involved extracting data from PROMIS on any further contact victims and offenders had with ACT Policing for DV in the six months following their index report. PROMIS event ID numbers for index DV reports were used to locate the records of any follow-up incidents involving the same victims and offenders.

Over half of offenders (63%) and victims (64%) were between the ages of 25 and 44 at the time of the index DV report, and less than a quarter were older than 55. Offenders were, on average, two years older (*M*=35.7, *SD*=10.6) than victims (*M*=33.7, *SD*=10.0). The vast majority of offenders were male (87%) and victims female (87%). Only four percent of incidents involved violence between current or former same-sex intimate partners. A small proportion of offenders and victims were Indigenous (6% and 5%, respectively) or members of culturally or linguistically diverse groups (12% of both). Equal proportions of cases received low (40%) and medium (38%) risk ratings on the FVRAT, while less than a quarter were rated as high risk (22%). Police arrested the offender in 39 percent of cases and charged the offender in 34 percent of cases, with 41 percent of cases resulting in at least one of these responses.

For the purposes of this study, a repeat DV incident was recorded if at least one additional report of DV was made to police within six months of the index report. If an offender was detained for any period of time following the index report, then the observation period commenced when they were released from custody. Reports were distinguished as DV based on a family violence flag and victim–offender relationship status recorded in PROMIS. Since protection order and bail conditions are applied more often to offenders in higher risk cases (Dowling et al. 2018b; Mazerolle et al. 2018), and it was not possible to include these responses as covariates, technical breaches not involving threatened or actual physical violence were excluded. Eleven percent of victims made at least one additional report of DV within the six month follow-up period. Two-thirds of these cases, or seven percent of the total sample, involved physical violence, as indicated by the presence of at least one offence classified under either Division 2 (acts intended to cause injury) or Division 3 (sexual assault and related offences) of the Australian and New Zealand Standard Offence Classification (Australian Bureau of Statistics 2011). Eighty-eight percent of repeat DV reports involved the offender who perpetrated DV in the index report.

Measures of predictive validity

Logistic regression was used to measure the association between FVRAT risk ratings and repeat DV. The AUROC score was used as the principal measure of the FVRAT's predictive validity. AUROC scores range from 0.50 (meaning the prediction tool is no better than chance at predicting an outcome of interest) to 1 (meaning the prediction tool perfectly predicts an outcome of interest). AUROC scores of 0.80 or above are commonly thought to indicate excellent accuracy, while scores of 0.70–0.79 indicate acceptable accuracy, and scores of 0.60–0.69 indicate moderate accuracy (Hosmer & Lemeshow 2004).

In addition to the AUROC, a number of more specific measures were used to assess the accuracy with which the FVRAT predicts repeat DV (see also Table 1):

- classification accuracy—the total proportion of correct ratings;
- sensitivity—the proportion of repeat DVs correctly predicted;
- specificity—the proportion of non-repeat DVs correctly predicted;
- positive predictive value—the proportion of correct high risk ratings; and
- negative predictive value—the proportion of correct low risk ratings.

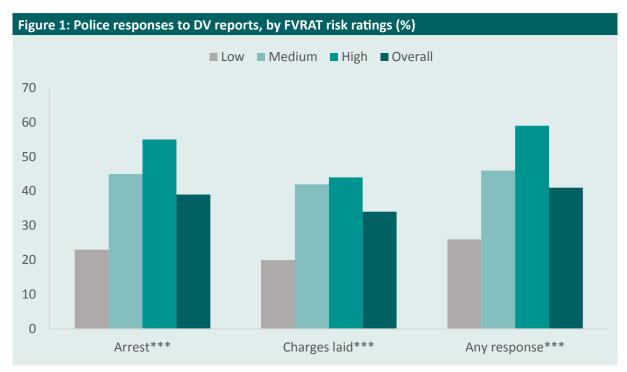| Table 1: Measures of predictive validity				
		Actual repeat DV?		
		Yes	**No**	
Predicted repeat DV?	**Yes**	True positive (tp)	False positive (fp)	PPV tp/(tp+fp)
	No	False negative (fn)	True negative (tn)	NPV tn/(fn+tn)
		Sensitivity tp/(tp+fn)	Specificity tn/(fp+tn)	Classification accuracy (tp+tn)/n

Because the FVRAT classifies cases into three risk ratings, these measures are calculated by comparing low-risk cases with medium- and high-risk cases. While risk assessment tools cannot be used as absolute predictors of repeat DV (high risk ≠ will experience repeat DV; low risk ≠ will not experience DV), these measures allow different variations of the FVRAT to be compared and the best fitting model to be determined.

Logistic regression models used initially to investigate the association between risk ratings and repeat DV include police responses to index DV reports as covariates. Individual measures of classification accuracy, sensitivity, specificity, PPVs and NPVs were calculated from predicted probabilities of repeat DV across cases at each risk rating, which were adjusted for covariates using marginal standardisation (Muller & MacLehose 2014). Covariate-adjusted AUROC scores, representing the weighted average of AUROC scores across cases with distinct covariate configurations (Janes & Pepe 2009, 2008), were also calculated.
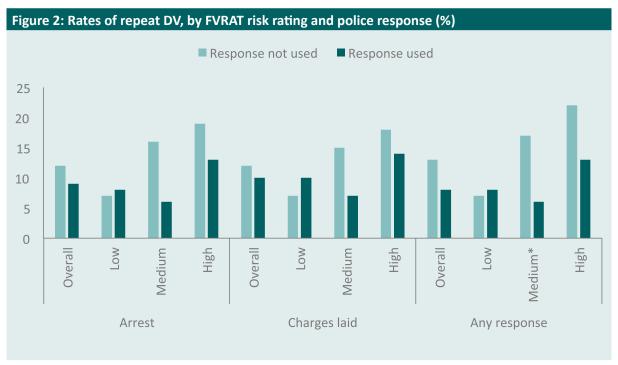
Results

Risk ratings, police responses and repeat DV

Police responses to index reports were more intensive for cases with higher risk ratings, a pattern which held for both arrests and charges when examined individually (Figure 1). For example, police arrested and/or charged the offender in 59 percent of high-risk cases, compared with 46 percent of medium-risk cases and 26 percent of low-risk cases (χ^2(5)=25.0, Cramér's V=0.26, p<0.001). Repeat DV rates were generally lower for cases where police arrested and/or charged the offender, both overall and among medium- and high-risk cases, although few of these differences were statistically significant (Figure 2). Meanwhile, rates of repeat DV were similar across low-risk cases that did or did not receive a police response.

Figure 1: Police responses to DV reports, by FVRAT risk ratings (%)

***statistically significant at $p<0.001$

Note: Arrest $\chi^2(5)=24.8$, Cramér's $V=0.26$. Charges laid $\chi^2(5)=18.4$, Cramér's $V=0.22$. Any response $\chi^2(5)=25.0$, Cramér's $V=0.26$

Source: ACT Policing 2018 [computer file]

Figure 2: Rates of repeat DV, by FVRAT risk rating and police response (%)
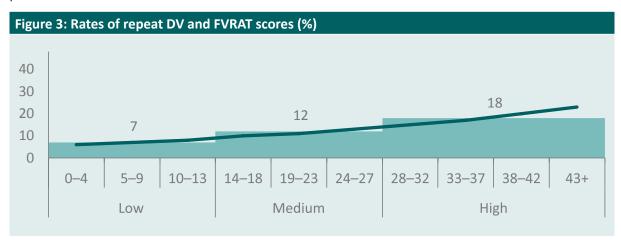

*statistically significant at $p<0.05$

Note: Any response for medium-risk DV cases $\chi^2(3)=3.8$, Cramér's $V=0.17$

Source: ACT Policing 2018 [computer file]

Predictive validity

Overall, DV cases that scored higher on the FVRAT, and received a higher risk rating, were slightly more likely to result in repeat DV, after adjusting for police responses (Figure 3). Eighteen percent of high-risk cases resulted in repeat DV, compared with 12 percent of medium-risk cases and seven percent of low-risk cases.

Figure 3: Rates of repeat DV and FVRAT scores (%)

Note: Due to the weighting of items on the FVRAT as either 2 or 3, an overall score of 1 is not possible. Rates of repeat DV are based on predicted probabilities, which take into account the police response to the index incident

Source: ACT Policing 2018 [computer file]

After controlling for police responses, the odds of repeat DV within the six-month follow-up period were significantly higher in cases rated as high risk than in cases rated as low risk (OR=2.96, CI=1.18–7.43; Table 2, Model 2). There was no significant difference in the odds of repeat DV between medium- and low-risk cases. Controlling for police responses resulted in a slight improvement to the FVRAT's AUROC score. However, even after controlling for police responses, the AUROC score indicates that the FVRAT is little better than chance at predicting repeat DV (AUROC score=0.60, CI=0.49–0.69). In other words, there is only a 60 percent probability that a randomly selected DV case where repeat DV occurred received a higher risk rating on the FVRAT than a randomly selected case where repeat DV did not occur.

The FVRAT correctly classified 67 percent of high- and low-risk DV cases. Of those cases where repeat DV occurred, 60 percent were correctly classified as high risk (sensitivity) and, of those cases where repeat DV did not occur, 68 percent were correctly classified as low risk (specificity). Additionally, of cases that received a high risk rating, 19 percent resulted in repeat DV (PPV). Among cases that received a low risk rating, 93 percent did not result in repeat DV (NPV).

Table 2: Logistic regressions predicting repeat DV from risk ratings, and AUROC scores

	FVRAT ratings (original)		FVRAT ratings (revised)	Empirical ratings	
	Model 1 OR (95% CI)	**Model 2** OR (95% CI)	**Model 3** OR (95% CI)	**Model 4** OR (95% CI)	**Model 5** OR (95% CI)
Risk ratings					
Medium (vs low)	1.60 (0.71–3.59)	1.82 (0.78–4.22)	3.72 (1.09–10+)*	3.07 (1.19–7.87)*	3.70 (1.37–9.93)**
High (vs low)	2.38 (0.1.01–5.44)*	2.96 (1.18–7.43)*	10+ (2.47–10+)**	10+ (3.96–10+)***	10+ (5.46–10+)***
Covariates					
Any response	–	0.51 (0.24–1.08)†	0.52 (0.26–1.07)†	–	0.36 (0.16–0.83)*
	Model 1 statistics	**Model 2** statistics	**Model 3** statistics	**Model 4** statistics	**Model 5** statistics
Constant	0.08 (0.05–0.16)***	0.09 (0.05–0.18)***	0.04 (0.01–0.15)***	0.04 (0.02–0.10)***	0.05 (0.02–0.12)***
Model χ^2 (df, n)	3.93 (2, 350)	6.39 (3, 350)	12.98 (3, 350)*	21.95 (2, 350)***	24.61 (3, 350)***
Nagelkerke R^2	–	–	0.07	0.12	0.16
AUROC score (95% CI)	0.59 (0.47–0.68)	0.60 (0.49–0.69)[a]	0.64 (0.55–0.72)[a]	0.70 (0.60–0.79)	0.73 (0.60–0.81)[a]

†$p<0.10$, *$p<0.05$, **$p<0.01$, ***$p<0.001$

a: Covariate-adjusted

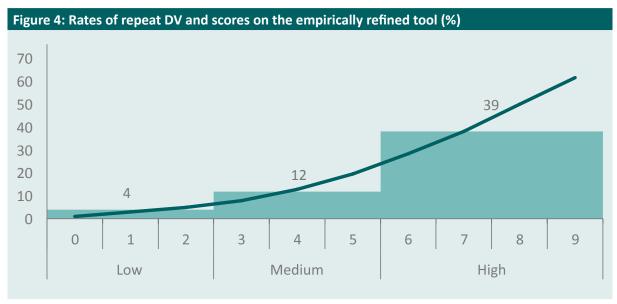Source: ACT Policing 2018 [computer file]

An empirical method of categorising continuous variables proposed by Barrio and colleagues (2017) was then used to identify the optimal cut-off points for the FVRAT's low, medium and high risk ratings. Starting with a population of potential cut-off points for a continuous predictor of a dichotomous outcome, this approach uses an evolutionary algorithm to successively identify gradually more refined generations of cut-off points with better AUROC scores than preceding generations until an optimal cut-off point is identified. This analysis was performed using the R package 'CatPredi' (Barrio et al. 2017), with police responses included as covariates. Revised FVRAT risk ratings based on this method (low 0–9=21%; medium 10–39=72%; high 40+=7%) returned only a marginally improved covariate-adjusted AUROC score of 0.64 (CI=0.55–0.72; Table 2, Model 3).

Additional analyses were undertaken to develop and test an empirically refined risk assessment tool using only the FVRAT items that were individually associated with a higher likelihood of repeat DV.

Controlling for police responses, 10 FVRAT items were positively associated with repeat DV. (Results are available from the authors on request.) These items are as follows:

- Offender has assaulted the victim during the most recent incident;
- Offender has injured or threatened to injure a family pet now or in the past;
- Victim was physically injured during the most recent incident;
- Victim and offender have recently separated after cohabiting;
- There is a pregnancy or new birth (less than 12 months) within the relationship;
- The offender has assaulted the victim in the past;
- The violence level in the relationship has escalated (violence includes property damage);
- Offender has committed offences of violence against any person in the past;
- Offender has breached a protection or any court order now or in the past;
- Offender has money problems at the moment.

Scores on these items were re-dichotomised (0=no, 1=yes) and summed for each case, returning a total score between zero and 10. Barrio and colleagues' (2017) empirical categorisation method was used to determine the optimal cut-off points for low, medium and high risk ratings on this empirically refined predictive tool. The results showed that cut-off points of 2.4 and 5.4 were optimal for categorising cases. These scores were rounded and used to classify cases into low (0–2=40%), medium (3–5=47%) and high (6+=11%) risk groups. Analysis of the cases in each of the risk categories created using this refined tool revealed a gradual and consistent increase in the covariate-adjusted rate of repeat DV (Figure 4).

Figure 4: Rates of repeat DV and scores on the empirically refined tool (%)

Note: No cases scored 10. Rates of repeat DV are based on predicted probabilities, which take into account the police response to the index incident

Source: ACT Policing 2018 [computer file]

There was a significant and positive association between police responses (both overall, and for arrest and charge) and empirical risk ratings, similar in magnitude to the association between these responses and the original FVRAT risk ratings ($p<0.001$ for all, Cramér's V=0.24–0.26). Additionally, the use of any response by police was significantly associated with a lower likelihood of repeat DV in cases classified as medium risk (17% vs 6%, $p<0.05$, Cramér's V=0.17) and high risk (53% vs 23%, $p<0.05$, Cramér's V=0.30), but not low risk (3% vs 5%, p=0.71).

Controlling for police responses, the odds of repeat DV within the six-month follow-up period were significantly higher in cases rated as medium risk (OR=3.70, CI=1.37–9.93) and high risk (OR=10+, CI=5.46–10+) than in cases rated as low risk on the empirically refined tool (Table 2, Model 5). There was a small improvement in the AUROC score for the empirical risk ratings once police responses were taken into account, from 0.70 (CI=0.60–0.79; Table 2, Model 4) to 0.73 (CI=0.60–0.81; Table 2, Model 5). This also represents a notable improvement in predictive validity over the original FVRAT risk ratings (AUROC=0.60; Table 2, Model 2). Based on the empirically refined tool, there is a 73 percent probability that a randomly selected case where repeat DV occurred received a higher risk rating than a randomly selected case where repeat DV did not occur.

Importantly, the AUROC for the three-category empirically refined tool was equal to or higher than the AUROC generated for versions of the tool using between two and eight risk categories. Given the impracticality of employing a large number of categories to guide police responses (which only marginally improves the AUROC) the three risk ratings were retained. (Results are available from the authors on request.)

Analysis of the sensitivity, specificity, PPV and NPV of the empirically refined tool also showed it significantly improved upon the original FVRAT. A comparison of medium- and low-risk DV cases indicates that half were correctly classified (50%). Of those cases where repeat DV occurred, 77 percent were correctly rated as medium risk (sensitivity), although only 48 percent of those where repeat DV did not occur were correctly rated as low risk (specificity). Predictions that repeat DV would occur (PPV=12%) or would not occur (NPV=96%) differed substantially in accuracy.

Further analysis indicates that over four-fifths of high-risk and low-risk DV cases were correctly classified (83%). Where repeat DV occurred, 73 percent were correctly classified as high risk (sensitivity), while 84 percent of cases where repeat DV did not occur were correctly rated as low risk (specificity). Additionally, the high risk category is better than the medium risk category at predicting repeat DV, relative to the low risk rating (PPV=39%).

Finally, items shown to be related to the risk of repeat violence (particularly fatal violence) were incorporated into the refined tool. These items included the involvement of alcohol and drugs, mental illness and strangulation. Adding these items, either separately or in combination, did not improve AUROC scores. (Results are available from the authors on request.)

Conclusion

How well does the FVRAT predict repeat domestic violence?

The findings of this study highlight the strengths and limitations of the FVRAT, along with potential avenues for its further development.

Finding: the FVRAT is not a strong overall predictor of further domestic violence.

The current FVRAT risk ratings are marginally better than chance at predicting further DV, as measured by repeat DV reported to police within a six-month follow-up period. The covariate-adjusted AUROC score (0.60) falls just within the range generally considered moderate, although it is lower than the scores of most other Australian and international tools (Lauria et al. 2017; Mason & Julian 2009; Messing & Thaller 2013; Millsteed & Coghlan 2016). Furthermore, empirically optimised risk ratings based on the full FVRAT produced an AUROC score only marginally better (0.64).

Finding: the FVRAT is better at predicting when repeat DV will not occur than when it will occur.

Examining only those DV cases classified as high or low risk using the FVRAT indicates that, while two-thirds were correctly classified, this was principally due to the correct classification of low-risk cases (specificity=68%, NPV=93%). Repeat DV occurred in only one-fifth of cases rated as high risk (PPV=19%). Importantly, 40 percent of cases where there was further DV were incorrectly classified as low risk (sensitivity=60%). False positives, where cases are incorrectly categorised as high risk, have significant resource implications for police and other agencies. However, it is important to note that, while police responses were controlled for in this study, other criminal justice and service responses were not, which may account for at least some of these false positives. Additionally, false negatives—cases classified as low risk where repeat violence does occur—mean that a victim may not be protected from further harm.

Finding: empirically refining the FVRAT substantially improves its predictive validity.

Using statistical analyses, the FVRAT was refined to its 10 most important items and optimal risk ratings were identified. This empirically refined tool was markedly more accurate than the original FVRAT. The AUROC score (0.73) is comparable to those of the better performing DV risk assessment tools such as ODARA (Graham et al. 2019; Messing & Thaller 2013), and by commonly accepted standards indicates acceptable predictive validity. This is particularly significant given the focus of this study was predicting repeat DV in the short term, when risk is highest (Morgan, Boxall & Brown 2018).

Finding: the empirically refined tool is better at predicting when repeat DV will occur, and just as accurate in predicting when it will not.

When DV cases classified as medium and high risk were compared with those classified as low risk, there was a notable increase in sensitivity. Around three-quarters to four-fifths of cases where repeat DV occurred were correctly classified into higher risk categories. Additionally, when examining high- and low-risk cases specifically, there is a substantial improvement in the PPV, with over a third of high-risk cases resulting in further DV (39%). These improvements, at least in relation to high- and low-risk cases, do not appear to have come at the expense of this tool's ability to correctly predict no further DV. Four-fifths of cases where there was no further violence were correctly classified as low risk (specificity=84%), and almost no low-risk cases resulted in further violence (NPV=96%).

Recommendations, limitations and future research

Findings suggest that the accuracy of the FVRAT can be improved by narrowing it down to a small number of important items. This would improve its predictive validity while giving frontline police a tool that is quicker and easier to use. For police, improving the tool's sensitivity is particularly important, as greater sensitivity means fewer victims 'slipping through the cracks' and experiencing further harm despite police assessing them as being at lower risk.

Many of the items identified in this report as important in predicting repeat DV were also found to be important in prior Australian research on DV risk assessment. These items relate to the severity of the violence in the most recent DV incident (eg physical assault), histories of violence (eg prior violence, escalating violence in the relationship, prior breaches of orders) and certain relationship stressors such as separation and financial difficulties (Lauria et al. 2017; Mason & Julian 2009; Millsteed & Coughlan 2016; Ringland 2018). This strengthens the argument for restricting the FVRAT to these items.

Importantly, a number of items that were found to be predictive in this prior research were not significant predictors of short-term repeat DV in the current study—namely alcohol/drug use and psychological health issues. While police responses aimed at deterring repeat DV were controlled for in this study, health and service responses that address these factors were not. Referrals to these services may be offered more persistently, and taken up more often, by offenders or victims who present with one or more of these risk factors, suppressing any direct association with repeat DV, particularly in the short term. Alternatively, the presence and extent of these complex factors may be difficult for frontline police to assess, potentially compromising the validity of this information.

Critically, this study did not assess the value of FVRAT items outside of their importance in predicting repeat DV in the short term. Items that did not emerge as predictive of short-term repeat DV may nevertheless be important in predicting longer term repeat DV, escalations in DV severity, or DV homicide. Additionally, items may still have investigative value to police even if they do not contribute to prediction, and could also inform the assessments and responses of others in the criminal justice system and services sector (Dowling & Morgan 2018). Items with some demonstrated or potential value to police decision-making, investigations or responses could be incorporated more explicitly into the currently under-utilised post-hoc procedures for reviewing and revising FVRAT risk ratings, or even as a supplementary tool to the FVRAT. For example, prior incidents of strangulation and suicide ideation by the offender have been identified as significant risk factors for DV homicide (Glass et al. 2008; Sherman et al. 2016), while alcohol use is generally associated with more severe acts of violence and escalating frequency and severity of violence (Sutherland, McDonald & Millsteed 2016). These items might be used to inform judgements regarding the need to prioritise offenders and victims for immediate intervention.

The results of this study provide further, albeit tentative, support for the importance of accounting for police and other short-term responses when measuring the accuracy of DV risk assessment tools. However, as the goal of risk assessment is not just to predict but to effectively target responses and ensure victim safety, evaluations of DV risk assessment tools should extend beyond prediction, and also examine whether they assist with risk-guided efforts to reduce DV (Backhouse & Toivonen 2018).

It is important to note the limitations of this study. Due to the sample size and low base rate of repeat DV, it was not possible to examine the FVRAT's ability to predict different types of repeat DV, or DV among victims or offenders of particular demographic characteristics such as gender or sexual orientation. While there are many similarities in the risk factors for repeat DV across these groups, some differences have been found, highlighting the importance of further research examining the validity of the FVRAT (eg Hilton et al. 2014; Lauria et al. 2017; Ringland 2018).

Further, this study does not analyse the reliability of the FVRAT to determine whether it gives a consistent measure of risk on different occasions (test–retest reliability) or for different users (inter-rater reliability) or information sources (victim vs administrative information). Any future validations of the FVRAT should include some form of reliability analysis, as limitations to its consistency will limit the accuracy with which it predicts repeat DV.

Finally, it should be noted that various forms of information readily available to police (eg criminal history patterns and demographic information) have been shown to predict repeat DV (eg Morgan, Boxall & Brown 2018; Richards et al. 2014). Additional research should compare the predictive validity of the FVRAT with the use of this information to ensure the former can enhance the police response over and above the information officers already have available to them.

References

URLs correct as at May 2019

Australian Bureau of Statistics (ABS) 2011. *Australian and New Zealand Standard Offence Classification (ANZSOC), 2011*, 3rd edition. ABS cat. no. 1234.0. Canberra: ABS. https://www.abs.gov.au/ausstats/abs@.nsf/mf/1234.0

Backhouse C & Toivonen C 2018. *National risk assessment principles for domestic and family violence: Companion resource*. Insights no. 9. Sydney: Australia's National Research Organisation for Women's Safety

Barrio I, Arostegui I, Rodríguez-Álvarez MX & Quintana JM 2017. A new approach to categorizing continuous variables in prediction models: Proposal and validation. *Statistical Methods in Medical Research* 26(6): 2586–2602

Belfrage H et al. 2012. Assessment and management of risk for intimate partner violence by police officers using the spousal assault risk assessment guide. *Law and Human Behavior* 36(1): 60–67

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 and the Law* 30: 420–430

Blumer C 2016. Australian police deal with domestic violence every two minutes. *ABC News,* 21 Apr. http://www.abc.net.au/news/2016-04-21/domestic-violence/7341716

Capaldi DM, Knoble NB, Shortt JW & Kim HK 2012. A systematic review of risk factors for intimate partner violence. *Partner Abuse* 3(2): 231–280

Crime and Misconduct Commission 2005. *Policing domestic violence in Queensland: Meeting the challenges*. Brisbane: Crime and Misconduct Commission

Dowling C & Morgan A 2018. Police investigations of domestic violence: What does the evidence say? *Police Science: Australia and New Zealand Journal of Evidence Based Policing* 3(2): 6–11

Dowling C, Morgan A, Boyd C & Voce I 2018a. *Policing domestic violence: A review of the evidence*. Research Report no. 13. Canberra: Australian Institute of Criminology. https://aic.gov.au/publications/rr/rr13

Dowling C, Morgan A, Hulme S, Manning M & Wong G 2018b. Protection orders for domestic violence: A systematic review. *Trends & issues in crime and criminal justice* no. 551. Canberra: Australian Institute of Criminology. https://aic.gov.au/publications/tandi/tandi551

Glass N et al. 2008. Non-fatal strangulation is an important risk factor for homicide of women. *Journal of Emergency Medicine* 35(3): 329–335

Graham LW, Sahay KM, Rizo CF, Messing JT & Macy RJ 2019. The validity and reliability of available intimate partner homicide and reassault risk assessment tools: A systematic review. *Trauma, Violence and Abuse.* Advance online publication. https://doi.org/10.1177/1524838018821952

Hilton NZ, Harris GT & Rice ME 2007. The effect of arrest on wife assault recidivism: Controlling for pre-arrest risk. *Criminal Justice and Behaviour* 34(10): 1334–1344

Hilton NZ, Popham S, Lang C & Harris GT 2014. Preliminary validation of the ODARA for female intimate partner violence offenders. *Partner Abuse* 5(2): 189–203

Hosmer DW & Lemeshow S 2004. *Applied logistic regression*. New Jersey: John Wiley & Sons

Janes H & Pepe MS 2009. Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristics curve. *Biometrika* 96(2): 371–382

Janes H & Pepe MS 2008. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: An old concept in a new setting. *American Journal of Epidemiology* 168(1): 89–97

Kane RJ 2000. Police responses to restraining orders in domestic violence incidents: Identifying the custody-threshold thesis. *Criminal Justice and Behavior* 27(5): 561–580

Kane RJ 1999. Patterns of arrest in domestic violence encounters: Identifying a police decision-making model. *Journal of Criminal Justice* 27(1): 65–79

Lauria I, McEwan TE, Luebbers S, Simmons M & Ogloff JR 2017. Evaluating the Ontario Domestic Assault Risk Assessment in an Australian frontline police setting. *Criminal Justice and Behavior* 44(12): 1545–1558

Mason R & Julian R 2009. *Analysis of the Tasmania Police Risk Assessment Screening Tool (RAST): Final report*. Hobart: Tasmanian Institute of Law Enforcement Studies

Mazerolle L et al. 2018. *Criminal justice responses to domestic and family violence: A rapid review of the evaluation literature*. Brisbane: University of Queensland

McEwan TE, Bateson S & Strand S 2017. Improving police risk assessment and management of family violence through a collaboration between law enforcement, forensic mental health and academia. *Journal of Criminological Research Policy and Practice* 3(2): 119–131

Messing JT & Thaller J 2013. The average predictive validity of intimate partner violence risk assessment instruments. *Journal of Interpersonal Violence* 28(7): 1537–1558

Millsteed M & Coghlan S 2016. *Predictors of recidivism amongst police recorded family violence perpetrators*. In Brief no. 4. Melbourne: Victorian Crime Statistics Agency

Morgan A, Boxall H & Brown R 2018. Targeting repeat domestic violence: Assessing short-term risk of reoffending. *Trends & issues in crime and criminal justice* no. 552. Canberra: Australian Institute of Criminology. https://aic.gov.au/publications/tandi/tandi552

Muller CJ & MacLehose RF 2014. Estimating predicted probabilities from logistic regression: Different methods correspond to different target populations. *International Journal of Epidemiology* 43(3): 962–970

Richards TN, Jennings WG, Tomsich E & Gover A 2014. A 10-year analysis of rearrests among a cohort of domestic violence offenders. *Violence and Victims* 29(6): 887–906

Ringland C 2018. *The Domestic Violence Safety Assessment Tool (DVSAT) and intimate partner repeat victimisation*. Crime and Justice Bulletin no. 213. Sydney: NSW Bureau of Crime Statistics and Research

Sherman L, Bland M, House P & Strang H 2016. *Targeting family violence reported to Western Australia Police 2010–2015: The felonious few vs. the miscreant many.* Somersham: Cambridge Centre for Evidence-Based Policy

Storey JE, Kropp PR, Hart SD, 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

Sutherland P, McDonald C & Millsteed M 2016. *Family violence, alcohol consumption and the likelihood of criminal offences.* In Brief no. 7. Melbourne: Victorian Crime Statistics Agency

Trujillo MP & Ross S 2008. Police response to domestic violence: Making decisions about risk and risk management. *Journal of Interpersonal Violence* 23(4): 454–473

Victorian Auditor-General 2009. *Implementing Victoria Police's code of practice for the investigation of family violence*. Melbourne: Victorian Auditor-General's Office

Dr Christopher Dowling is a Principal Research Analyst at the Australian Institute of Criminology.

Anthony Morgan is a Research Manager at the Australian Institute of Criminology.