Title: Police are influenced by anchoring and risk when allocating resources for scenario-based intimate partner violence cases.

Abstract

Sixty-six police officers were given four intimate partner violence (IPV) scenarios to rate for risk of future violence. At the start of the experiment, participants were provided with either a low risk or high risk ‘anchor’ scenario of police attending an IPV incident. Next, participants were given three counterbalanced scenarios: high, medium, and low risk. Half the participants were given a structured professional judgment tool to guide their decisions. Participants given the low risk anchor rated the following scenarios as being of greater risk than those given the high risk anchor. Participants were consistent in identifying high, medium, and low risk scenarios and the tool made no difference to these ratings. Participants were more confident in their higher risk judgments than their lower risk judgments. Officers distributed a disproportionately high amount of resources to the high risk offenders and the results suggest that police officers can make decisions consistent with Risk-Need-Responsivity principles. However, anchoring effects, and working in a context where violence is more severe and frequent has the potential to bias perceptions and make officers less sensitive to risk.

Key words: Domestic Violence; Intimate Partner Violence; Risk Assessment; Police.
Intimate partner violence (IPV) accounts for a great deal of police work (Devries et al., 2013; Robinson, Myhill, Wire, Roberts & Tilley, 2016; Sherman, 2018) and police resources are finite. This means that the police must prioritize their resources. While each perpetrator of IPV is unique, patterns and trajectories emerge (e.g., Holtzworth-Munroe & Stuart, 1994). For some perpetrators the violence may be a one-off event at the end of a relationship; for others the violence may continue to increase in frequency and severity. In some cases the violence may be so extreme that a person is killed in a future attack. If police were able to predict who would do what, they could effectively target their resources where they would be most impactful (Ogloff & Davis, 2004). For these reasons many police services around the world are attempting to use formal risk assessment processes to predict IPV.

**Risk Assessment**

Prior to the development of formal risk assessment processes, police officers arriving at an incident of IPV determined the level of risk based on both their experience and the information available to them. This is unstructured risk assessment. Clinical predictions made by psychiatrists and psychologists using unstructured risk assessments are generally shown to be inferior to purely statistical or structured methods of prediction (Ægisdóttir et al., 2006; Grove, et al., 2000; Quinsey, Harris, Rice, & Cromier, 1998). Possible explanations for this include: (i) biases the assessor might have, (ii) not looking for the most useful information to guide an assessment, (iii) not receiving feedback with regard to the accuracy of a judgment in order to calibrate one’s judgment and, (iv) comparison with a previous case that provides an extreme anchor.

Concerns about the validity of unstructured risk assessments have led to a shift towards formal ‘actuarial’ or ‘structured-professional-judgment’ (SPJ) approaches with regard to IPV (for
reviews see, Messing & Thaller, 2013; Nicholls, Pritchard, Reeves, & Hilterman, 2013; Wheller & Wire, 2014). These approaches have been taken up by many police services (Robinson, Myhill, Wire, Roberts, & Tilley, 2016; Storey, Kropp, Hart, Belfrage, & Strand, 2014). Actuarial approaches are based on statistics and remove personal decisions entirely. Actuarial risk is usually calculated from large datasets and analyzed for factors that are related to risk. In comparison, SPJ requires an assessor to collect specific information to guide his or her own decisions about risk.

The Ontario Domestic Abuse Risk Assessment (ODARA; Hilton, Harris & Rice, 2010) is an example of the actuarial approach with, arguably, the strongest evidence-base (see, for example, Wheller & Wire, 2014). Other actuarial approaches include the Danger Assessment (DA; Campbell, Webster, & Glass, 2009) and the shortened version of the DA, the Lethality Assessment (Messing, Campbell, Sullivan Wilson, Brown, & Patchell, 2017). The ODARA is a 13 item tool for identifying the risk of future IPV for men who have already committed at least one act of IPV (Hilton, Harris, & Rice, 2010). The items require a review of police records as well as the most recent IPV incident. They are as follows: has a prior domestic assault (against a partner or child) in police Record Management System (RMS), has a prior non-domestic assault (against anyone other than a partner or child) in RMS, has a prior sentence to a term of 30 days or more, has a prior failure on conditional release including bail, parole, probation, and no-contact order, has threatened to harm or kill anyone during the index offence, has unlawfully confined the victim during the index offence, the victim fears the repetition of violence, the victim and/or the offender have more than one child altogether, the offender is in a stepfather role in this relationship, the offender is violent outside the home (to people other than a partner or child), the offender has more than one indicator of a substance abuse problem, the offender
has ever assaulted the victim when she was pregnant, and, the victim faces at least one barrier to support. These items are scored as present or absent and then added up. If the score is seven or more the person is in the highest risk category. There is evidence to support the reliability and validity of the ODARA (for reviews see: Graham, Sahay, Rizo, Messing, & Macy, 2019; Kebbell, 2019).

The Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER) is an exemplar of the SPJ approach for police (Kropp & Hart, 2004). It is an adaptation of the Spousal Abuse Risk Assessment (SARA) - a clinical tool which has been modified so it can be used by police without the qualifications needed to administer the SARA (Belfrage, et al., 2012). The items are: serious physical or sexual violence against a partner, serious violent threats or thoughts about violence to a partner, escalation of physical or sexual violence, violations of civil or criminal court orders, negative attitudes about IPV, other serious criminality, intimate relationship problems, employment or financial problems, substance abuse, and mental health problems. The B-SAFER seems to have similar validity to the ODARA (Storey, Kropp, Hart, Belfrage, & Strand, 2014) and there is evidence to support the validity of the B-SAFER (for reviews see: Graham, Sahay, Rizo, Messing, & Macy, 2019; Kebbell, 2019). Direct comparisons of risk assessment tools ‘like for like’ are difficult across different studies due to the various methodologies used. Some studies look at offending over 12 months whilst others look at offending over five years. Further, some studies may use charges as evidence of reoffending while others use convictions or even allegations. For these reasons, caution must be exercised when saying there are definite differences between risk assessment tools because they have not been compared with identical criteria in the same study. Indeed, compared with forensic risk assessment tools such as the
Static-99 (Hanson, Thornton, Helmus, & Babchishin, 2016) and Level of Service Inventory (Olver, Stockdale, & Wormith, 2014) all IPV risk assessment tools have a limited evidence-base.

Another SPJ tool, similar to the one used in this study, is the Domestic Abuse, Stalking and Harassment and Honour-based violence risk identification, assessment and management model (DASH; Robinson, Myhill, Wire, Roberts, & Tilley, 2016). The DASH consists of 27 items that are scored primarily by asking the victim about the perpetrator (see Robinson, Myhill, Wire, Roberts, & Tilley, 2016). The items were developed by amalgamating other tools that were in use in England and Wales, and with reference to academic literature. For these reasons, many of the items that are used in the DASH overlap with those of the ODARA and the B-SAFER. The DASH does not have a strong empirical base with regard to its validity, for example, Almond, McManus, Brian, and Merrington (2017) found that most of the items in the DASH do not predict future IPV.

**Defining and communicating risk**

Defining and communicating risk is difficult. For example, Hilton, Carter, Harris, and Sharpe, (2008) found that 60 forensic clinicians were inconsistent in their ratings of risk when assessing the same scenarios using percentiles, putting a point on a line from 1-100, or using terms such as ‘high risk’. Similarly, Barbaree, Langton, and Peacock (2006) show that the administration of different validated risk assessment tools can give different risk ratings to the same offenders because they rely on different measures or use different criteria for reporting risk. Indeed, risk may mean different things to different people. For example, Scurich (2018) and Hogan (2020) both point out that people use categorical risk judgments (e.g., high, medium, etc.) inconsistently, interpreting the terms in different ways. Further, Hogan (2020) argues that time frames are critical, pointing out that a recidivism rate of 5% over one year is very different to a
similar percentage over five years. He uses the analogy that a family income of $100,000 for one year is very different to $100,000 for five years. Time can be critical for police where officers may need to respond immediately rather than over five years.

**Factors that may bias risk assessment**

There are multiple ways that decision-making can be biased to increase errors (see for example, Kahneman, 2011; Klein, 2017). Studies have shown that some of this research can be applied to police decision-making. For example, a priming-bias against a person who is identified as an early suspect can occur (see Ask & Granhag, 2005). One potential bias that was mentioned by police officers during the pilot stage of this research was the potential for police officers to be influenced by their previous experiences with responding to IPV. Officers suggested that police who had worked in environments where IPV was highly prevalent, and there was a great deal of extreme violence (e.g. stabbings, beatings with weapons), had a higher tolerance for lower-level IPV (e.g. shouting, smashing furniture), than those without this experience.

This issue of context has not been studied with police experimentally. The influence of prior experience in this way would be a manifestation of the anchoring and adjustment heuristic (Tversky & Kahneman, 1974). A classic study of anchoring and adjustment bias tested adjustments from externally provided anchors (Jacowitz & Kahneman, 1995). People were first given an anchor by the experimenter and then asked to make their decisions. They were asked to estimate the average number of babies born in the United States per day, either after they considered whether it was higher or lower than 100, or after they considered whether it was higher or lower than 50,000. The median responses were estimates of 1,000 in the former case and 40,000 in the latter case, clearly showing that the anchor makes a difference to the estimate
The implication of this is that police officers, when confronted with a very high risk case, may use this as an anchor and compare the next case they deal with to the previous one. Thus, if the next case is low risk it may be perceived as being particularly low risk because of the previous anchor and vice versa.

**Risk-Need-Responsivity**

Assessing risk accurately and communicating the results is not enough. For a risk assessment to be useful the results need to be applied. If all offenders are treated the same, regardless of the assessed risk they pose, there is little point in conducting a risk assessment. High risk offenders should receive more attention than low risk offenders. The Risk-Need-Responsivity (RNR) model is relevant here and has a strong evidence-base for reducing offending (e.g., Andrews & Bonta, 2017; Andrews, Bonta, & Hoge, 1990; Andrews, Zinger, et al., 1990; Gendreau, Little & Goggin, 1996; Ogloff & Davies, 2004). Put simply, the RNR model requires that the level of intervention an individual receives matches the risk that they pose to the community. The Risk principle means that high risk offenders will generally require more intensive and extensive services than those offenders who have been assessed to be at a lower risk of recidivism. Therefore, priority is given, and resources directed, to higher risk offenders. In comparison, those offenders considered to be at a lower risk of reoffending will receive minimal, or even no, intervention. The Need principle means identifying what needs the offender has that must be addressed to reduce the risk of offending. For IPV offenders these might include attitudes towards women and using violence against women, and abuse of alcohol or other drugs. The Responsivity principle concerns the delivery of evidence-based services in a manner which increases the likelihood of offenders responding positively. This principle involves tailoring services to match the learning style, motivation, abilities, and strengths of the offender.
There is evidence that the RNR approach applies to IPV (Radatz, & Hilton, 2019; Radatz, & Wright, 2016) and that some IPV offenders may have more risk factors than other violent offenders (Hilton & Radatz, 2018). While it may seem obvious that offenders who pose a greater risk to the community require more resources, it may be less clear that there are problems with giving too many resources to low/medium and low risk people. Stewart, Flight, and Slavin-Stewart, (2013) point out that committing too many resources to medium and low risk offenders dilutes the resources available to those who need them most – the high risk offenders. For interventions to be successful they require a great intensity of treatment in the order of 200 hours or more (see for example, Bourgon & Armstrong, 2005). Misidentifying people as high risk can increase the likelihood of people offending. For example, removing pro-social supports (such as family and employment because of imprisonment) and forcing low risk offenders into contact with high risk offenders via a treatment program may increase offending (Bonta, Wallace-Capretta, & Rooney, 2000). This is important as there is research to suggest that when people override a risk assessment it is to make the assessment higher, rather than lower. For instance, Orton, Hogan, and Wormith (2020) found that when a ‘professional override’ was used, assessors increased risk in 94.1% of cases on the Level of Service Inventory - Ontario Revision. These overrides generally inflated risk disproportionately and made the predictions of true recidivism less accurate (see also, Monahan, 1981).

The current study

The current study was an examination of police officers’ perceptions of risk and how this influenced their allocation of resources. Police officers were given IPV scenarios and asked to rate risk and allocate resources. They were either given a SPJ tool or not. They started the task with either a high risk or a low risk anchor.
The hypotheses were as follows:

1. Consistent with Tversky and Kahneman (1974), it was predicted that police officers’ risk ratings would be biased by anchoring effects. Scenarios following a very high risk case would be perceived as of lower risk of violence. Scenarios following a very low risk case would be perceived as of higher risk of violence.

2. Consistent with the risk assessment literature (e.g., Ægisdóttir et al., 2006; Grove, et al., 2000; Quinsey, Harris, Rice, & Cromier, 1998), unstructured risk assessment by police officers was predicted to be less accurate than structured risk assessment.

3. In line with Andrews and Bonta (2017), police officers were predicted to allocate more police resources to those they perceived to be of higher risk.

Method

Participants

Participants were 66 police officers. All were in front-line, general duty, roles that included dealing with IPV incidents. Of the participants, 40 identified as constables, 20 identified as senior constables and four identified as sergeants. Forty-three identified as male and twenty as female, the remainder did not identify their gender. The average age was 37 years (SD = 7.1, range 23 to 51 years). The average length of service was 7.5 years (SD = 6.4, range 3 months to 30 years). A police member contacted police stations and asked the sergeant or senior sergeant in charge if a researcher could attend at a convenient time. The researcher would then attend and ask any available officers if they would participate.
An a priori power analysis using the G*Power software package (Faul et al., 2007), using a large effect size and power of .80 ($p < .05$), indicated a sample size of 48 was needed. Power of .80 is generally considered sufficient (Geher & Hall, 2014).

**Materials**

Five IPV scenarios were created. To create the scenarios the first author read through 100 IPV incidents picked at random from a police database. The database included police accounts of what had happened, previous police contact with the different people involved (e.g., the aggrieved and the respondent) and other evidence such as court orders. From these incidents, a consideration of the risk assessment literature, and discussions with police officers with experience of IPV incidents, five scenarios were developed. To control for length each scenario was exactly 275 words. These scenarios are available on request.

To try to obtain an objective measure of risk, notwithstanding the fact that even the best risk assessments are far from perfect, risk items were included that were in both the ODARA and the B-SAFER that would approximate to the assigned risk categories. For example, for the high risk case of Richard and Katherine, some factors in the scenario were ‘previous incidents/contraventions’, ‘frequency of domestic violence’, and ‘drugs/alcohol misuse’ – all risk factors included in both the ODARA and B-SAFER. These scenarios were piloted with five experienced police officers who reviewed them and rated them for risk and plausibility which provides some convergent validity. At the end of this process the five scenarios were: a high risk anchor; a low risk anchor; a high risk scenario; a medium risk scenario; and a low risk scenario.

An IPV risk assessment tool was given to half the participants. This was the tool used by the Police Service. The tool was based on the DASH with the following DASH items removed: ‘presence of child abuse’, ‘stepfather relationship to child’, ‘injuries’, ‘child custody dispute’,...
and, ‘other people threatened victim’. There was an additional item included: ‘prior failure of conditional release’.

For each scenario participants were asked to rate: whether there would be another call to police concerning this couple in the next year, if there was another call how serious would it be, and how confident they were in their answers. All these questions required responses on 10-point Likert scales as are described in the results section. Participants were also asked to rate risk from the categories: Unknown, Medium, High, and Extreme. These are the categories of risk used by the Police Service. They were also asked to rate their confidence in their assessment.

**Design**

The design was a 2 (high risk anchor/low risk anchor) X 2 (no risk assessment tool/structured professional judgement tool) X 3 (low risk offender/medium risk offender/high risk offender) ANOVA with repeated measures on the third factor. The dependent variables were ratings on the risk assessment tasks such as likelihood of another call for service.

**Procedure**

Participants were identified by the police sergeant or senior sergeant on duty and asked if they would participate. It was made clear that they did not have to do so and could stop participating at any time. An information sheet was provided concerning the Study and details of the University Ethics Board and how it could be contacted if there were any concerns. Participants were then given the test booklets and asked to work their way through. The test booklets included four vignettes concerning IPV men. Only men were the perpetrators in these scenarios to control for potential differences in perceptions of male or female IPV offenders. For half the participants the first vignette was ‘John and Susan’. In this vignette John was a low-risk
offender. For the other half the participants received ‘Richard and Katherine’ as the first vignette. In this vignette Richard was a high-risk offender. Thus, participants started the task with either a low or a high risk anchor. This was then followed by the remaining three scenarios that were counterbalanced for order effects. Participants were also asked to allocate 100 hours of police time between the four IPV scenarios. Finally, demographic information was collected (age, gender, length of service, and rank).

Results

Manipulation check

As a manipulation check, participants’ ratings of the high and low risk anchor scenarios were compared with between-subjects $t$-test on the dependent variables of likelihood of another call for service (1 = not at all likely; 10 = extremely likely) and seriousness of the next call for service (if it occurs) (1 = very minor non-violent domestic dispute; 10 = extreme levels of violence/homicide). Both were significant, $t(65) = 18.56, p < .001$ and $t(65) = 14.71, p < .001$, respectively. Participants identified the high risk offender as being more likely to be involved in a call for service (high anchor $M = 9.70, SD = 0.53$ versus low anchor $M = 3.73, SD = 1.77$) and for any future call for service to be serious (high anchor $M = 8.21, SD = 1.27$ versus low anchor $M = 2.80, SD = 1.69$).

To check that the SPJ tool could be scored consistently Intraclass Correlation Coefficients (ICC) were calculated using a two-way random effects model with absolute agreements across the three scenarios. These were very strong for the high-risk offender (average ICC = .97), the medium-risk offender (average ICC = .97), and the low-risk offender (average ICC = .94).
Ratings of risk and confidence in ratings of risk

To test the main hypotheses a series of 2 (high risk/low risk anchor) X 2 (no risk assessment tool/risk assessment tool) X 3 (high risk offender/medium risk offender/low risk offender) ANOVAs with repeated measures on the third factor were conducted. The first ANOVA was on the likelihood of another call for service in the next year. There was a significant effect of the initial anchor, $F(1,61) = 10.82, p < .01, \mu^2 = .151$. Participants who were given the low risk anchor rated the next offenders as being of greater risk ($M = 6.47, SD = 0.87$) than those given the high risk anchor ($M = 5.70, SD = 1.01$). There was no significant effect of the SPJ tool, $F(1,61) = 1.35, p = .25, \mu^2 = .022$. There was a significant effect of the risk of the offender, $F(2,122) = 309.07, p < .001, \mu^2 = .835$. Follow up $t$-tests indicated that the participants identified the high risk offender as being of much higher likelihood ($M = 9.00, SD = 1.03$) of having another call for service than the medium risk offender ($M = 6.12, SD = 1.89$) who in turn was rated higher than the low risk offender ($M = 3.15, SD = 1.48$). None of the interactions were significant. This data is shown in Figure 1.

A second ANOVA used the same 2 (high risk/low risk anchor) X 2 (no risk assessment tool/risk assessment tool) X 3 (high risk offender/medium risk offender/low risk offender) design as the first ANOVA but with seriousness of next call for service as the dependent variable. The results mirrored the first ANOVA. There was a significant effect of the initial anchor, $F(1,61) = 4.89, p < .05, \mu^2 = .074$. Participants given the low risk anchor rated the offenders that followed afterwards as being of greater risk ($M = 5.61, SD = 0.93$) than those given the high risk anchor ($M = 5.09, SD = 0.93$). There was no significant effect of the SPJ tool, $F(1,61) = 1.28, p = .26, \mu^2 = .021$. There was a significant effect of the risk of the offender, $F(2,122) = 507.38, p < .001, \mu^2 = .820$. Follow up $t$-tests indicated that the participants consistently identified the likelihood of
the high risk offender committing a more serious offence next ($M = 8.52, SD = 1.03$) than the medium risk offender ($M = 4.25, SD = 1.66$) who in turn was rated higher than the low risk offender ($M = 3.28, SD = 1.60$). Again, none of the interactions were significant.

Participants’ confidence in their judgments were initially analyzed with separate 2 (high risk/low risk anchor) X 2 (no SPJ tool/SPJ tool) X 3 (high risk offender/medium risk offender/low risk offender) ANOVAs with repeated measures on the third factor for both officers ‘confidence in their estimate of likelihood of another call for service and the officers’ confidence in their estimate of seriousness of the next call for service. Confidence was measured on a Likert scale (1 = not at all confident; 10 = extremely confident). The results were very similar and so for brevity both confidence judgments were combined to produce a composite score of ‘confidence’. There was no significant effect of the initial high or low anchor, $F(1,61) = 1.27, p = .264, \mu^2 = .020$ and there was no significant effect of the SPJ tool, $F(1,61) = 0.67, p = .42, \mu^2 = .011$. However, there was a significant effect of the risk of the offender, $F(2,122) = 27.42, p = .001, \mu^2 = .310$. Follow up $t$-tests indicated that participants were more confident the accuracy of their judgements of the high risk offender ($M=8.95, SD=1.09$) than they were in the accuracy of their judgments of the medium risk offender ($M=7.59, SD=1.74$) or the low risk offender ($M = 7.71, SD = 1.75$). Participants did not differ in their confidence in their risk judgments between the medium and low risk offenders.

**Allocation of resources**

Participants were asked to allocate 100 hours of police time to the four different IPV scenarios (including the initial anchors). These allocations were analyzed with a 2 (high risk/low risk anchor) X 2 (no SPJ tool/SPJ tool) X 4 (high or low risk offender/high risk offender/medium risk offender/low risk offender) ANOVA with repeated measures on the third factor. There was
no significant effect of the initial anchor, $F(1,59) = 1.60, p = .21, \mu^2 = .026$, or of the SPJ tool, $F(1,59) = 0.00, p = .99, \mu^2 = .000$. However, there was a significant effect of the risk of the offender, $F(1,59) = 73.62, p < .001, \mu^2 = .460$, and an interaction between offender and if there was an initial high or low risk anchor offender, $F(3,177) = 29.78, p < .001, \mu^2 = .335$. None of the other interactions were significant. These results are displayed in Figure 2. Follow up $t$-tests ($p < .05$) showed that there was a significant difference between the allocations of resources to each offender depending on anchor – except for the low risk offender where both conditions allocated few hours. The presence of the high risk anchor meant that participants took resources from the other high-risk offender and the medium risk offender in that condition to allocate to the high-risk anchor.

**Discussion**

In the current study, police officers’ perceptions of risk were tested with IPV scenarios to determine how this influenced their allocation of resources. The study also examined the effect of a high risk or low risk anchor on officers’ judgments and compared the use of a structured professional judgment tool with unstructured risk assessment.

**Anchoring effects**

As expected, there was an effect of anchor. Those who started with a high risk anchor rated subsequent offenders as less risky. In contrast, those who started with a low risk anchor rated subsequent scenarios as riskier. This has implications for policing. Officers’ day-to-day experiences of offenders may mean they make relative judgments about risk depending on their previous experience. Officers working in violent areas may have a greater tolerance for violence than those who work in more peaceful areas. In turn, this means that people who live in violent areas may have a lesser police response than those in less violent areas. This may compound
problems that exist in high violence areas as residents may have less access to justice. Conversely, more affluent areas with lower levels of violence may be over-policed, leading to a resource disparity that disadvantages those in lower-socioeconomic areas and, as a result, cycles of poverty, crime, abuse and racism are continued (Adams, Tolman, Bybee, Sullivan, & Kennedy, 2012; Voith, 2019).

If an aim is equal access to justice in a diverse communities, police resources should be targeted in a transparent way to where they are most needed. Indeed, some argue that unequal levels of policing across diverse jurisdictions is one of the most critical issues facing police today (Leon-Moreta, 2018). To prevent anchoring effects from reducing police responses in more violent environments, and to ensure people are treated fairly, attempts might need to be made to overcome this anchor-bias. Research from cognitive psychology suggests this is difficult, for instance giving people a monetary incentive for more accurate predictions does not seem to be effective (Simmons, LeBoeuf, & Nelson, 2010), and making people aware of the effect does not seem to assist either (Wilson, Houston, Etling, & Brekke, 1996). The use of actuarial tools such as the ODARA (Hilton, Harris & Rice, 2010) might be useful because this would remove anchoring as a potential source of bias. Further, the implication of this is that at an aggregate, district-level, police resources might need to be focused where there are the most IPV incidents and harm (see for example, Sherman, Bland, House, & Strang, 2016).

**Structured risk assessment**

The police officers in this experiment distinguished between the cases designated as high, medium and low risk of IPV in a similar way to the levels of risk that would be predicted by risk assessment tools (Hilton, Harris & Rice, 2010; Storey, Kropp, Hart, Belfrage, & Strand, 2014) and the subject-matter-experts who helped design the scenarios. This finding is also consistent
with a survey of what police officers’ rate as being the most important (Robinson, Pinchevsky & Guthrie, 2016). Robinson, Pinchevsky and Guthrie (2016) found that using or threatening to use a weapon, strangulation, and physical assault resulting in injury were all rated as being particularly high-risk behaviors. These factors were included in the high risk scenarios in this experiment that were categorized as such by participants. The absence of risk factors in other scenarios was associated with lower perceptions of risk and is, again, consistent with what one would find from a risk assessment tool - an absence of risk factors is associated with lower risk.

One reason for this positive finding might be that the officers in the police service used for this study had all been trained to investigate IPV and use an IPV risk assessment tool. Consequently, they may have been so attuned and practiced in identifying risk factors that they did not need the tool to prompt them concerning risky behavior. Nevertheless, there was no significant effect of the tool on anchor effects showing that, even with an SPJ tool, biases can be introduced although this finding may not generalize to risk assessment tools other than the one used here.

Relatedly, participants were more confident in their ratings of high risk offenders than the medium and low risk offenders. In the case of high risk offenders, police have enough information to decide that the offender is high risk. For instance, if an offender has stabbed and strangled his partner, it is self-evident that he can commit a serious offence. For this reason, police can be confident in their high risk judgements. However, unpredictable situational factors mean risk assessments should always be treated with some caution. The risk someone poses is situationally dependent on factors that are often beyond the control of the police (Wortley, 2001). For instance, one IPV offender may separate from a former wife and then meet a new partner and move to a different city. In this case the offender may be of low risk of offending against his ex-wife. In another instance, an offender with the same risk characteristics may not find a new
partner, become inebriated, then encounter his ex-wife and hit her. The situation has influenced 
the outcome and is difficult to predict. Thus, even if police assess risk accurately, they are not be 
able to account for all situational factors.

Farrell, Petersen, Nicholls, and Roesch (2020) point out that there is little research that 
has explored the relationship between confidence and accuracy in risk judgments. This appears a 
topic worthy of future research because the current findings have implications for the adoption of 
a RNR framework as police would need to be confident in low risk assessments and not assign 
too many resources to them. As mentioned previously, risk assessors tend to use overrides to 
increase risk levels (Orton, Hogan, & Wormith, 2020) and incorrectly assigning a low risk 
person as higher risk can increase the risk of the person by reducing protective factors, such as 
making them homeless by removing them from their home (Bonta, Wallace-Capretta, & Rooney, 
2000). Confidence in a low risk judgment is important, therefore. Some factors that may increase 
confidence that someone is low risk might include belief in the validity of the risk assessment 
tool and belief that information on a police management system is accurate. In addition, there 
may be some need for officers to understand that offenders may reduce in risk over time and so 
risk assessments should decrease over time if the person has not offended (for an example of this 
effect with sex offenders see Hanson, Harris, Letourneau, Helmus, & Thornton, 2018: Olver, et 
al., 2018). Thus, reduction in risk seem a critical issue worthy of future research.

 Allocation of resources

In this study police put by far the most resources into those whom they perceived to be 
the highest risk offenders. In line with Andrews and Bonta (2017), police officers allocated more 
police resources to those they perceived to be of higher risk. There is some evidence to support 
this happens in practice. Belfrage and Strand (2012) found that police put most of their resources
into the highest risk IPV offenders. This reduced high risk IPV offending (Belfrage & Strand, 2012). This is entirely consistent with the offender rehabilitation literature and RNR (Radatz, & Hilton, 2019; Radatz, & Wright, 2016; Stewart, Flight, & Slavin-Stewart, 2013).

The allocation of far more resources to the highest risk offenders compared with the low risk offenders is consistent with findings that high risk offenders need a very high intensity of treatment (Bourgon & Armstrong, 2005) while no or minimal intervention is appropriate for managing the lowest risk offenders. Overall, the current study demonstrated that once a judgment of risk is made, police officers can prioritize resources consistent with a RNR framework for reducing future IPV offending. This suggests police are likely to be supportive of an evidence-based RNR approach, potentially. This allocation of resources, including how much time should be spent investigating properly, and how much effort should be spent protecting victims, seems deserving of being a focus of future research.

Limitations

There are limitations to the current study. One limitation is that the scenarios were easier to conceptualize and rate than real cases - the scenarios were short, and all the relevant information was spelt out whereas many police are reporting they have an overload of information and a high workload (Duxbury, Higgins, & Halinski, 2015). This might be very different in real cases where officers may need to investigate and interrogate police databases to make decisions. This would make the decision-making more complicated for police officers, and mean much more information to go through, increasing the likelihood of missing risk factors or putting too much weight on irrelevant information. Further, police may have an additional bias because the consequences for the officer may be lower for a false positive (incorrectly
identifying someone as high risk) than a false negative (incorrectly identifying someone as low risk).

To conclude, an anchor of a high or low risk offender biased police officers’ decisions in this experiment. Officers distributed a disproportionately high amount of resources to the high risk offenders and so made decisions consistent with Risk-Need-Responsivity principles. Taken as a whole the current study helps shed light on how police make decisions with regards to IPV and provides opportunities for enhancing police responses.
References


doi:10.1037/law0000135


doi.org/10.1080/13218719.2020.1767719


Cardiff & London: Cardiff University, College of Policing and UCL Department of Security and Crime Science.


doi.org/10.1177/1524838017717744


Figure 1. Police officers’ perceptions of the likelihood of another call for service for the different offenders with a high- or low-risk anchor.

Figure 2. Police officers’ distribution of 100 hours of police time and the influence of a low or high-risk anchor.