

# Police Attempts to Predict Domestic Murder and Serious Assaults: Is Early Warning Possible Yet?

Sara Thornton<sup>1</sup>

Published online: 28 August 2017

© The Author(s) 2017. This article is an open access publication

## Abstract

*Research Question* What facts known to police, if any, could have predicted the 118 domestic and family murders and near-murders in the Thames Valley (UK) police area over a recent three-year period?

*Data* Thames Valley Police records were examined for 118 victims and 120 offenders in all 118 cases of “deadly” domestic violence: murder, attempted murder, manslaughter, and grievous bodily harm with intent in calendar years 2007, 2008 and 2009. Police National Computer (PNC) and other records of police contact with those persons prior to the cases identified for this prediction study were also coded, as well as the PNC records of a case-control sample of 150 male and 100 female offenders in less serious violent crimes. Local police risk assessments, including the DASH protocol, were coded for deadly violence cases with prior police contact.

*Methods* Risk assessments made during prior police contacts with domestic violence cases during the time period were coded for accuracy of forecasts. Characteristics of offenders and victims in deadly violence were also compared to characteristics of the case-control sample, calculating the relative-risk ratios between the two groups for the presence of factors that further research could explore as possible predictors of deadly domestic violence.

*Findings* In 55% of deadly violence cases, there was no prior recorded contact by police with the victim; a further 21% of victims had only one prior police contact. Not one of the 13 murder cases with any prior police contact had ever been assessed by responding officers as “high risk” based on DASH or its predecessor, for a 100% false negative rate of prediction on risk of murder and an 89% false negative rate for any deadly violence (only 6 cases with prior contact had been classified as high risk). In contrast, during the same time period, 2721 other domestic violence cases yielded “high-risk” assessments but were not actually followed by deadly violence. Yet the case control study found male offenders who committed serious domestic assaults were

---

✉ Sara Thornton  
Sara.Thornton@npcc.pnn.police.uk

<sup>1</sup> National Police Chiefs’ Council, England, UK

over three times more likely than other violent offenders to be have had markers of suicidal tendencies.

*Conclusions* Predicting deadly domestic violence based on intelligence from prior police contacts does not appear possible at present, given these findings that less than half of these cases had prior police contact, and that when contact did occur, the relationships were assessed by the DASH protocol or its predecessor as *not* high risk in 89% of those cases. A more promising predictor for male offenders in these cases, even without prior contact, may be a prior marker for suicide (ideation, threats or attempts), but use of that marker requires further research on the extent to which it can be identified before the deadly violence occurs, either from police contacts or other sources, so that more intensive efforts can be undertaken to prevent deadly violence in the truly high-risk cases.

**Keywords** Domestic homicide · Risk assessment · Prediction accuracy · DASH

## Introduction

Domestic murder and other serious domestic assaults can often look predictable with hindsight, so many claim they could have been prevented. But how does murder look with foresight? To what extent is it possible to identify few cases which will result in the most serious harm, out of many thousands of other cases that look very similar but never become so serious?

In 2003 and 2004, there were three particularly tragic cases of domestic murder in the Thames Valley. In every case, the victim's families asked whether the death could have been predicted and, if so, whether it could have been prevented. In response to these and other cases, Thames Valley Police made a significant investment in training officers, creating specialist units, joining in multi-agency arrangements and applying a new risk assessment model. This article will focus on the risk assessment model which was implemented in 2005–2010.

There is a considerable body of literature on using prediction in criminology. Yet the evaluation of risk assessment tools is also in its infancy. While risk factors have been identified, their accuracy in terms of predictions has been low, with many false positives and false negatives (Heckert and Gondolf 2004). A risk assessment approach known as SPECSS+ (Separation, pregnancy, escalation, community issues, stalking and sexual assault) was introduced in most forces including Thames Valley during 2005 and 2006. This was then replaced by DASH (Domestic Abuse, Stalking and Harassment) in 2009. The assessment is completed by operational officers attending any domestic crime or incident, although in some forces, this is restricted to domestic crimes only. The officer asks the victim a series of questions in order to complete a form, and on the basis of those answers, cases are then assessed as standard, medium or high. The tool is based upon analysis of previous domestic murders and seeks to assess the potential dangerousness of cases so that the high risk case can be identified and then managed within a multi-agency panel known as a Multi-Agency Risk Assessment Conference (MARAC) (Richards et al. 2008).

Despite the literature on the reliability of risk assessment, DASH and its predecessor were introduced without any form of effective evaluation of its use in the pilot sites. This lack of evaluation raises basic questions about the accuracy and reliability of the forecasts produced by the risk assessment. The present study is a first effort to answer those questions with respect to deadly domestic and family violence, at least for one police force.

The study was completed in 2011 as a Cambridge University thesis examining cases of domestic murder and serious assault. At the time of the crimes included in the present study, the Association of Chief Police Officers' (ACPO) definition of domestic violence was "any incident of threatening behaviour, violence, or abuse (psychological, sexual, financial or emotional) between adults, aged 18 or over who are or have been intimate partners or family members regardless of gender or sexuality". The definition therefore included family relations other than spouse or partner, as well as victims of either gender.

The specific research objectives of the study were (1) to establish whether the victims were known to the police prior to the deadly (including non-fatal) assaults and (2) if they were, how accurate the risk assessment tool was in predicting serious harm. The study also asked whether (3) the offenders in these deadly assaults differed in predictive ways from a case control sample of other violent offenders in the Thames Valley.

## **Earlier Studies on Prior Record of Domestic Abuse Predicting Deadly Assaults**

The accepted wisdom is that escalating domestic violence provides an opportunity for the police and other agencies to intervene to prevent further harm. However, the prevalence of a prior history in cases of domestic murder varies widely across the many studies that have been undertaken. These studies also vary in their units of analysis, from addresses to offenders to couples.

**Addresses** Early research in Kansas City in 1971–1972 showed that police had been called to the *address* of the victim or suspect in domestic homicide cases over the last 2 years in 90% of the cases and that police had been called to the address more than five times in 50% of the cases (Breedlove et al. 1977). Research in Minneapolis in 1985–1989 showed that given the 52 homicides in the sample, the rate of domestic homicide was 16.83 per 1000 for addresses with more than nine call outs (and this was considerably more than the .28 per 1000 where there had been no call outs). While the risk of deadly violence is greater at addresses with more call outs, it remains the case that over 98% of the time any prediction arising out of prior reporting would be incorrect (Sherman 1992, p7).

**Offenders** Neil Websdale's influential work on domestic murders in Florida in 1994 showed that 86.6% of offenders had a prior history of battering (Websdale 1999).

**Couples** A study based in Atlanta in 1984 showed that 30% of all domestic homicides had been preceded by an offence involving either the victim or the offender in the last

4 years (Saltzman et al. 1992). Research in Milwaukee in 1987–1989 found that of 33 domestic homicides, only one couple had a prior report of domestic violence, but that study did not examine criminal history records for other kinds of crime as the Atlanta study did (Sherman et al. 1991). Research in Victoria, Australia, in 1988–1992 found that 90% of the couples in which an intimate partner was killed had had no prior contact in respect of the 82 cases of domestic homicide studied; conversely, couples who report domestic abuse incidents almost never experienced a homicide (Sherman and Strang 1992). Analysis by the Canadian Bureau for Justice Statistics in 1991 found that 42% of cases had prior contact (Canadian Bureau for Justice Statistics 1992).

## **The Development of Risk Assessment Models for Domestic Homicide: USA**

Risk assessments for domestic violence were first developed in the USA about 30 years ago. Barbara Hart, who had a legal background, was the first to develop a lethality assessment based upon her work as a practitioner (Hart 1988). Jacqueline Campbell's Danger Assessment (Campbell et al. 2003) is probably the most widely used. Campbell's Danger Assessment tool was then developed into a "lethality screen for first responders" for law enforcement officers in Maryland. This assessment tool has 11 questions and is used to predict the danger and potential for lethality in situations. If a victim is assessed as high risk, then they are referred to local programmes. Evaluation of this approach in 2006 and 2007 showed that 57% of victims screened were assessed as high risk and 54% of this group spoke to a programme counsellor on the phone (Sargent and Campbell 2008). However, the numbers are very small and this could reflect random variation.

Campbell worked with practitioners on National Institute of Justice research on risk assessment to create an empirical base and use a multi-site case control study (Campbell et al. 2003). This research was then used to revise the Danger Assessment. All the cases of femicide in 11 cities between 1994 and 2000 were examined to identify 545 closed cases where the perpetrator was a current or former intimate partner. In each of these cases, a knowledgeable family friend or relative was identified from police records, approached and asked to participate. In 373 (68%) of cases, a proxy was identified, and in 307 cases, they agreed to participate. Two criteria were used to exclude cases—age and no previous abuse—which left 220 cases in the study. A control group of 343 abused women was then identified, and logistic regression was used to estimate the association between identified risk factors and the risk of femicide.

The strongest demographic risk factor for intimate partner femicide was the abuser's lack of employment; similarly, a university education was a protective factor. Ethnicity was not independently associated with risk of intimate partner femicide after controlling for other demographic factors. At an individual level, the abuser's access to firearms and use of illicit drugs were strongly associated with intimate partner femicide, but excessive abuse of alcohol was not. Having left an abusive partner after living together or having a child at home who was not the abuser's biological child also increased the risk of femicide. There was a ninefold increase in the risk when a combination of separation and a highly controlling abuser existed. Threats with a weapon and threats to kill were not surprisingly also associated with higher risks of femicide.

While stalking, strangulation, abuse during pregnancy, escalating violence, suicide, perceptions of danger and child abuse were more closely associated with femicide in the bivariate analysis, this was not the case in the multivariate analysis. Campbell explains this by arguing that “these characteristics of abuse are associated with previous threats with a weapon and previous threats to kill the victim, factors which more closely predict intimate partner femicide risk” (p1092). Importantly, previous arrest of the abuser for domestic violence was associated with a reduced risk of femicide, but this correlation was not experimental and could have been caused by other factors.

**Evaluation of US Risk Assessments** Campbell’s Danger Assessment has been independently tested in two studies (Bennett et al. 2000 and Heckert and Gondolf 2004). There was some predictive validity for re-assault, rather than lethality, but there was also a high rate of false positives. Low base rates make lethality prediction very difficult, and Campbell accepted that “it is always difficult to predict, with our current statistical models and limited resources for longitudinal research, a seldom occurring event” (Campbell 2005a:1210).

Roehl et al. (2005) conducted a prospective study of the accuracy of four models—the Danger Assessment, DV-MOSAIC, the Domestic Violence Screening Instrument and the Kingston Screening Instrument for Domestic Violence. The study showed the assessments were better than chance, but there was a high level of false positives and sufficient false negatives to be of great concern. The study did not compare the tools with expert practitioners’ unstructured assessments but does conclude that the ideal approach would be to have an experienced practitioner using a validated tool. Compared with other fields, this level of prospective evaluation is extremely limited and very few have been evaluated by those other than their authors (Campbell 2005a). But what are the acceptable levels of sensitivity? Clearly as far as the victims are concerned, any false negative is problematic. The current models would appear to work in progress, and it is important that they are subject to continuing rigorous evaluation. As Kropp (2004: 682) argues, there is ‘an unresolved schism between science and practice’.

## **Risk Assessment for Domestic Homicide in England and Wales**

In England and Wales, the first predictors of domestic homicide were identified from a Metropolitan Police study of 30 domestic violence murders between 2001 and 2002 (Richards 2006). The cases were analysed and characteristics identified with the intention of “identifying certain patterns and characteristics that could indicate potential lethality” (Richards 2004, p33). Six high-risk identification markers were identified: separation, pregnancy, escalation, community issues, stalking and sexual assault. It is not at all clear how the analysis of the 30 murders led to the identification of the risk factors. For example, two cases out of 30 involved pregnancy or new birth and this has become a risk factor. Similarly, “community issues and isolation” are a factor in 14 (47%) cases and become a high-risk factor without any explanation of how such a judgement is being made and what is being counted in what is a very broadly drawn descriptor.

However, there is no comparison of this group of cases with a control group of domestic violence cases that did not lead to murder. Without such a comparison, it is

not clear how the conclusions can be valid. It is also unclear whether the 30 cases selected were the whole sample in the period or, if they were not, whether there were any criteria for selection of those 30 cases.

The initial analysis of 30 murders was supported by a subsequent analysis of 400 other assaults termed “near miss” incidents and a review of international practice (Richards 2004). The analysis of near miss incidents does not make any comparison between the 30 cases of murder and this group of 400 as a control, but it does look at 241 serious sexual assaults in some detail. It then uses a group of other offences in a similar period in what is termed a control group. This control group included all cases of grievous bodily harm, actual bodily harm, kidnap, murder and attempted murder in January and February 2001 in the Metropolitan Police Area. From the published work, it is not clear how this control group was used to draw conclusions.

In this analysis, 241 sexual assaults with a domestic violence flag were analysed. The offences had all happened in the Metropolitan Police Area between January and April 2001. In 130 cases (54%), the victims had reported previous domestic violence to the police, and in 49% of cases, the offender had a previous criminal history (but as 44 offenders could not be identified that number is likely to be higher).

This analysis is used to support a claim to corroborate the six high risk identifiers. Review of the reports, however, shows no evidence to link the six to deadly violence.

South Wales Police developed its own risk assessment from a study of 47 local homicides and relevant literature. A checklist of 15 yes/no questions was developed, and any victim scoring over seven was considered to be high risk (Robinson 2006). This resulted in a similar but different risk assessment to that developed in the Metropolitan Police, but again, there was no comparator group.

In 2009, the SPECSS+ model was developed into the DASH (Domestic Abuse, Stalking and Harassment) risk assessment model which in 2010 consisted of 27 questions covering 15 high risk factors. The complete questionnaire used in Thames Valley can be found in Thornton (2011). When the risk assessment is complete, the case is then categorized in the following way (Table 1).

There was limited evaluation of the pilot sites, and there has been very little evaluation of DASH or its predecessor SPECSS+ (Sully and Greenaway 2004; Humphreys et al. 2005; Richards et al. 2008). DASH does not claim to be able to predict violence, but that it aims to prevent violence—a claim made by other approaches which can be described as structured professional judgement. This is

**Table 1** Description of risk levels

Standard	Current evidence does not indicate likelihood of causing serious harm.
Medium	There are identifiable indicators of risk of serious harm. The offender has the potential to cause serious harm but is unlikely to do so unless there is a change in circumstances, for example, failure to take medication, loss of accommodation, relationship breakdown and drug or alcohol misuse.
High	There are identifiable indicators of risk of serious harm. The potential event could happen at any time, and the impact would be serious. “A risk which is life threatening and/or traumatic, and from which recovery, whether physical or psychological, can be expected to be difficult or impossible” Risk of serious harm (Home Office 2002 and OASyS 2006)

particularly relevant in respect of false positives where although the risk may be assessed as high, successful intervention and management may have prevented escalation and further violence. This research will undertake a brief analysis of the false positives in Thames Valley.

## Data and Methods

Two distinct methods were used for this research. The first method was to select a number of Thames Valley cases where domestic murder or serious assault had taken place (the numerator). These cases were then analysed to answer the first two research questions. In effect, this provided an evaluation of whether the SPECSS+ and DASH risk assessment protocol was accurately predicting domestic murder or serious assault in Thames Valley.

The second method was to compare the deadly domestic violence cases with a case control sample selected from defined sample of non-deadly Thames Valley violence cases.

**Numerator** Using the Cedar data base (the Thames Valley Police crime recording system), all the cases of domestic violence in the three calendar years between 1 January 2007 and 31 December 2009 were selected in the following categories: murder, attempted murder, manslaughter and grievous bodily harm with intent. One hundred and eighteen cases were identified. There were only 13 cases of murder in the period, yet the difference between murder and attempted murder may be as much about the speed and quality of medical care as the intent to harm.

The ACPO definition of domestic violence was used to flag these offences, so they were not limited to female victims or to those who had ever had an intimate sexual relationship. The cases included 51 (43%) male victims and 67 (57%) female victims.

Prior contact was then identified by searching the Cedar database for all cases where the victim had contacted Thames Valley Police since 1 January 2000 until the date of the offence in the sample. The vast majority of victims for whom there was prior contact recorded had reported offences by the same offender (46 of the 53 cases or 87%). While there are cases of victims who have been abused by many offenders, the cases in the sample of deadly violence showed that the most offending was done in the same relationship.

**Risk Assessment** The 118 cases were then examined to see what the prior risk assessment was in order to assess the accuracy of such assessments using the DASH model or its predecessors. Given that the prior contact was assessed to the year 2000 which was before DASH or its predecessor SPECSS+ was introduced in six out of 53 cases, there had been no risk assessment at the time of prior contact. SPECSS+ was introduced over a period in Thames Valley because it was first piloted in Oxfordshire in 2004 after a woman had been murdered in the police station car park as she came for help. The approach was then introduced force wide at a later date (DASH was introduced in 2009).

A further search of the Cedar data base was then carried out to identify the number of high risk assessments made in all domestic violence cases between 1 January 2007 and 31 December 2009. There were 2721 cases which had been assessed as high risk which are used as the denominator to calculate the rate of false positives.



**Case Control Samples** To select a case control sample, a sampling framework consisting of all those arrested for violence from 1 January 2007 to 31 December 2009 in Thames Valley was produced. This included offences of violence with or without a domestic violence flag, and the Home Office category of violent crime listed above was used. This was the same definition that was used to identify the numerator group. A sample was the randomly selected from this group of offenders using stratified sampling within age groups. Age was restricted to those cases where the age of the suspect fell within the range of the 120 domestic violence offenders. The age range is however very wide. The sampling framework found 49,000 cases with male offenders and over 9000 cases with female offenders. The female offender control sample was set at 100 cases, and the male offender control sample was set at 150 cases. This was approximately double the number of numerator cases in each gender block and follows the advice of Schlesselman and Stolley (1982) who suggested that a case control should be two to three times the number of the cases.

These control samples were then tested for exposure in respect of the following criteria in comparison with exposure in the deadly domestic violence cases and a bivariate analysis completed: N of all prior arrests, arrests for violence, prior convictions and cautions for all offences (and for violent crimes), age at first arrest and first conviction for all crims and separately for violence, employment, weapons, drug use, firearms, self-harm (not suicide), suicidal ideation or attempts and mental health issues.

## Findings

### Descriptive Analysis

Table 2 shows the cases by crime category, in which 51 (43%) of the victims were male and 67 (57%) female. The average age of the victim was 37, and the average age of the offender was 36.

Table 3 below shows that the types of offences varied greatly between the male and female victims. As a consequence, for women, the rate of death per attacks was 1 in 6 but for men it was 1 in 25. While some of this may be explained by the fact that men are on average stronger and larger than women, there is also research which shows that men are more likely to use a gun or a knife (Brown 1987).

**Table 2** Deadly domestic violence offence reports in Thames Valley

Offence	Number from 2007/2009	% of sample
Murder	13	11%
Attempted murder	16	13.5%
Manslaughter	1	0.8%
GBH with intent	88	74.5%
Total	118	100%



**Table 3** Outcomes of attacks by victim gender

	Murder	Attempted murder	Manslaughter	GBH with intent	All offences
Female victims	11 (85%)	15 (94%)	1	40 (45%)	67 (57%)
Male victims	2 (15%)	1 (6%)	0	48 (55%)	51 (43%)
Total	13 (100%)	16 (100%)	1	88 (100%)	118 (100%)

In 51 (43%) cases, the relationship was recorded as spouse or cohabitee (or ex); in 52 (44%) cases, as lover, boyfriend or girlfriend (or ex); in 13 (11%) cases, as parent or family; and in two (2%) cases, as civil partner (or ex).

Table 4 shows that there was prior recorded contact with police for 53 (45%) cases and there was little difference between male and female victims, 46% and 42%, respectively. In 16 of the 53 cases (30%), the most recent previous contact was a non-crime domestic incident, and of the remaining cases, 16 were actual bodily harm (30%). In 87% of the cases, the previous offences were committed by the same offender.

Table 5 shows that where there had been prior contact, nearly half of the cases involved only one prior contact for a domestic incident with the police. Overall, 76% of the sample had only one or no prior contact with police prior to the deadly violence.

Table 6 shows that the risk assessments for 21 cases (40%) with prior police contact had been classified by the responding officer as standard risk, 21 (40%) assessed as medium risk, five as high risk (9%) and there are six cases (11%) where the assessment is unknown. Only one of the 13 murders had been assessed as medium risk with the remaining assessed as standard risk. Not one of the murder cases with prior contact had been assessed as high risk. In respect of seven other murders, there was no prior contact. This initial analysis suggests a high number of false negatives for the risk assessment tool as it was applied in Thames Valley. In respect of murder, the false negative rate is 100%, and in the case of non-deadly assault, the false negative rate is 87%. The combined false negative rate is 90%.

**Table 4** Most recent prior victim contact with the police for domestic incidents ( $n = 53$ )

Victim's most recent prior domestic incident contact with the police	Number	% of sample
Non-crime domestic incident	16	30.2
Actual bodily harm	16	30.2
Assault without injury	7	13.2
Criminal damage	4	7.5
Threats to kill	4	7.5
Harassment	3	5.7
Racist incident non-recordable crime	1	1.9
GBH with intent	1	1.9
Arson with intent to endanger life	1	1.9
Total	53	100

**Table 5** Prior incident contacts between victim and police

Total number of contacts	Number of victims	Cumulative %
20+	3	2
10–19	5	7
2–9	20	24
1	25	45
Sub-total contacts	53	45
No prior domestic incident contact with police	65	55
Total	118	100

**Hampshire Comparator** A neighbouring police force (Hampshire Constabulary) provided its own data on prior police contact with cases later developing into deadly violence, using exactly the same offence definitions. No prior contact was found to have been 48% in Hampshire compared with 55% in Thames Valley.

**Case Control Samples: Males** Table 7 shows the comparisons between the male offenders in the domestic murders and serious assaults studied and the 150 males in the case control sample drawn from the broader population of those arrested for violence. There are some striking similarities between the groups but also some differences.

There are few differences of any magnitude in Table 7, but these are of great interest. One falsifies the commonly held view is that the most serious domestic assaults are a result of escalating harm and violence. In sharp contrast to the escalation claim, the evidence shows that offenders in the case control have significantly more arrests and convictions for violence than the offenders who committed domestic murder and serious assault. This evidence contradicts the hypothesis upon which much of the risk assessment tools are based.

The average age of first criminal justice contacts with the male deadly violence offenders is somewhat older than for the case control males. The average age at first arrest for violence is significantly higher in the cases than in the control sample,  $t(98.51) = -3.13$ ,  $p = 0.002$ ,  $d = 0.35$ . Also the average age at first conviction is significantly higher in the cases than in the control sample,  $t(83.47) = -2.19$ ,  $p = 0.03$ ,  $d = 0.48$ . Lastly, the age at first conviction for violence is significantly higher in the

**Table 6** Prior risk assessment

Risk assessment for victim on most recent prior domestic incident	Number	Percentage	Valid %
High	5	4.2	9.4
Medium	21	17.8	39.6
Standard	21	17.8	39.6
Prior contact but risk assessment not known or not recorded	6	5.1	11.3
N/A as no prior domestic incident contact with police	65	55.1	-
Total	118	100	100

**Table 7** Case control comparison: male offenders

	Offenders in the Thames Valley cases ( <i>standard deviation</i> ) <i>N</i> = 67	Offenders in the case control sample ( <i>standard deviation</i> ) <i>N</i> = 150	Effect size Cohen's <i>d</i>
Average number of prior arrests	9.32 (12.64)	9.29 (12.56)	0.002
Average number of prior arrests for violence	2.48 (3.32)	3.80 (6.53)	-0.25
Average number of prior convictions and cautions	6.78 (8.46)	6.69 (9.09)	0.01
Average number of prior convictions and cautions for violence	1.21(1.45)	2.77 (4.59)	-0.45
Average age at first arrest	25.69 (15.15)	22.09 (9.79)	0.28
Average age at first arrest for violence	30.49 (14.32)	24.47 (9.99)	0.35
Average age at first conviction	24.52 (14.89)	20.29 (7.43)	0.48
Average age at first conviction for violence	28.51 (12.86)	22.62 (8.20)	0.54
			Relative risk ratio
Unemployed	41.7%	38%	1.09
PNC warning for weapons	33.3%	23.3%	1.43
PNC warning for firearms	6.9%	4.6%	1.5
Drugs— this describes a very loose association with controlled drugs	13.8%	16.6%	0.83
Self-harm—other than suicide	12.5%	7.3%	1.71
Suicidal	19.4%	6%	3.23
Mental health	16.6%	9.3%	1.78

deadly assault cases than in the control sample,  $t(78.90) = -3.18, p = 0.002, d = 0.54$ . All three have a small to medium effect size.

These three comparisons show that the average age of onset of criminal career is later for those offenders who commit serious domestic assault than for those who commit all violence. Again, the evidence does not support the commonly held view that serious domestic assaults result from escalation over the years but suggest that these attacks are often much less predictable and perpetrated by those with less of a criminal history and a criminal history with a much later onset. In the cases with no prior violence, the serious assault came “out of the blue”.

In respect of the warning marks on PNC there are some differences between the two groups. In respect of most categories, the offenders in the Thames Valley deadly violence cases were more likely to have the warning marker than the case control. Thirty-three percent had a warning marker for weapons compared with 23% in the case control sample. Six percent had a warning marker for firearms compared with 4% in the case control

**Table 8** Case control comparison—female offenders

	Offenders in the Thames Valley cases ( <i>standard deviation</i> ) N = 51	Offenders in the case control sample ( <i>standard deviation</i> ) N = 100	Effect size Cohen's <i>d</i>
Average number of prior arrests	4.41 (8.98)	4.75 (12.16)	-0.03
Average number of prior arrests for violence	1.15 (1.81)	2.11 (5.88)	-0.22
Average number of prior convictions and cautions	3.17 (5.52)	3.36 (8.60)	-0.02
Average number of prior convictions and cautions for violence	0.80 (1.30)	1.64 (4.70)	-0.24
Average age at first arrest	29.24 (13.87)	28.37 (10.91)	0.06
Average age at first arrest for violence	33.21 (14.04)	30.70 (10.95)	-0.32
Average age at first conviction	24.03 (11.21)	27.64 (11.26)	0.19
Average at first conviction for violence	30.14 (14.29)	30.09 (11.57)	0.003
			Relative risk ratio
Unemployed	54.3%	62%	0.87
PNC warning for weapons	32.6%	7%	4.65
PNC warning for firearms	2.1%	2%	1.05
Drugs—this describes a very loose association with controlled drugs	13%	7%	1.85
Self-harm—other than suicide	6.5%	13%	0.5
Suicidal	13%	10%	1.3

sample. There were smaller differences in respect of markers for drugs, and the levels of employment varied slightly. However, the largest relative risk ratio is found with the PNC suicide marker, which was collected some years after the act of deadly violence. While it is not possible to separate suicide markers in these data entered on PNC after the deadly violence from those entered before, the indication of high prevalence of mental health issues seems worthy of substantial further research. This finding on the prevalence of mental health issues also replicates the review in South Wales Police (Robinson 2006).

#### *Female Block*

Table 8 shows the comparisons between the female offenders in the domestic deadly violence cases and the case control sample drawn from the broader population of females arrested for violence. In none of the cases was the difference significant and the effect sizes were small or less than small, with the exception of PNC warning for weapons. While this difference was not significant due to small sample size, the prevalence of PNC warning for weapons among the female serious assault offenders was substantial. Almost one third of these offenders had a PNC weapons warning, compared to 7% for controls. The effect size is well beyond large. The relative risk ratio is 4.75 meaning that offenders who are known

to have used weapons are nearly five times more likely to commit domestic murder and serious assault than those who do not.

## Discussion

This study finds very little success of current risk-assessment tools in providing a reliable early warning for a deadly domestic assault. It found that, by the broadest possible definition, there had been a false negative rate of 90% in forecasting a deadly attack with a “high-risk” classification.<sup>1</sup> The study found that 2721 domestic violence cases had been initially assessed as high risk. There were many repeat cases in this group, and the total number of victims was 1745. In terms of false positives, five victims out of 1745 were correctly assessed as high risk and were part of the initial numerator group of cases—a false positive rate of 99%.

It may be argued by practitioners that false positives are indeed examples of where the MARAC (Multi-Agency Risk Assessment Conference) process of developing prevention plans have succeeded. This study has not included an assessment of the MARAC process, which is largely about information sharing and safety planning on high-risk cases. It is therefore not possible to comment on effectiveness of the MARAC process, nor to assess the accuracy of the claims. While Robinson’s study of MARACs in South Wales (2006) did find that those high risk cases which were referred to MARAC did not have any contact in the subsequent 6 months in 30% of cases, it is hard to make any conclusions as there was no comparison with high-risk cases which were not referred to the MARAC.

Overall, the results suggest that DASH and its predecessor SPECSS+ do not accurately assess risk. The literature review identified that there has been little evaluation of SPECSS+ and DASH, but these results do corroborate the few studies that have taken place. Kerry Nixon, in her unpublished doctoral thesis, commented on the use of SPECSS+ in Merseyside Police, “Empirical tests of the SPECSS based risk assessment shows it to be unreliable.” Robinson’s evaluation of the South Wales Police model looked at 146 high-risk cases and checked for re-assault 6 months later. She found that only one risk factor, injury, significantly predicted repeat abuse (Robinson 2006). The findings also corroborate the high rate of false positives found in the use of structured professional judgement models in use in the USA (Bennett et al. 2000 and Heckert and Gondolf 2004).

There are many possible reasons for these inaccurate results which will be considered in this discussion. Some are in respect of the model itself— is it intrinsically flawed or just poorly implemented? More fundamentally, is the wrong research question being asked?

It was clear from the earlier analysis of SPECSS+ and DASH that the methodology was weak. The risk factors were identified from 30 cases of murder in London; there was no comparison with a broader risk pool, and it is not entirely clear why some factors were selected and others not. Yet even Jacqueline Campbell’s Danger

<sup>0</sup> A false negative has been calculated by dividing the number of cases that were not assessed as high risk and therefore not considered by the MARAC, by the total number of cases where there was both prior contact and a risk assessment completed.

Assessment model, based on rigorous research evidence, still had a high level of false positives and some false negatives (Bennett et al. 2000).

DASH is a blunt tool which assumes that there is one unitary phenomenon of domestic abuse. Yet any tool for risk assessment should be more reliable if it is disaggregated into different types of violence and relationships? There are offenders of both genders, there are same sex relationships, and there are parent child and sibling relationships. Some may be about intimate terrorism, but others will be about situational couple violence. Arguably, the tool was developed with attacks on female sexual partners in mind and based upon an understanding of intimate terrorism rather than an appreciation of the breadth of domestic violence encompassed by the ACPO definition.

However, even if the model is sound, the inaccurate results may be caused by poor implementation in Thames Valley. Originally, the risk assessment in SPECSS+ was completed by specialist staff in the Domestic Abuse Units. With the advent of DASH, this is now completed by front line officers. This raises yet another question about the risk assessment process: how appropriate is it to rely upon the professional judgement of generalist operational officers rather than domestic violence professionals?

It is very clear when talking to operational officers that they do not feel comfortable completing the risk assessment. Questions of a very personal nature are asked and officers feel very embarrassed to ask them (Macvean and Ridley 2007). In Thames Valley, the risk assessment needs to be completed for all cases of domestic incidents as well as crimes. This means that in 2009/2010, 18,386 risk assessments were completed for incidents and 12,490 risk assessments were completed for crimes. Some of these incidents may be a call from a third party such as a neighbour to a noisy argument.

However, there is a broader explanation which needs to be addressed—that this study has been asking the wrong question. The developers of SPECSS+ and DASH consistently argue that the tools are not about predicting murder but about deciding which cases are suitable for proactive intervention. The accompanying literature on DASH frequently repeats the assertion that it is not a predictive model; its purpose is preventative (Richards et al 2008, p108). Does that interpretation undermine the logical premise that prevention depends on reliable prediction?

The rejection of the word “prediction” is probably meant in a very narrow sense—that is to say the model does not aim to predict where harm will occur. But a more appropriate comparison is with a weather forecast for rain which is a prediction based on all that is known but does not mean that it will actually rain. And the parallel does not end there—as forecasts have improved as more information has been used so domestic violence prediction should improve as more information is available.

The evidence of false negatives and false positives found in this study presents a serious challenge to the current approach to risk assessment. While it may be that the implementation in Thames Valley has been problematic, it is more likely that the weak methodology of the development of these risk assessments lies at the root of this problem. Risk factors have been identified because of their presence in the numerator, but they have never been compared with their presence in a wider risk pool or denominator. The concept of risk depends upon the presence of a denominator but these risk assessments have overlooked that fact and fallen into the hindsight fallacy.

The case-control approach greatly expands the power of the logic behind making predictions. Yet in some ways, the results were disappointing. They did not identify any

clear risk factors that might make it possible to predict which cases will escalate—and only one promising prospect (suicidal indications by the offender) which require further research.

The results in respect of the presence of PNC markers for suicidal, mental health and self-harm in respect of the male offenders may be powerful clues for developing more accurate prediction models. Those who commit serious domestic assault are nearly three times more likely to be suicidal than other violent offenders. They are also nearly twice as likely to have mental health problems. This finding on the prevalence of mental health issues replicates the review in South Wales Police (Robinson 2006) and works in the USA where those who murder their partners were more than four times likely to have mental health problems than the wider pool of murderers (Zawitz et al. 1994).

As the mental health findings suggest, any risk assessment completed by police officers in respect of the risk faced by a third party is bound to have significant information gaps. However, even if we knew all the information, it would always be very hard to predict occurrences with such a low base rate.

If a structured professional judgement model is not providing an accurate forecast, might a sophisticated actuarial model involving non-linear data mining across huge data sets identify those elusive risk identifiers? If information was not limited, would it be possible to predict accurately? Given the diversity of situations and the dynamic nature of the domestic violence, it is hard to see how a model will be found that applies to all domestic violence. Both hardware and software, however, are developing rapidly, and a big data approach might be used to supplement professional judgement.

## **Conclusion: Investment in Evidence-Based Policing**

The study has raised several significant challenges for the current approach to domestic violence. Overall, there is the need to ensure that policy is based upon evidence rather than theory alone. This is a significant challenge to the National Police Chiefs' Council (NPCC, the successor to ACPO). How can a model be endorsed without a sound evidence base and lacking any form of peer review and minimal evaluation? On the basis of this model, the resources of domestic violence units have been rationed and reassurance has been given to victims. In this way, there is significant potential to undermine the legitimacy of policing. The endorsement of DASH was a leap of faith based on the work of a few analysts and police officers working in isolation from the wider research community. One of the most striking findings of this study has been the significant difference between the approach to evidence-based practice in the UK and that in the USA. In the USA, there has been a substantial piece of research with a large National Institute of Justice grant which has led to an evaluated tool.

These observations do not just apply to the approach to domestic violence but to many areas of strategy and policy development. The need to focus precious resources on evidence-based practice is more important than ever. In its desire to do something to stop the tragic loss of life to domestic violence, ACPO put its faith in the use of a risk assessment model based on scant evidence. This study has falsified the hypothesis upon which the DASH approach is based. There is an opportunity to learn from this and to develop a new approach which is based upon evidence rather than theory.



**Acknowledgements** Several former colleagues when from my period as Chief Constable of the Thames Valley Police showed immense patience as I sought data from computer systems to complete this study as an MSt thesis at the University of Cambridge. I am particularly indebted to Peter Milne and Penny Fraser for all those many spreadsheets and to Michelle Olufsen for completing many checks on the Police National Computer. Dr. Katrin Muller-Johnson and Dr. Barak Ariel at Cambridge gave sound advice on anything statistical. We also thank all those who expressed so much interest in this study, especially those who have replicated it or may do in future.

Sara Thornton is Chief Constable and Chair of the National Police Chiefs' Council for England and Wales. She previously served from 2007 to 2015 as Chief Constable of Thames Valley.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bennett, L., Goodman, L., & Dutton, M. (2000). Risk assessment among batterers arrested for domestic assault: the salience of psychological abuse. *Violence against Women*, 6, 1190–1203.
- Breedlove, R., Kentish, J., Sandker, D. and Sawtell, R. (1977). Domestic violence and the police: kansas city. In: *Domestic violence and the police: Studies in Detroit and Kansas City*, Washington DC: Police Foundation.
- Brown, A. (1987). *When battered women kill*. New York: The Free Press.
- Canadian Bureau for Justice Studies (1992). Homicide in Canada. In: *Juristat Service Bulletin*, vol 14, no 4.
- Campbell, J., Webster, D., Koziol-McLain, J., Block, C., Campbell, D., Curry, M., Gary, F., Glass, N., McFarlane, J., Sachs, C., Sharps, P., Ulrich, Y., Wilt, S., Manganello, J., Xu, X., Schollenberger, J., Frye, V., & Laughan, K. (2003). Risk factors for femicide in abusive relationships: results from a multisite case control study. *Am J Public Health*, 93, 1089–1097.
- Campbell, J. (2005a). Commentary on Websdale, lethality assessment approaches: reflections on their use and ways forward. *Violence against Women*, 11, 1206–1213.
- Hart, B. (1988). Beyond the duty to warn: a therapists duty to protect battered women and children. In K. Yllo & M. Bogard (Eds.), *Feminist perspectives on wife abuse*. California: Sage.
- Heckert, D., & Gondolf, E. (2004). Battered women's perceptions of risk versus risk factors and instruments in predicting repeated reassault. *Journal of Interpersonal Violence*, 19, 778–800.
- Humphreys, C., Thiara, R., Regan, L., Lovett, J., Kennedy, L., & Gibson, A. (2005). *Prevention not prediction? A preliminary evaluation of the Metropolitan Police Service Domestic Violence Risk Assessment model (SPECSS+) unpublished report by the Centre for the Study of Safety and Wellbeing*. London Metropolitan University: University of Warwick and the Child and Woman Abuse Studies Unit.
- Kropp, P. (2004). Some questions regarding spousal assault risk assessment. *Violence against women*, 4, 676–697.
- Krug, E., Dahlberg, L., Mercy, J., Zwi, I., & Lozano, R. (2002). *World report on violence and health*. Geneva: World Health Organisation.
- MacVean, A. and Ridley, N. (2007). An evaluation of the Thames valley response to domestic violence. unpublished report to the Chief Constable of Thames Valley Police.
- Richards, L. (2004). Homicide prevention: findings from the multi-agency domestic violence homicide reviews. *The Journal of Homicide and Major Investigation*, 2, 53–72.
- Richards, L. (2006). *Getting away with it: a strategic overview of domestic violence, sexual assault and serious incident analysis*. London: Metropolitan Police.
- Richards, L., Letchford, S., & Stratton, S. (2008). *Policing domestic violence*. Oxford: Oxford University Press.
- Robinson, A. (2006). Reducing repeat victimization among high risk victims of domestic violence. *Violence against Women*, 12(8), 761–788.
- Roehl, J., O'Sullivan, C., Webster, D. and Campbell, J. (2005). Intimate partner violence risk assessment validation study: the RAVE study practitioner summary and recommendations; validation tools for assessing risk from violent intimate partners. Report submitted to the US Department of Justice.
- Sargent, D. and Campbell, J. (2008). Assessing lethality in domestic violence cases. *Gazette* vol 70 No 2.
- Saltzman, L., Mercy, J., & Rhodes, P. (1992). Identification of nonfatal family and intimate assault incidents in police data. *American Journal of Public Health*, 82(95), 1018–1020.

- Schlesselman, J. J. & Stolley, P. D. (1982). *Case-control studies; design, conduct, analysis*. Oxford: Oxford University Press.
- Sherman, L., Schmidt, J., Rogan, D., & DeRiso, C. (1991). Predicting domestic homicide: prior contact and gun threats. In *Women battering: policy responses*. Cincinnati Ohio: Anderson Publishing.
- Sherman, L. and Strang, H. (1992). Predicting domestic violence from police records. Unpublished report to the Chief Commissioner Victoria Police.
- Sherman, L. (1992). *Policing domestic violence: experiments and dilemmas*. New York: Free Press.
- Websdale, N. (1999). *Understanding domestic homicide*. North Carolina: Northeastern University Press.
- Zawitz, M., Klaus, P., Bachman, R., Langa, P., Graziadei, H., & Harlow, C. (1994). *Violence between intimates*. Rockville: National Institute of Justice.