forensic sciences

MDPI

Article

The Domestic Abuse, Stalking and Harassment and Honour-Based Violence (DASH) Risk Assessment Instrument in Predicting Deadly or Persistent Domestic Abuse.

Karen M. Caulfield [1,2], Nicola S. Gray[2,3], Andrew Edwards[1], and Robert J. Snowden[4,*]

[1] Dyfed-Powys Police Force
[2] Swansea University
[3] Swansea Bay University Health Board.
[4] Cardiff University
* Correspondence: Robert Snowden, School of Psychology, Cardiff University, Cardiff CF10 3AT. Snowden@cardiff.ac.uk

Abstract

The DASH risk assessment scheme is used across the UK to identify and manage instances of domestic abuse. Recent studies have questioned whether the scheme can identify offenders who go on to commit further acts of domestic abuse, in particular serious violence, and therefore whether it is fit for purpose. We therefore tested the ability of DASH to predict future instances of deadly or persistent domestic abuse. From a database of ≈25,000 incidents, we compared DASH assessments which preceded an incident of "deadly violence" or was the first in a series of "persistent abuse". These groups were compared to a control group where there was no further incident of domestic abuse. The proportion of "high-risk" stratifications were approximately 5-times higher in the deadly violence group compared to the control group. Prediction accuracy assessed via signal detection theory showed DASH was a moderate predictor of deadly violence (AUC = 0.67). DASH also showed predictive accuracy in identifying persistent offenders (AUC = 0.62). While these results are encouraging and are similar in efficacy to other risk assessment schemes used in the prediction of domestic violence, the results identified that many individual items of DASH were not predictive. The inclusion of non-predictive items within DASH adds 'noise' and error into the risk evaluation. The development of a shortened version of DASH, removing these ineffectual items, was shown to have even higher predictive value for deadly violence (AUC = 0.80). We stress, however, that the role of risk assessment is not to *predict* violence per se, but to *prevent* violence via the accurate identification of dangerous perpetrators and via effective intervention and safeguarding of victims. Despite this, research such as this is imperative to evaluate if the risk assessment schemes selected by practitioners and police are fit for purpose.

Keywords: Domestic Abuse, DASH, Risk assessment.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

Citation: To be added by editorial staff during production.

1. Introduction

The DASH (Domestic Abuse Stalking and Harassment and Honour-Based Violence risk assessment: [1]) has been used by most UK Police Forces to manage their approach to risk surrounding domestic abuse cases since 2009. However, recent events have suggested that DASH has "obvious problems" [2]. In this paper we review the effectiveness of DASH

in the prediction of which perpetrators of domestic abuse go on to commit a further act of "deadly" domestic violence or are "persistent" perpetrators of domestic abuse.

The DASH Risk Assessment is a multi-agency tool used by police, health professionals, housing officers, social workers, and domestic abuse specialists. The DASH checklist includes 27 questions covering issues such as coercive control, physical and sexual violence, stalking and harassment, impact on children, and 'honour-based' violence. The DASH Risk Assessment is completed through a structured, face-to-face or telephone interview with the victim of the offence. Consent is normally required, but in cases of perceived high-risk or safeguarding situations, DASH may be completed without consent. Responses are recorded as "Yes", "No", or "Refused". Practitioners may add contextual notes to clarify or expand on answers. A risk-level is then assigned to the victim for any given incident as Standard, Medium, or High. This risk-level is determined by the responses victims gave to the DASH, officers' knowledge of previous incidents of domestic abuse, safeguarding plans already in place, and the officer's professional judgement. Typically, a high-risk category is given if 14 or more "Yes" answers are given, or if there is evidence of escalation, serious threats, or the practitioner's professional judgment indicates concern even if fewer risk factors are disclosed. Hence, DASH is designed to be objective, but the addition of professional judgment is crucial—especially if it is believed that the victim has minimized or omitted details.

The use of DASH has been reviewed by a number of authors [3-11]. While its original inception was that of a structured professional judgment scheme and authors have argued that DASH is not a predictive tool, in practice this is what it is being used as by practitioners [11]. Indeed, Ariza et al. [4] argued that putting people into risk categories (Standard, Medium or High) is a predictive task in itself. Further, the evidence that has called into question the effectiveness of DASH (see below) is based upon its poor ability to predict future instances of domestic abuse.

Previous research on DASH

Chalkley and Strang [11] identified 107 cases of high violence (domestic murders and near murders) and compared these cases to 214 control cases where "less-deadly" violence was perpetrated. Sixty-seven of the high violence cases had a previous DASH assessment. Of these 45 (67%) were not classified as high-risk at the previous assessment stage, with only 33% being assessed as high risk. The proportion of high-risk in the matched control sample is not provided.

Chalkley and Strang [11] conclude that DASH failed to predict the majority of deadly domestic violence cases (see also [8, 12]). This is clearly true but could be rectified by altering the criteria of classification of "high-risk" to a far lower threshold thus capturing far more offenders who commit high violence. Of course, such a change in threshold would come at the cost of producing more false positive results (labelling someone high risk who does not go on to commit such an act of high violence), would create a far greater workload, and would spread limited resources even more thinly. In turn, this may lead to poorer management of the high-risk offenders and more, rather than less, instances of further violence. Hence, merely changing thresholds of categorization does not "improve" the risk assessment!

To judge whether DASH (or any other risk assessment scheme) is affective in identifying high risk offenders, information is needed about performance in *both* the group that did go on to commit a violent act, and in those that did not commit further offences. Signal detection theory can then be used to provide a bias-free (one that is not contingent on a particular threshold being used to make a "high" vs "low" decision). This is done by constructing the Receiver Operating Characteristic the plot of true positives (correctly predicting the act of violence) against the false positives (incorrectly predicting an act of violence)) and calculating the area under the curve (AUC). Such analyses are routinely used

in many aspects of medical science, including prediction of violence, as well as many other areas (see [13]). AUCs range from 0 to 1.0, with scores near 0.5 indicate an instrument with no predictive value, and AUCs > 0.70 as a large effect size [14]

Turner et al. [10] looked at whether officers in a large metropolitan police force in the UK were able to distinguish cases of "serious harm" using risk ratings based on DASH. Only 5.7% of these future serious harm perpetrators were classified as high-risk. This study did use ROC analysis and found an AUC of 0.54 in identifying these future serious harm events. This suggests little value in the predictions based on DASH.

Turner et al. [9] performed similar analyses of DASH and compared these to alternate methods using information already available in police databases. They also found that the DASH assessment had little predictive accuracy (AUC ≈ 0.55) while an algorithm based upon police records was far more successful (AUC ≈ 0.75). Adding the DASH information to the predictions from the police records alone did not produce any additional increase in predictive efficacy (see also Grogger et al. [15]).

The extant literature on DASH therefore suggests little value in the DASH assessment procedure when viewed purely from the perspective of a "risk prediction tool" (see Discussion) and it is these findings that have called into question its continued use.

However, any instrument might fail if it is not being used in the correct manner. Robinson et al. [6] focused on the implementation of DASH, in particular it's use in the classification, identification and assessment of risk of domestic abuse. They carried out a UK wide exercise using in-depth fieldwork with three police forces. The DASH was used inconsistently, if at all, and there were also inconsistencies in how responses were recorded. For example, they uncovered instances of DASH being used inappropriately, such as an officer cutting and pasting from a previous assessment or only completing it if they felt a prosecution was likely. Sebire and Barling [16] used intraclass correlation tests to establish DASH's inter-rater reliability. They used information from known police history, initial police reports and responses of victims to complete the DASH. They conclude that the location and volume of cases which officers are carrying could be impacting on the decision making of officers, and that officers in a real-life setting would use other extraneous factors aside from the DASH. This suggests responses to DASH items were not the deciding factor when it comes to assessing risk.

The present study was commenced to establish if the existing use of DASH was warranted in a particular police force in the UK. In response to a Her Majesty's inspectorate review, Dyfed-Powys Police implemented changes in their handing of domestic violence cases. This included the setting up of a Vulnerability Desk to aid frontline officers in their response to domestic abuse incidents, and a commitment to "100%" in DASH completion for all reported domestic abuse incidents. Furthermore, training for all frontline staff was introduced in "Domestic Abuse Matters", to educate officers and staff on the complexities of domestic abuse from the perspective of the victim. Officers were given a separate input from internal force trainers on how to complete a DASH assessment. A Secondary Risk Assessment Unit was introduced to provide consistency of approach to risk identification, assessment, and management. When a call was received in the force control room an officer was dispatched to the scene (where appropriate). The Vulnerability Desk, made up of officers and police staff, are informed of an ongoing domestic incident. They can interrogate police systems and inform an officer in live time of any previous criminal history and other relevant information concerning those involved in the incident. This provides officers with a more comprehensive understanding of previous incidents to help inform the overall risk assessment. The information provided by the Vulnerability Desk is added to the System for Tasking and Operational Resource Management (known as STORM) log (the log of the initial call). This information provides the officer with background knowledge which they could explore during the completion of the DASH with the victim.

The officer at the scene completes a DASH with the victim and enters the responses into their Mobile Data Terminal. The attending officer then assigns a risk rating of "standard", "medium", or "high" risk according to the responses the victim gives to the DASH items, the information provided by colleagues on the Vulnerability Desk, and their own professional judgement of the situation having attended the scene. All this information is logged on the crime management system, along with the officer's rationale for their risk stratification. The officer's supervisor reviews the risk rating before the end of their shift and will either agree with officer's risk grading or change it, documenting their reasons for the change. Given these attempts at improvement in the use of DASH, Dyfed Powys Police aimed to evaluate if DASH was able to predict future cases of serious domestic violence. This was the overall objective of the current research.

2. Materials and Methods

Participants

24,889 domestic abuse crimes and incidents occurred within the Dyfed- Powys Police area between 1st January 2016 and 31st December 2019. Dyfed- Powys Police is mandated by the Home Office to collate and record Domestic Abuse data. It is analysed by Dyfed-Powys Police to profile offenders and victims and to inform intelligence-led policing and was not specifically collated to inform this study. The lead researcher (KC) had access to this data to provide this analytical function specifically for Dyfed- Powys Police and held the relevant vetting level to access the data and subsequence intelligence in relation to this study. KC did all the data cleaning and analysis to provide an anonymised database of the information used in the present analyses. As the study only involved archival data, which was anonymized, informed consent was not sought. Ethical permission for the study was given by Swansea University School of Psychology Ethics Committee (Ref: 5145, 15/04/2021).

There were 25 variables for each case detailing the victim and offender demographics, the nature and location of the incident, the offence committed, and the risk level assigned to that incident. Any occurrence which did not fit the UK government's definition of domestic abuse was removed from the sample.

Materials

Domestic Abuse, Stalking and Honour Based Violence (DASH)

Domestic Abuse, Stalking and Honour Based Violence (DASH [1]) is a risk identification and assessment model administered by police and other safeguarding and public protection agencies to record victim responses to a set of 27 main questions. The victim's responses to each question are recorded as Yes, No, or Refused. Victim responses to the DASH administered by police officers and civilian domestic abuse officers are recorded and electronically linked to each domestic abuse occurrence detailing their response to each question. Where a Yes reply was given, officers recorded additional qualitative information in relation to this response.

A risk-stratification is then assigned to the victim as either Standard, Medium, or High. This risk-stratification is determined by the victim's response to each item on the DASH, officers' knowledge of previous incidents, safeguarding plans already in place, and the officer's professional judgement.

Assignment to Groups

Figure 1 illustrates how incidents/offenders were assigned into the groups. The deadly group was defined as offenders who had committed offences which had either caused death or had the potential to cause death or serious physical harm. Those offences were: Arson Endangering Life, Attempted Murder, Assault with Intent to Cause Serious Harm, Manslaughter, or Murder. 141 offenders were identified as being in the deadly group.

The persistent group were defined as offenders with five or more domestic abuse incidents in the three and a half year research period. 859 offenders fit this definition. A random sample of 427 of these offenders were chosen for this group.

The control group were defined as offenders who had committed only one non-deadly offence between 1st January 2016 and 30th June 2019. Those who offended after this date were excluded from the study so to allow a reasonable period of time for possible recidivism. To ensure the control group were true non-recidivists for domestic abuse within the research period, each potential control group offender was checked against a list of offenders from the DPP area who were serving a prison sentence at any time between the period of 1st January 2016 and 31st December 2019 and these people were excluded from the sample if imprisoned during these dates. A total of 9,201 offenders fit the criteria for the control group. A random sample of 450 offenders were selected to serve as the control group. If the Index offence had a refused DASH, that offender was not included in the control sample. The final control group sample consisted of 432. Demographic details for the three research groups are given in Table 1.

Table 1. Demographic information of the groups.

	Control N = 432	Deadly N = 141	Persistent N = 427
Gender (% male)	69.7	66.0	88.0
Age (mean, SD)	36.2 (13.9)	35.3 (13.1)	35.2 (13.4)
White (%)	88.4	95.7	97.7
Mixed (%)	0.9	1.4	0.7
Asian (%)	1.2	2.1	0.7
Black (%)	0.2	0.0	0.2
Other (%)	0.8	0.0	0.0
Unknown (%)	9.3	0.7	0.5

Assignment of Index and Target Offence.

To determine which of the several offences committed by an offender would be subject to scrutiny to assess the effectiveness of each DASH item, a target and an index offence were identified. For the Deadly group, the target offence was the most recent occurrence of a "deadly offence". The index offence was defined as the offence immediately preceding the target offence. If the DASH in the immediately preceding occurrence was refused, the index offence was taken as the next most recent occurrence with a completed or partially completed DASH. For the Persistent group, the target offence for the persistent group is the fifth offence committed since the beginning of the study period. The index offence was the fifth offence back from the date of the target offence (inclusive). If,

however, the first offence had a refused DASH, the index offence was the second offence if it had a completed DASH, and the target offence would then move to the sixth offence, thus ensuring that a total of five offences was maintained between index and target offences. The control group were one-time domestic abuse offenders within the research period. Hence, their offence defined the index offence and there was no target offence.

Data Analysis

Cases where less than 24 of the 27 DASH items were answered were eliminated. Signal Detection Theory via the receiving operator characteristics and area under the curve (AUC) was used to assess the performance of DASH. The AUC is the widely accepted measure of predictive accuracy [13] as it is unaffected by base rates (and can therefore be used to compare performance for both rare and frequent events), different selection ratios, and non-normal distributions [14]. Magnitudes of AUC were compared via the methods described by Hanley and McNeil [17, 18] We also present the false negative rates (FNR) for comparison with previous studies and a Chi-square analysis of risk stratification across the groups. Individual items of the DASH were assessed by binary logistic regression to calculate the odds ratio between the comparison groups.

3. Results

Sixty-two offenders in the Deadly group had an index offence DASH of which 16 (25.8%) were rated as high risk. For the Persistent offenders, 427 had an index offence DASH of which 69 (16.2%) were rated as high risk. For the control group, 432 offenders had an index offence DASH of which 23 (5.3%) were rated as high risk. A Chi-square test showed this pattern of results to be highly significant, $\chi^2(4, 921) = 65.6$, $p < .001$. This was also the case when the Deadly group was compared to the Control group, $\chi^2(2, 494) = 38.4$, $p < .001$, or when the Persistent group was compared to the Control group, $\chi^2(2, 859) = 52.6$, $p < .001$. Thus, DASH ratings were far higher for Deadly and Persistent groups compared to the Control group, with a rate of high-risk ratings approximately five times greater for the Deadly group compared to the Control group.

The risk ratings were used to predict group membership using signal detection theory. When comparing the Deadly and Control groups, the AUC was 0.67; 95% CI[0.59, 0.75], $p < .001$, illustrating that the DASH risk stratification procedure performed well at predicting the Deadly group in comparison to the Control group. The analysis was repeated for the male offenders (N = 44) only as there were too few Deadly female offenders (n = 18). This analysis produced a similar result, AUC = 0.69; 95% CI[0.59, 0.79], $p < .001$. An analysis of the Persistent offenders in comparison to the Control group produced an AUC of 0.62; 95% CI[0.58, 0.66], $p < .001$ illustrating a modest ability to distinguish between Persistent offenders and Controls. The AUC for the male offenders (N = 376 persistent, N = 302 control) was 0.60; 95% CI[0.56, 0.65], $p < .001$, and for the female offenders (N = 51 persistent, N = 130 control) it was 0.64; 95% CI[0.54, 0.73], $p = .003$. A similar analysis was performed using the DASH tool in a purely actuarial manner. This was done by adding the number of endorsed items rather than using the risk ratings. Nearly identical powers of prediction were achieved.

Performance on Individual Items of the DASH

The ability of each of the DASH items to distinguish between the Deadly and the Control offenders was assessed by examining the rate of endorsement for that item (the Persistent offenders were omitted from this analysis, but these data are available in the Supplementary Materials). Table 2 illustrates these results.

Fifteen of the 27 items (55.6%) did not produce a significant difference (defined here as $p < .05$) between the two groups. This finding is consistent with previous research suggesting that only a few of the DASH items are predictive of future serious harm (see Discussion).

DASH Short Form.

The finding that over half of the DASH items are not predictive of future serious domestic violence suggests some of these items could be eliminated and the performance of the instrument might improve. Adding in non-predictive items to the risk analysis would only serve to add noise and error into the risk evaluation and make the decisions of the police officers regarding the level of risk of the offender more difficult to make. To examine this issue, we performed a multiple logistic regression in which all 27 DASH items were entered to predict deadly violence (vs control). In the resulting model only five of the items were significantly predictive and had large odds ratios. These were:

- Q27. "*Do you know if the alleged offender has ever been in trouble with the police or has a criminal history?*"
- Q24. *"Has the alleged offender had problems in the past year with drugs (prescription or other), alcohol or mental health leading to problems in a normal life?"*
- Q26. *"Has the alleged offender ever breached bail/an injunction and/or any agreement for when they can see the Injured person (IP) and/or the children?"*
- Q16. *"Has the alleged offender ever used weapons or objects to hurt the IP?"*
- Q6. *"Has the IP separated or tried to separate from the A/O within the past year?"*

The scores from these five items were then added together to produce a DASH-short form (DASH-SF) which has a range of scores from 0 to 5. Analysis of this shortened version to predict deadly violence showed an AUC of 0.80, 95% CI[.73, .87], $p < .001$. This is significantly greater than the AUC obtained from the full DASH items ($p < .001$). The DASH-SF was also a predictor of persistent offending, AUC = 0.71, 95% CI[.64, .75], $p < .001$, which was also greater than the full DASH items ($p < .001$).

Prediction from Criminological Variables

Much research (see, for example, Kroner et al. [19]) has shown that criminological variables are strong predictors of future violence. Recent research has shown that such variables are also effective in domestic abuse cases. For instance, Turner et al. [9] demonstrated the value of such variables (which produced an AUC of 0.78 in their study) and found that the addition of DASH information did not increase the magnitude of prediction produced by these criminological variables alone. We conducted a similar analysis in the present sample where four criminological variables (number of previous offences, number of previous violent offences, age of first offence, and age of first violent offence) were taken from the Police National Computer records for each offender. We used a regression analysis to predict Deadly violence (vs control) for these items and found strong effects for the number of offences and number of violent offences, but not for the age-related variables. We then produced a simple coding scheme where the number of offences and number of violent offences were coded (e.g., for violence offences 0 = 0 offences, 1 = 1 offence; 2 = 2-3 offences, 3 = 4-9 offences, 4 = 10 or more offences) to produce a score from 0 – 8. We labelled this the CV-score (criminological variable). A signal detection analysis showed that this CV-score was a strong predictor of deadly violence with an AUC = 0.82, 95% CI[0.76, 0.89], $p < .001$. CV-score was also predictive of persistent offending, AUC = 0.70, 95% CI[0.66, 0.74], $p < .001$.

These analyses show that both the DASH-SF and the CV-score are both predictive of deadly violence. We therefore tested whether together they may be more predictive than either alone. We first performed regression analyses to predict membership of the Deadly violence group where we first added one of the variables at step 1 and then the other at step 2 to see if its addition improved the model's fit. Addition of the CV-score at step 2 improved the model's fit when DASH-SF was entered at step 1 ($p < .01$). Likewise, DASH-SF improved the model's fit when entered at step 2 when CV-score was entered at step 1.

Both CV-score (Exp(B) = 1.54, $p < .001$) and DASH-SF (Exp(B) = 1.73, $p < .01$) were significant predictors in the final model. Hence CV-score and DASH-SF make non-redundant predictions to future deadly violence.

Finally, we produced a predictor variable that combined DASH-SF with CV-score by the simple addition of the two which we termed DASH-CV. This was strongly predictive of membership of the Deadly violence group (AUC = 0.84, 95% CI[0.78, 0.91], $p < .001$) although the AUC was not significantly greater than DASH-SF or CV-score alone. Despite being based on items that predicted deadly violence, DASH-CV was also strongly predictive of membership of the Persistent Offending group (AUC = 0.74, 95% CI[0.70, 0.78], $p < .001$).

Table 2. Percentage "yes" responses to each of the DASH items and the odds ratio between the deadly and control groups. #

Question	Control (% yes)	Deadly (% yes)	Deadly vs control Odds ratio [95% CI]
1. *Has the current incident resulted in Injury?*	23.6	35.1	1.75 [0.96, 3.17]
2. *Is the IP very frightened?*	37.4	41.5	1.19 [0.66, 2.13]
3. *What is the IP afraid of? Is it further injury or violence?*	29.0	35.9	1.37 [0.75, 2.51]
4. *Does the IP feel isolated from family/ friends?*	19.6	25.9	1.44 [0.74, 2.79]
5. *Is the IP feeling depressed or having suicidal thoughts?*	25.6	38.9	1.85 [1.02, 3.37]*
6. *Has the IP separated or tried to separate from the A/O within the past year?*	40.9	63.0	2.46 [1.36, 4.45]**
7. *Is there conflict over child contact?*	12.7	11.1	0.86 [0.35, 2.13]
8. *Does the A/O constantly text, call, contact, follow, stalk or harass the IP*	18.6	17.0	0.90 [0.42, 1.93]
9. *Is the IP currently pregnant or have they recently had a baby (in the past 18 months)?*	4.5	7.7	1.76 [0.57, 5.49]
10. *Are there any children, stepchildren that aren't the A/O in the household?*	18.5	12.7	0.64 [0.28, 1.48]
11. Has the A/O ever hurt the children/dependants?	5.4	5.8	1.08 [0.31, 3.78]
12. H*as the A/O ever threatened to hurt or kill the children/dependants?*	2.8	7.6	2.79 [0.84, 9.25]
13. *Is the abuse happening more often*?	24.9	43.3	2.04 [1.12, 3.73]*

14. *Is the abuse getting worse?*	25.6	43.7	2.22 [1.23, 4.03]**
15. *Does the A/O try to control everything the IP does and/or are they excessively jealous?*	35.3	47.9	1.83 [1.04, 3.22]*
16. *Has the A/O ever used weapons or objects to hurt the IP?*	8.1	27.3	3.23 [1.57, 6.65]***
17. *Has the A/O ever threatened to kill you or someone else and you believed them?*	11.8	22.6	2.19 [1.07, 4.51*
18. *Has the A/O ever attempted to strangle/choke/suffocate/drown the IP?*	12.3	29.6	3.01 [1.55, 5.84]***
19. *Does the A/O do or say things of a sexual nature that makes the IP feel bad or that physically hurts the IP or someone else?*	8.1	1.9	0.22 [0.03, 1.66]
20. *Is there any other person that has threatened the IP or that the IP is afraid of?*	5.2	7.6	1.50 [0.49, 4.62]
21. *Does the IP know if the A/O has hurt anyone else?*	12.6	23.1	2.09 [1.02, 4.28]*
22. *Has the A/O ever mistreated an animal or the family pet?*	8.4	5.8	0.67 [0.20, 2.27]
23.Are there any financial issues?	25.8	24.5	0.94 [0.48, 1.83]
24. *Has the A/O had problems in the past year with drugs (prescription or other), alcohol or mental health leading to problems in leading a normal life?*	37.8	75.5	5.07 [2.62, 9.84]***
25. *Has the A/O ever threatened or attempted suicide?*	25.2	32.7	1.44 [0.77, 2.70]
26. *Has the A/O ever breached bail/an injunction and/or any agreement for when they can see the IP and/or the children?*	3.5	14.3	4.63 [1.73,12.40]**
27. *Do you know if the A/O has ever been in trouble with the police or has a criminal history?*	35.2	78.9	6.86 [3.40,13.85]***

*$p < .05$, **$p < .01$, ***$p < .001$ in comparison to the control group.

4. Discussion

The study found that DASH had predictive validity for both deadly violence and for persistent domestic abuse. The AUC of 0.67 for deadly violence is regarded as a "moderate" effect size [14]. The present results therefore differ from previous examinations of DASH that failed to find it had any predictive validity [8-12].

Whilst the predictive validity of DASH was modest, it does not appear to be out of line with other instruments designed to predict domestic violence. The meta-analytic review of van Der Put et al. [20] looked at a range of instruments designed to predict violence in general or domestic violence. Overall, they found an AUC of 0.65 (though when correcting for possible missing studies this reduced to 0.60), with those specifically designed for domestic violence also producing an AUC of 0.65. Limiting studies to those that looked at severe/near fatal violence (as in the present study) produced an AUC of 0.66. Looking at some of the most used and studied domestic violence risk instruments across the world, they showed that the DVSI [21] had an AUC of 0.61, the B-SAFER [22] an AUC of 0.60, the DA [23] an AUC of 0.66, the ODARA [24] an AUC of 0.69, and the SARA [25] an AUC of 0.64. Hence, DASH appears to perform as well as these other domestic violence instruments and there is no obvious evidence that one instrument is a better predictor than any another.

Why might there be a discrepancy between our study showing that DASH is as predictive of deadly violence commensurate with other domestic violence risk prediction schemes, and the previous studies finding little validity for DASH? Any scheme designed to predict domestic abuse (or other forms of violence) is dependent on the correct usage of the instrument. As discussed earlier, Robinson et al. [7, 8] have shown that DASH is often not used in the correct manner, with instances of cutting and pasting from previous reports, and only completing DASH if it was felt worthwhile, etc. The current research was partly driven by the recognition of the increasing importance of the accurate assessment of the risk of domestic abuse and improvements being made in Dyfed Powys Police force to improve the quality of the police response to domestic incidents. This included greater training on the use of DASH. These improvements may have produced a higher quality of DASH, allowing it to be predictive of future incidents of domestic violence. We stress that this possible reason for DASH's efficacy in this study compared to the previous studies is speculative as a direct comparison to other forces was not conducted.

The current study also found that DASH performed moderately well in distinguishing persistent offenders from the control group. The AUC of 0.62 is also a moderate effect size [14] but is less than that for deadly violence (but this difference was not statistically significant). We also show that modifications to DASH improved this predictive validity (DASH-SF; AUC = 0.70, DASH-CV: AUC = 0.74) even though these modifications to the DASH scheme used information about item prediction of deadly violence. It seems likely that further development of DASH based on items that are related to the prediction of persistent offending may well produce even stronger predictive efficacy. No previous studies have looked at the efficacy of DASH in predicting persistent offenders. These persistent offenders may commit lower-level abuse/offences than the deadly group but do so with a frequency that could inflict serious psychological harm to their victims. As such, identification of such individuals and an understanding of the reasons for such action is important for risk management and intervention, and in safeguarding victims.

Despite the "success" of DASH in predicting deadly violence, it was notable that many of the individual DASH items were not predictive. Again, this should not be regarded as specific to the DASH scheme as it has been shown that many items in other violence risk assessment tools are also not predictive [26]. Previous studies have also shown that many of the DASH items are not useful in terms of predicting future violence [3, 10]. It is therefore of interest to examine which items are predictive.

The present study found that Q27 *(criminal history)* was the largest predictor of a deadly offender, which is consistent with the findings of Almond et al. [3] and Turner et al. [10]. However, Thornton [8, 12]and Chalkley and Strang [11] found that males in the control sample had significantly more arrests and convictions than those in the deadly domestic offender's sample. This discrepancy could be due to differences in the control

groups used in the studies. Q26 *(breach of bail or injunction)* was also a strong predictor of deadly violence.

Q24 *(Alcohol, drugs, or mental health)* had the next greatest effect on predicting deadly offenders. Chalkley and Strang [11] found more drug abuse and mental health difficulties in female deadly offenders compared to the control group. They also found self-harm to be present at a higher rate for deadly male offenders compared to the control group.

Two questions probing the type of violence previously used were also predictive of future deadly violence. Q16 *(Use of weapons)* had a large effect size in identifying deadly offenders. This finding concurs with Thornton [8, 12] that female offenders who have used weapons were almost five times more likely to be deadly domestic abuse offenders. Q18 *(strangle/choke/suffocate/drown)* had an odds ratio of 3.01 in the current study. Almond et al. [3] calculated it to have an odds ratio of 2.00 in identifying violent recidivists, compared to non-recidivists.

Q6 *(Separation from the offender)* had a large effect on predicting deadly offenders in this current study, with an odds ratio of 2.46. This is also consistent with the findings of Almond et al. [3] who found an odds ratio of 2.23 of having been separated from their victim.

Q15 *(Controlling or excessively jealous)* was found to have a medium effect on predicting deadly offenders in the current study. However, none of the other studies that looked at the DASH items found this item to have a significant effect. This may be due to the definition of harm in Turner et al. 10] which eliminated any non-physical abuse, such as coercive and controlling behaviours. Thus, it is not surprising that if we eliminate non-physical abuse and controlling behaviours in our definition of harm that an item evaluating controlling behaviour in the perpetrator may not be an effective predictor. Almond et al. [3] also did not find this item to have an effect in predicting outcome in any of their groups of offenders. The fact that the current research has highlighted controlling or excessively jealous behaviour as having a medium effect size may be due to the impact of the training of front-line officers and staff in Dyfed Powys Police on "Domestic Abuse Matters". This training programme educated staff on the various guises of domestic abuse, including the insidious nature of coercive control and how to recognize this.

Many of the DASH items were not predictive of deadly violence. Including these items would serve to add in noise and error to the evaluation and make the task of the police officers more difficult. We therefore wanted to evaluate if eliminating these non-predictive items would lead to a more accurate and effective instrument for the assessment of domestic abuse. To evaluate this, we took five of the most predictive items of deadly violence to produce a "DASH-shortform" assessment (DASH-SF). As predicted, this instrument was able to predict with a higher level of accuracy (AUC = 0.80) than the full DASH alone. We also show that merely using two pieces of information from the PNC records (the CV-score) can also produce an effective deadly-violence predictor (AUC = 0.82). Finally, if the two are combined (DASH-CV) an instrument with a large effect size (AUC =0 .84) was achieved.

Although this preliminary work is encouraging, these initial results should be taken with caution. The development and testing of a risk prediction instrument using just the construction sample can take advantage of random variations that are in its favour, and which might not be replicated in another sample of domestic abuse offenders. It is important to test the DASH-CV in a new independent sample to see if the improvement in prediction accuracy over the standard DASH indeed reflects an instrument with better accuracy or was merely due to the inflation caused by selecting items that were chosen because of the efficacy in this particular sample.

This paper focuses solely on the predictive ability of DASH. However, the point of risk assessment is not merely to predict future events but to devise a safety plan to prevent

harm, safeguard potential victims, and mitigate against any future adverse events. Indeed, DASH was originally devised as an example of a structured professional judgment where the items of the DASH are a starting point for the risk assessment. The use of a few key risk indicators, such as DASH-SF, is unlikely to help the assessor understand the needs of the victim/perpetrator and the reasons (both distal and proximal) that mediate domestic abuse. It is appreciated that most police forces, and many other agencies and sectors, have only limited resources and time to be able to perform in-depth psychological formulations to understand the drivers of domestic abuse (see for example Snowden et al. [27]), and that the use of checklists such as DASH may be needed to assess the need for intervention and to determine when urgent safeguarding action is required. We would note, however, that DASH was originally put forwards as an example of a "structured professional judgement" scheme (which uses psychological formulation as its basis), and where the data gathered from such an assessment is then subjected to professional scrutiny to determine risks and needs. However, in practice this does not appear to be how DASH is being used.

While we have emphasized that DASH, at least in this study, has a predictive accuracy commensurate with other risk assessment schemes, it is still the case that the majority of the perpetrators of deadly violence were not regarded as "high risk" in their previous DASH assessment. This "low risk paradox" often leads to the idea that the risk assessment instrument does not work. The low risk paradox has been identified in many risk assessment schemes and is known as the "prevention paradox" [28]. In most risk assessment schemes, the large majority of cases are given a "low" rating with only a few getting "high" ratings (in the present study only 5.5% of the total cases were placed in the "high risk" category). Hence, even if the classification scheme is good (but not perfect) the fact that there are many more "low risk" cases produces the finding that more people in the low risk category have the target outcome (such as heart disease, breast cancer, suicide, deadly domestic violence, etc.) even though the *proportion* of target events is far higher in the high-risk group. Hence, evaluation of an instrument based solely on the proportion of "correct" identification of those that commit the act of violence, or on a comparison of the *number* of people who commit the act in each group, is inappropriate and misleading.

Limitations and Future Directions

The study is based solely on official police data. It is appreciated that a large proportion of domestic violence incidents are not reported, with recent figures suggesting only 20-24% of such incidents being reported to the police [29]. Such a situation leads to inaccuracies in our dependent variable. It seems probable, however, that a greater percentage of "deadly violence" is likely to be reported and to be present in official records than that of lower-level violence. Thus, the effect of this low level of reporting of domestic violence would be that our control group may be contaminated by having people who did commit a further act of domestic abuse, but that this was not reported/recorded. Any effect that serves to put "noise" into the dependent variable will reduce the ability of any instrument to correctly predict group membership. Thus, it is likely that the figures presented here, and in many other studies using official police and conviction data, will be lower that the "true" figures achievable if all such incidents were reported.

The study was well powered for its intended purpose but was not powered to be able to look at group differences related to the effectiveness of DASH in identifying risk of domestic abuse. Larger scale studies are needed to examine differential predictiveness for perpetrators of different genders, in heterosexual vs non-heterosexual relationships, in different cultural settings, and in groups such as older adults [31], etc. Likewise, it would be of interest to examine different forms of domestic abuse, such as physical assaults vs coercive control, etc.

The data presented was based on cases where a DASH was completed (or at least the majority of the DASH was completed) as our aim was to examine if a completed DASH

was predictive of future domestic abuse. However, the exclusion of cases where the DASH was incomplete may have removed higher-risk and/or chaotic cases and does not therefore represent true operational conditions. Future research is needed to examine the implications of incomplete DASHs (and the reason for the non-completion) for future domestic abuse.

5. Conclusions

The DASH scheme, as implemented by the Dyfed Powys Police force, appears to be predictive of future acts of deadly violence, and to also be predictive of offenders who show persistent domestic abuse. Its levels of predictive accuracy are in accord with other risk assessment instruments for predicting domestic violence (and other forms of violence). The predictive efficacy of DASH can be improved by eliminating items that do not have predictive value, and this would reduce the workload on police officers and save resources. However, risk assessment is more than just the mere prediction of who might commit a further criminal act and greater emphasis is needed on understanding the reasons behind the incident(s) of domestic violence so that effective risk management strategies can be put in place [3]. Hence, we hope our demonstration of the efficacy of the overall DASH score will encourage its use as the starting point for a deeper structured professional judgment assessment in line with the DASH's original conception.

No risk assessment scheme is ever going to achieve perfect accuracy in its predictions, and high profile "misses" (labelling someone low risk who then commits an act of deadly violence) will continue to attract (media) attention while successful management of high-risk perpetrators will remain silent. This should not subvert efforts to produce risk assessments that are based on science and have an appropriate evidence-base.

Author Contributions: Conceptualization, CK, NG, AE & RS; methodology, KC, NG, RS; software, KC; formal analysis, KC, KC.; investigation, KC; resources, AE; data curation, KC, RS; writing—original draft preparation, KC, RS; writing—review and editing, KC, NG, AE.; supervision, NG; funding acquisition, NG, AE. All authors have read and agreed to the published version of the manuscript..

Funding: KC was funded by the Dafyd Powys Domestic Abuse Gold Group and Swansea University.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and Ethical permission for the study was given by Swansea University School of Psychology Ethics Committee (Ref: 5145, 15/04/2021).

Informed Consent Statement: Patient consent was waived due to this being archival data that was anonymized for the research team.

Data Availability Statement: The datasets presented in this article are not readily available because of the sensitive nature of the subject and on-going studies. Requests to access the datasets should be directed to RJ Snowden.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DASH	Domestic Abuse, Stalking and Harassment and Honour-based Violence risk assessment
AUC	Area under the curve
CI	Confidence interval

References

1. Richards, L. (2009). *Domestic Abuse, Stalking and Harassment and Honour Based Violence (DASH, 2009-16) Risk Identification and Assessment and Management Model IA-IP on behalf of NPCC and in partnership with Safe Lives*. https://www.dashrisk-checklist.co.uk/wp-content/uploads/2022/02/DASH-2009-2016-with-quick-reference-guidance.pdf
2. Domestic violence screening tool should be replaced, Jess Phillips says. (2025, August 26). *The Guardian*. https://www.theguardian.com/politics/jess-phillips/2021/dec/28/all
3. Almond, L., McManus, M., Brian, D., & Merrington, D. P. (2017). Exploration of the risk factors contained within the UK's existing domestic abuse risk assessment tool (DASH): do these risk factors have individual predictive validity regarding recidivism? *Journal of aggression, conflict and peace research*, *9*(1), 58-68.
4. Ariza, J. J. M., Robinson, A., & Myhill, A. (2016). Cheaper, faster, better: Expectations and achievements in police risk assessment of domestic abuse. *Policing: a journal of policy and practice*, *10*(4), 341-350.
5. Myhill, A., & Hohl, K. (2019). The "golden thread": Coercive control and risk assessment for domestic violence. *Journal of Interpersonal Violence*, *34*(21-22), 4477-4497
6. Robinson, A. L., Myhill, A., Wire, J., Roberts, J., & Tilley, N. (2016). Risk-led policing of domestic abuse and the DASH risk model. *What Works: Crime Reduction Research. Cardiff & London: Cardiff University, College of Policing and UCL Department of Security and Crime Science*.
7. Robinson, A. L., Pinchevsky, G. M., & Guthrie, J. A. (2016). Under the radar: Policing non-violent domestic abuse in the US and UK. *International journal of comparative and applied criminal justice*, *40*(3), 195-208.
8. Thornton, S. (2017). Police attempts to predict domestic murder and serious assaults: Is early warning possible yet? *Cambridge Journal of Evidence-Based Policing*, *1*(2), 64-80.
9. Turner, E., Brown, G., & Medina-Ariza, J. (2022). Predicting domestic abuse (fairly) and police risk assessment. *Psychosocial Intervention*, *31*(3), 145.
10. Turner, E., Medina, J., & Brown, G. (2019). Dashing hopes? The predictive accuracy of domestic abuse risk assessment by police. *The British journal of criminology*, *59*(5), 1013-1034.
11. Chalkley, R., & Strang, H. (2017). Predicting domestic homicides and serious violence in Dorset: A replication of Thornton's Thames Valley analysis. *Cambridge Journal of Evidence-Based Policing*, *1*(2), 81-92.
12. Thornton, S. (2011). *Predicting serious domestic assaults and murder in the Thames Valley* (Master's dissertation, University of Cambridge).
13. Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological science in the public interest*, *1*(1), 1-26.
14. Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d* and *r*. *Law and Human Behavior*, *29*, 615-620.
15. Grogger, J., Gupta, S., Ivandic, R., & Kirchmaier, T. (2021). Comparing conventional and machine-learning approaches to risk assessment in domestic abuse cases. *Journal of Empirical Legal Studies*, *18*(1), 90-130.
16. Sebire, J., & Barling, H. (2016). Assessing the assessors: An analysis of the consistency of risk grading by police when conducting domestic abuse investigations. *Policing: a journal of policy and practice*, *10*(4), 351-360
17. Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristi (ROC) curve. *Radiology 143*, 29-36.
18. Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under the Receiving Operating Characteristic curves derived from the same cases. *Radiology*, *148*, 839-843.

19. Kroner, D. G., Mills, J. F., & Reddon, J. R. (2005). A coffee can, factor analysis, and prediction of antisocial behavior: The structure of criminal risk. *International Journal of Law and Psychiatry*, *28*(4), 360-374.

20. van Der Put, C. E., Gubbels, J., & Assink, M. (2019). Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools. *Aggression and Violent Behavior*, *47*, 100-116.

21. Williams, K. R., & Houghton, A. B. (2004). Assessing the risk of domestic violence reoffending: A validation study. *Law and Human Behavior*, *28*(4), 437-455

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

23. Campbell, J. C., Webster, D. W., & Glass, N. (2009). The danger assessment: Validation of a lethality risk assessment instrument for intimate partner femicide. *Journal of Interpersonal Violence*, *24*(4), 653-674.

24. Hilton, N. Z., Harris, G. T., Rice, M. E., Houghton, R. E., & Eke, A. W. (2008). An indepth actuarial assessment for wife assault recidivism: The Domestic Violence Risk Appraisal Guide. *Law and Human Behavior*, *32*(2), 150-163.

25. Kropp, P., Hart, S., Webster, C., & Eaves, D. (1999). Manual for the spousal assault risk assessment guide (Version 3). *Vancouver: British Columbia Institute against Family Violence*.

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

27. Snowden, R. J., Holt, J., Simkiss, N., Smith, A., Webb, D., & Gray, N. S. (2019). WARRN–a formulation-based risk assessment process: Its implementation and impact across a whole country. *The Journal of Mental Health Training, Education and Practice*, *14*(6), 399-410.

28. Rose, G. (1981). Strategy of prevention: lessons from cardiovascular disease. *British medical journal (Clinical research ed.)*, *282*(6279), 1847.

29. Office for National Statistics (ONS), released 27 November 2024, ONS website, article, Domestic abuse prevalence and trends, England and Wales: year ending March 2024.

30. Franceschetti, L., Merelli, V. G., Margherita, M., Maggioni, L., Barbara, G., Kustermann, A., & Cattaneo, C. (2022). Older adult abuse in a service for sexual and domestic violence: Medico-legal implications from the experience of an Italian center. *Forensic Science International*, *338*, 111383.