College of
Policing

Vulnerability and Violent Crime Programme

Evaluation of using the Spousal Assault Risk Assessment (SARA) v3 and the Stalking Assessment and Management (SAM) tools to assess and manage risk

Key findings and implications for practice

Contents

Vulnerability and Violent Crime Programme (VVCP)

The College of Policing was awarded a grant through the Home Office Police Transformation Fund to develop the evidence base on vulnerability and serious violence. The programme focused on key areas of interest to policing, including knife crime, gangs, county lines, criminal exploitation of young people, and child sexual abuse and exploitation. This is one of nine summaries accompanying ten reports delivered as part of the VVCP.

If you have any questions about the VVCP, please email:
research@college.pnn.police.uk

Overview

In 2019 two structured professional judgement tools, one for assessing risk of domestic violence (Spousal Assault Risk Assessment (SARA) Version 3) and one for assessing the risk of stalking (Stalking Assessment and Management (SAM)) were piloted in three police forces. The intervention aimed to provide police officers and staff with tools to support them in assessing the risk and managing the behaviour of domestic violence and stalking offenders. These tools are an example of structured professional judgement that combine the strengths of professional judgement and actuarial prediction while mitigating against their respective limitations. The use of SARA and SAM was evaluated in three forces to test how offender managers would use these tools in practice. Training was provided prior to the tools being implemented to improve police understanding of risk management.

Does it work?

While the training provided before implementation of the tools was viewed positively, participants felt it did not provide them with enough knowledge to complete the two risk assessment tools or use them effectively in their work. Offender managers had too many difficulties with the forms for them to feel there was benefit in using them. There was little consistency in the way offender managers completed the forms and the risk ratings of offenders frequently differed between managers. Other barriers identified to effective use of SARA or SAM include the time commitment required to complete a SARA or SAM, the availability of information needed for their completion and offender managers feeling isolated and in need of more support.

Background

About this report

This report summarises the findings of the full independent evaluation of SARA and SAM undertaken by the University of Birmingham as part of the College's Vulnerability and Violent Crime Programme (VVCP). This summary describes how SARA and SAM work in practice and outlines key findings from the impact, process and cost analysis aspects of the evaluation. Emerging implications for practice are also discussed.

Read the full SARA and SAM reports

What is SARA and SAM?

The SARA and SAM tools are used to assess offenders' risk of domestic violence and stalking respectively. The tools were originally developed for use by psychologists or clinicians, and part of the aim of this pilot was to test their use in a police setting. These tools were piloted in West Midlands Police (WMP), Cumbria Constabulary and Lancashire Constabulary in 2019. The intervention aimed to provide police officers and staff with tools that support them in assessing the risk and managing the behaviour of domestic violence and stalking offenders. Training was provided prior to the tools being implemented to support their use. The training is important as previous research on the use of SARA identifies that assessors using the tool are intended to have advanced training (Messing and Thaller, 2013).

To identify a prioritised list of individuals for assessment, a Recency, Frequency, and Gravity (RFG) algorithm was developed. Due to similarities in the nature of domestic abuse and stalking offences it was felt that an RFG algorithm would be appropriate for prioritising individuals for assessment with both tools. The algorithm would have produced a list of suspects/offenders for risk assessment with those scoring highest being prioritised. In practice, the algorithm was only used by one force (WMP), as there were difficulties in using the approach in Cumbria and Lancashire (see Moderator section).

When completing a SAM or a SARA, officers need to consult a range of sources of information including intelligence systems, incident logs, case files (such as documents prepared for the Crown Prosecution Service or court), safeguarding information, minutes from multi-agency meetings, (for example Multi Agency Risk Assessment Conferences (MARAC); Multi-Agency Public Protection Arrangements (MAPPA); One Day One Conversation (ODOC), the Police National Computer and custody records). In Cumbria and Lancashire, information gathered from these systems was supplemented (where possible and where appropriate) with interviews with the perpetrator and/or victim(s).

How was the intervention evaluated?

The evaluation assessed whether the SARA and SAM tools helped officers to create more defensible risk assessment decisions and management plans to inform their mitigation of ongoing risk. The evaluation included contributions from police offender managers (those who were trained in the tools and those who were not), the intervention leads, experts in the use of the SARA and SAM, and partner agencies. A mixed methods approach was used to gather information on both the impact of the pilot, and the process of its implementation. A logic model[1] was developed for the SARA and SAM intervention that informed the evaluation (see **Figure 1**).

The **impact evaluation** focused on quantitative data and aimed to investigate whether the use of both tools was associated with improved outcomes (such as less reoffending). Force data was used to produce statistics that described the perpetrators in the sample, rates of reoffending (overall and DA-related) and harm caused through offending. Unfortunately, for a number of reasons, no comparison group was available for this analysis. Chief among these was that prior to the pilot, both Cumbria and Lancashire were not managing their domestic violence offenders in a structured manner, meaning no data was available. Furthermore, due to the additional operational constraints that were imposed by the COVID-19 pandemic, WMP were unable to provide a comparison sample for analysis in a timely manner.

The **process evaluation** was informed by both quantitative and qualitative data. Interviews and focus groups were conducted with offender managers and intervention leads. All of the SAMs and SARAs completed over the evaluation period were reviewed by the evaluation team. Demographic information, previous offending and reoffending data were also collected on these offenders. Information pertaining to the training and to offender managers' confidence in their judgements was also obtained and analysed.
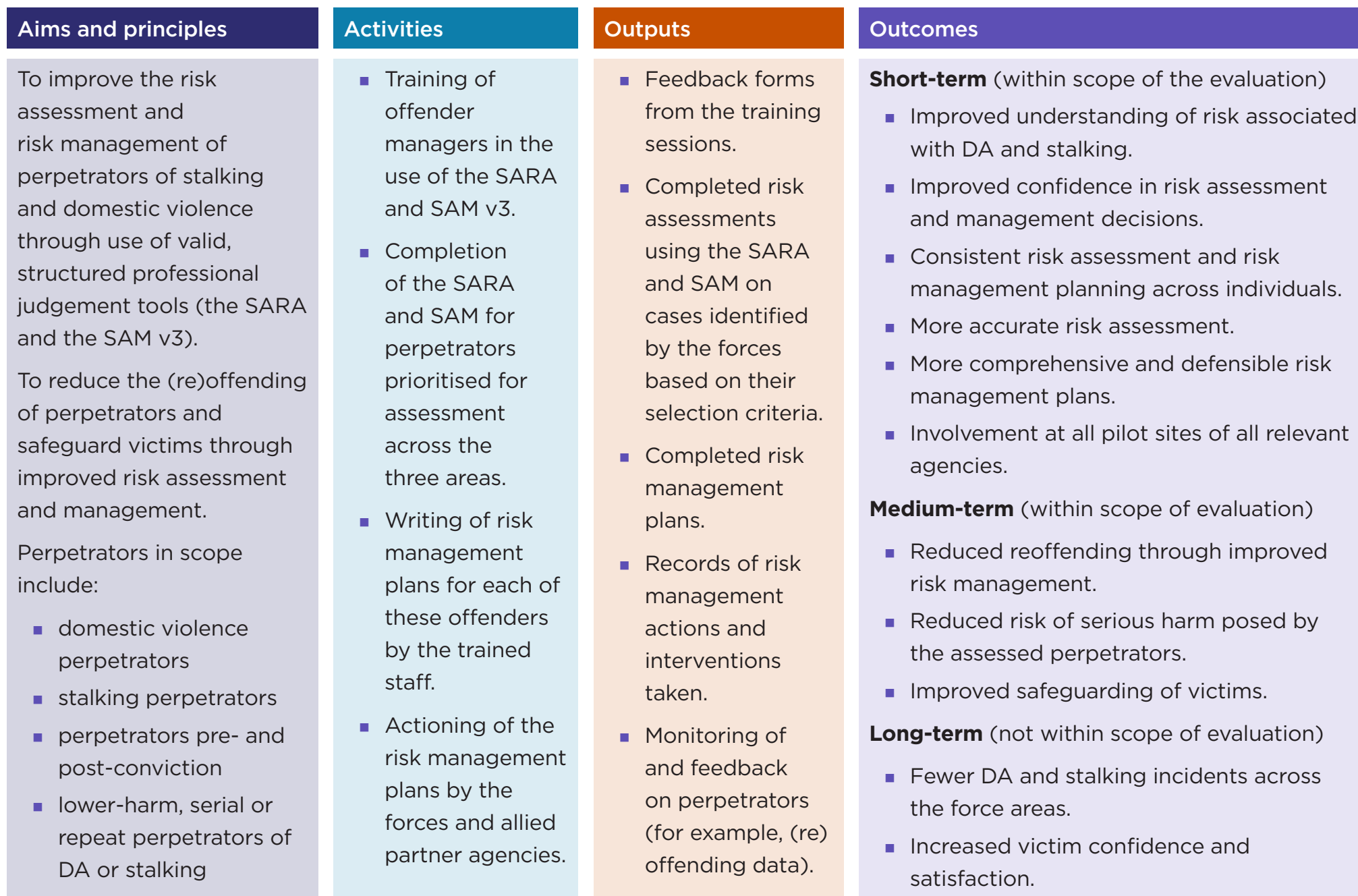
---

1   A logic model helps you think critically about the links between your problem, your intervention and your measures of success to show how and why your intervention might work. More information can be found at: **whatworks.college.police.uk/Support/Pages/Research-guidance.aspx**

At three points during the evaluation, all offender managers were asked to complete a risk assessment on a specific case to evaluate the consistency and reliability of their ratings. Reliability refers to whether a tool produces the same results at different times. This can include whether two people applying a tool to the same event or person get the same results. This type of reliability is called inter-rater reliability. As well as being compared to each other's ratings, offender managers' ratings were also compared to those of an expert rater.

Two case studies of domestic abuse were provided to assess offender managers' use of SARA (one in August 2019 and one in February 2020), and one case of stalking was provided for offender managers to conduct a SAM (October 2019). For each assessment of inter-rater reliability, a real but anonymised case study of a perpetrator of intimate partner violence (or stalking, in the case of the SAM) was provided by a police force not involved in the pilot of the SARA tool. For further details on each specific case, please refer to the main report.

Finally, it was not possible to conduct a full **cost analysis** as the intervention makes use of existing staff within the force. However, indicative costs for staff and training have been provided where appropriate.

Figure 1: SARA and SAM logic model

Aims and principles	Activities	Outputs	Outcomes
To improve the risk assessment and risk management of perpetrators of stalking and domestic violence through use of valid, structured professional judgement tools (the SARA and the SAM v3). To reduce the (re)offending of perpetrators and safeguard victims through improved risk assessment and management. Perpetrators in scope include: ■ domestic violence perpetrators ■ stalking perpetrators ■ perpetrators pre- and post-conviction ■ lower-harm, serial or repeat perpetrators of DA or stalking	■ Training of offender managers in the use of the SARA and SAM v3. ■ Completion of the SARA and SAM for perpetrators prioritised for assessment across the three areas. ■ Writing of risk management plans for each of these offenders by the trained staff. ■ Actioning of the risk management plans by the forces and allied partner agencies.	■ Feedback forms from the training sessions. ■ Completed risk assessments using the SARA and SAM on cases identified by the forces based on their selection criteria. ■ Completed risk management plans. ■ Records of risk management actions and interventions taken. ■ Monitoring of and feedback on perpetrators (for example, (re) offending data).	**Short-term** (within scope of the evaluation) ■ Improved understanding of risk associated with DA and stalking. ■ Improved confidence in risk assessment and management decisions. ■ Consistent risk assessment and risk management planning across individuals. ■ More accurate risk assessment. ■ More comprehensive and defensible risk management plans. ■ Involvement at all pilot sites of all relevant agencies. **Medium-term** (within scope of evaluation) ■ Reduced reoffending through improved risk management. ■ Reduced risk of serious harm posed by the assessed perpetrators. ■ Improved safeguarding of victims. **Long-term** (not within scope of evaluation) ■ Fewer DA and stalking incidents across the force areas. ■ Increased victim confidence and satisfaction.

How did the intervention perform?

Evidence is presented using the EMMIE framework, which was developed to help practitioners and decision-makers understand and access the evidence base quickly and easily. The EMMIE framework describes findings across five dimensions:

Effect	Impact on crime or offending	Does the evidence suggest that the intervention led to an increase or decrease in crime or offending, or that it had no impact?
Mechanism	How it works	What aspect(s) of the intervention could explain this effect?
Moderators	Where it works	In what circumstances and contexts is the intervention likely (or unlikely) to work?
Implementation	How to do it	What conditions should be considered when implementing an intervention locally?
Economic cost	How much it costs	What direct or indirect costs are associated with the intervention, and is there evidence of cost benefits?

Effect – what was the impact of the intervention?

Evidence on the overall impact of the intervention is limited by both the duration of the evaluation period and available data sources. In the absence of longer term data, the best available measures of change were used to give an indication of potential impact. Future follow-ups using longer term data would help us to understand better the overall impact of the intervention.

The results of the evaluation indicate that the training itself was viewed positively and that participants were satisfied with the content and the way in which it was delivered. Offender managers also rated their overall confidence with the tools as higher after the training than before. The offender managers found the tools helpful in terms of being able to better structure their risk assessments and management judgements.

However, while the training was well-received, participants often felt the training was too short and that it did not necessarily improve their understanding of or skill in risk assessment and management. While their confidence in the new tools rose, participants felt the training did not provide them with enough knowledge to complete the two risk assessment tools effectively or use them properly in their work. Offender managers also reported finding the tools themselves very difficult to complete, which was reflected in the inconsistency of ratings between offender managers and in the views expressed in interviews.

The tests of inter-rater reliability for the risk assessments showed that offender managers did not agree with one another in terms of the risk factors present or relevant for an offender, or in the interventions the offender needed. The statistics calculated for inter-rater agreement did not reach an adequate level for large portions of both risk assessment tools. Since reliability (of which inter-rater reliability is part) is an essential component for valid risk assessment, these findings are concerning. As the tools' designers state: 'If raters cannot agree on the presence of individual risk factors or the implications that can be drawn from them, there is little point in conducting risk assessments' (Kropp and Hart, 2000, p 109).

Across all forces in the pilot, the rate of reoffending for all offences was 54% after six months. The highest reoffending rate was in WMP and the lowest rate in Lancashire[2]. A similar pattern can be seen for DA-related reoffending, which was 40% across the whole sample after six months, with Cumbria and WMP reporting the same rate, and a lower rate found in Lancashire. The findings were also mixed when assessing whether the tools had validity in predicting future reoffending. Offenders who were rated as higher risk on the SARA summary scores went on to commit more DA-related offences in the follow-up periods but not more offences overall. Summary and total scores were also significantly associated with harm scores but not at all time points. When testing associations between different subsections of the SARA and these outcomes, the victim vulnerability scores were the only subsection scores to be significantly associated with some of these outcomes (for example, general reoffending at three months, and harm at three and six months).

2   The reoffending rate is much lower for Lancashire. It has been suggested that this may result from a large proportion of offenders who were subject to SARAs being released from prison and therefore receiving intensive supervision from other services.

This is relatively positive but it is an unexpected finding that scores on the SARA were not associated with the level of intervention planned or actioned for perpetrators, as has been found with previous studies.

Mechanism – how did it work?

The training was designed with the aim of improving understanding of risk assessment and management among officers, leading to improved skills in risk assessment and management. These improved skills should have resulted in 'better' risk management decisions, that lead on to actionable interventions that prevent reoffending.

As mentioned above, while the training was referred to favourably by attendees, participants felt the training was too short and that it did not necessarily improve their understanding of or skill in risk assessment and management. The offender managers felt that some bridging training between training in the tool and its application in practice was needed. Any such bridging training would need to be co-designed, requiring input from the tool creators and forensic psychologists, with expertise about the use of risk assessments tools, and from the police offender managers who have expertise on the manner in which such tools would be used in practice.

Offender managers also reported finding the tools themselves very difficult to complete, which was reflected in the inconsistency of ratings between OMs and in the views expressed in interviews. The tests of inter-rater reliability for the risk assessments showed that offender managers did not agree with one another in terms of the risk factors present or relevant for an offender, or in the interventions the offender needed. These findings undermine the impact of the training and the tools. Further training might lead to improved inter-rater reliability and a reduction in time taken to complete assessments.

Moderator – where did it work best?

Due to a lack of standardised operating procedures, there were several aspects of the pilot that differed between the three force areas, in particular, the way offenders were chosen for risk assessment, and the forces' capacity to manage these offenders once they had been risk assessed.

For example, it was decided that the most suitable approach to selecting cases for the intervention was the use of an RFG of offending algorithm. While this was implemented by WMP, during the evaluation it became clear that Cumbria were not using the same algorithm, and a change in IT system for Lancashire meant they could not use any algorithms at the time of the evaluation. Instead, Lancashire had to underpin their definitions of high harm/high risk that were used to select cases with the principles of RFG. Cumbria produced a prioritised list of perpetrators based on the number of DA-related calls made to them about each nominal in a 12-month period. This list was then considered by the offender management team alongside the risk ratings given to these incidents (High, Medium, Low or No Risk) when deciding which nominals to select for intervention.

All forces also took referrals from partner agencies that do not feature in police systems and thus would not be picked up by any algorithms. Cases were considered at a multi-agency meeting to determine whether police offender management was appropriate. Following this, the force would have selected those cases that met their local criteria for inclusion in the pilot.

As mentioned in the Effect section, there were a number of differences in the reoffending outcomes and inter-rater reliability across the three pilot forces. Due to the inconsistencies in the implementation of the tools and how forces processed SARA or SAM offenders (as described above), we cannot be sure if the tools were more effective in one force compared to others.

Implementation – how to do it

There were several barriers to the implementation of the SARA and SAM risk assessment tools. The findings from this report have informed a set of recommendations for other forces seeking to implement SARA or SAM. There were four elements of the intervention where implementation could be improved.

Firstly, the tools took much longer to complete than was originally expected by the intervention leads (eight hours compared to the expected two hours) and the offender managers felt that this was too much of a time commitment. While some of the changes mentioned

here could reduce the time taken, many participants felt that a simpler tool would be needed. Faster completion of tools would likely be supported by:

- a timeline to demonstrate when assessment tools should be completed

- sufficient processes for timely sharing of data required to complete the tools

- improved workforce planning to ensure enough staff were trained to meet the time commitments required, including planning for staff turnover

- allowing sufficient time to complete the assessments (and revisit them)

Secondly, tests of inter-rater reliability for the risk assessments showed that offender managers often did not agree with one another in terms of the risk factors present or relevant for an offender, or on the interventions the offender needed. Confusion about who the tools were to be used on and when, as well as which information they needed, compounded these issues as well as the issues with timeliness mentioned above. Establishing the following improvements to using SARA and SAM would improve clarity and ensure consistency:

- an agreed set of criteria of cases that are supposed to be risk assessed

- clearer and more systematic methods for identifying individuals that fit the criteria for risk assessment

- whether offender managers should complete risk assessments for offenders being managed by their colleagues, particularly where their colleagues are not yet trained in the use of SARA or SAM

- whether the risk assessments were supposed to be live documents and, if so, when they were supposed to be reviewed and where they should be stored for ease of access

- how offender managers should actively manage the offenders they have risk assessed

Thirdly, as described above, the training was generally discussed positively. However, it was criticised for being conducted too early in the pilot, as it occurred significantly before the intervention was fully implemented in each force. The timing of training delivery was recognised by the intervention leads as an issue, but had been unavoidable due to the availability of funding and the expert trainer. The evaluation team identified the following improvements to the SARA and SAM training that would ensure staff have the best chance of putting their skills into practice:

- If there is likely to be a large time gap between initial training and implementation, consider planning refresher training.

- Plan bridging training between the official SARA or SAM training and its use in practice (which covers things such as how offenders should be selected, how information should be gathered for the assessment, how to use the electronic forms and where to store them, that a risk assessment is a living document).

- Using more real-life examples of cases that are similar to those that the offender managers will encounter in practice would strengthen the officers' abilities to apply the knowledge gained during training to their role.

- Discuss with delegates, potentially in further training sessions, in more detail about how the SARA or SAM differ from other tools and explain what they can contribute above and beyond other tools.

- Consider if additional training is needed prior to the SARA or SAM training or should those trained have pre-existing qualifications. Offender managers in this evaluation suggested interviewing skills are key. Comments from the proformas and some of the inter-rater reliability findings suggest that training in some psychological concepts is needed (such as personality disorder, mental health) to use the tools effectively.

Fourthly, participants also felt the pilot was not managed well and that they lacked support from their supervisors, in part, because supervisors were not sent on the training course. This meant it was difficult for supervisors to quality control assessments of offender managers later in the pilot. Future implementations should consider extending support networks for all staff, including supervisors.

- Carefully consider who should be trained and where they will be placed post-training. Once trained, offender managers should work in units with similarly trained personnel so they have access to a peer support network and, ideally, supervisors who understand the use of the tool and can support offender managers with quality assurance.

- Consider the need for a 'support network' where trained offender managers can seek support from peers and bring challenging cases to the group. Consider having this facilitated by a trained SARA or SAM expert user.

Ongoing internal evaluation of the process as it was implemented may have highlighted these issues and led to more effective implementation. Further information around quality assurance of SARA and SAM and the use of specialist expertise can be found in the full evaluation report.

Economic cost – how much is it?

It was not possible to undertake a cost benefit analysis as part of this evaluation but indicative costs are provided. This is because the cost of the SARA and SAM intervention makes use of existing offender managers within the force, some of whom had additional responsibilities. Similar resources may exist in other forces but are working in a number of different ways.

Each SARA or SAM assessment was found to take on average eight hours to complete, and the average annual salary for offender managers was approximately £40,000. Training costs (including travel and accommodation) associated with each person trained in using SARA and SAM were in the range of £2,000 to £4,000 per force, due to the small number of officers trained in each force in this pilot.

While we cannot provide quantitative estimates of the costs of devoting this time to risk assessment in comparison to any other work, interviews with offender managers who were part of the pilot suggest that they think they are high. At present, there is insufficient data to make an assessment of benefits in terms of reduced reoffending and reduced harm.

Conclusion

While the rationale for the intervention was sound and a lot of effort was invested by the intervention leads, overall the SARA and SAM tools were not well received by the offender managers. Offender managers did not see these tools as suitable for use in their work, as they took much longer to complete than was originally expected by the intervention leads. The offender managers felt the tools required too much of a time commitment and that a simpler tool was needed. As such, the intervention cannot be deemed to be sustainable as it currently stands. There were also concerns about the reliability of the tool and how it was being completed as there were inconsistencies between practitioners' ratings. There may be alternative tools that would be more suitable for use in a policing context and, regardless of which tool is used, it is important that sufficient time is allocated to offender managers to gather information for the risk assessment and to complete the tool itself.

About the College

We're the professional body for the police service in England and Wales.

Working together with everyone in policing, we share the skills and knowledge officers and staff need to prevent crime and keep people safe.

We set the standards in policing to build and preserve public trust and we help those in policing develop the expertise needed to meet the demands of today and prepare for the challenges of the future.

college.police.uk

Follow us
@CollegeofPolice