Abstract

This study examined 91 Active Risk Management System (ARMS) assessments from four police areas across England and Wales. ARMS is a tool that guides criminal justice practitioners to assess and develop formal risk management plans based on the risks and strengths of individual clients convicted of sexual offending. This present study is particularly concerned with the application of this new tool and the quality of subsequent risk assessment as a result of police practitioner assessment. Findings indicate the quality of ARMS assessments were not to the expected standard. The study found while there were acceptable levels of detail and evidence documented by practitioners across individual areas; overall, assessor risk ratings and risk management plans were poor. This paper provides an outline of these findings, making suggestions for areas of improvement, along with recommendations for policy, practice, and research.

Key words: registered sexual offender; risk assessment; police assessment; risk management

Introduction

When working with people convicted of sexual offending criminal justice practitioners/professionals (herein practitioners), are responsible for their management and treatment are tasked with preventing recidivism by first, assessing the presence and strength of risks (Cortoni, 2009) and, more recently, the presence of protective factors (de Vries Robbé, de Vogel, Koster, & Bogaerts, 2015). Following assessment, practitioners develop supervision, treatment, or management plans to control, prevent, or reduce identified risks (Wood & Kemshall, 2010). Andrews and Bonta (2010) argue that the administration of inappropriate treatment or management strategies, for example, in the overtreatment of low-risk or under-treatment of high-risk offenders, can have a detrimental effect on the individual, including the unintended consequence of increasing risk (Andrews & Dowden, 2006; Bonta & Andrews, 2007). Thus, it is essential that to ensure people receive an appropriate level of treatment, they must first be assessed correctly.

The process of formal risk assessment and management is however, fraught with problems, such as, assessor override (where the assessor makes a clinical judgement to inflate or deflate the risk assessment score), limited resources, combining information from multiple assessments (Bonta & Wormith, 2007); inadequate assessor training and supervision, or assessor misinterpretation of classifications (Studer, Aylwin, Sribney, & Reddon, 2011). Indeed, establishing which client might recidivate is an impossible task; yet, the consequences of incorrect or poor assessment can result in harm to both clients and future victims (Craig, Browne, & Beech, 2008).

In recent decades, those tasked with assessing the needs and risks posed by people convicted of sexual offending have been guided by a growing body of literature, in which, a combination of static and dynamic assessment is carried out using structured clinical

judgment (Ireland & Craig, 2015). While much work is supported by empirical research, there is still plenty to learn. Actuarial risk assessment tools are statistically driven, usually measuring a small number of static variables, enabling practitioners to categorise and classify clients into groups of predicted risk. The most routinely used of these tools (Kelley, Ambroziak, Thornton, & Barahal, 2018) include the Static-99R (Helmus, Thornton, Hanson, & Babchishin, 2012); the Stable-2007 (Hanson, Harris, Scott, & Helmus, 2007); Violence Risk Scale–Sex Offense version (VRS-SO) (Olver, Wong, Nicholaichuk, & Gordon, 2007); and the Risk Matrix 2000 (RM2000) (Thornton et al., 2003). It is not the authors aim to discuss these tools in detail here, other papers, such as Craig et al., (2008) and Kelley et al., (2018) do an excellent job of this. For this paper, attention is given to the RM2000 measure, as up until very recently, across England and Wales; the RM2000 has been one of the main tools used to help classify potential sexual reoffending. For police practitioners, this has been the tool used to determine the allocation of resources for the management of those convicted of sexual offending (Kewley & Blandford, 2017).

While tools, such as the RM2000 can help with risk classification, and thus, enable criminal justice managers and practitioners to determine the level of resources needed to manage and intervene in a person's sentence or parole conditions; when used in isolation, there are some limitations. First, while RM2000 provides an estimate of the likelihood of recidivism, the RM2000 classification indicates that the person shares the same characteristics as a particular group of offenders who reoffended over a particular period, not that they, as an individual, have a particular chance of reoffending (Craig et al., 2008). Second, RM2000, when used in isolation does not enable the practitioner to consider dynamic risk in a person's life, which might be related to sexual offending. Although, debate is underway regarding the efficacy of dynamic factors (Ward & Beech, 2015). Finally, when using the RM2000 in isolation, practitioners are limited in their ability to develop meaningful

treatment or risk management strategies, as items measured are unchangeable through treatment or intervention.

In response to the limitations of actuarial tools, development of structured clinical judgement scales is underway. Such scales allow practitioners to provide opinion and judgement of the risk a client presents by considering the presence of both actuarial classifications and dynamic factors. Dynamic factors are either stable or acute (Beech, Craig & Browne, 2009). For example, an acute dynamic factor would be a factor, in the short term, placing the person at risk of sexually offending, these include marital breakup, bereavement, loss of employment or housing, change in substance use, and so on. Although not linked to longer-term recidivism rates, if they have a propensity to sexually offend, such factors clearly have the potential to disrupt a person's ability to cope and desist. Stable dynamic factors are factors that pervade a person's life. They include for example a sexual interest in children, sexual arousal to violent stimuli, or intimacy difficulties (Hanson, Harris, Scott & Helmus, 2007; Helmus, Hanson, Babchishin & Thornton, 2015). These factors are known to be related to sexual recidivism and by their dynamic nature, are changeable.

There are a number of structured clinical judgement scales that include the assessment of these factors, for example, the Sexual Violence Risk-20 (SVR-20) (Boer, Hart, Kropp, & Webster, 1997); the Risk for Sexual Violence Protocol (RSVP) (Laws & Kropp, 2003); and, the Assessment of Risk Manageability for Individuals with Developmental and Intellectual Limitations who Offend (ARMIDILO) (Boer, Haaven, Lambrick, & Lindsay, 2006). Again, it is not the purpose of this paper to discuss each tool in turn, instead please see Ireland and Craig, (2015) for a thorough review. Caution is required as theoretical foundations of dynamic risk factors are questionable (Ward & Beech, 2015). Dynamic factors are not yet well defined, and it is unclear if they effectively explain offending. Likewise, there is limited

clarity regarding their causal role in crime, and their role in treatment is also vague (Heffernan & Ward, 2017; Ward, 2016; Ward & Fortune, 2016).

In the development of structured clinical judgement scales and actuarial scales, interest in the role of protective factors (de Vries Robbé, de Vogel, & Douglas, 2013) and how they might assist in the assessment and management process of both general offenders and those convicted of sexual offending (de Vries Robbé, Mann, Maruna, & Thornton, 2015; Ward, 2016) is an emerging area. With it, brings debate regarding issues of definition, conceptual modelling, and of course the need for factors to be empirically tested and evidenced. However, tools are being developed and tested to help fill the knowledge gap, strengthen theory, and of course assist practitioners in developing meaningful and effective risk assessments and management plans.

One such tool is the Active Risk Management System (ARMS) (College of Policing, 2016). The ARMS tool is an instrument used to assist practitioners in assessing and managing the risks and needs of clients convicted of sexual offending. It incorporates the actuarial tool, RM2000, along with structured clinical judgment of dynamic risks and protective factors. The tool, designed for the police, probation, and prison services across England and Wales, became operational in 2013. For an account of the development of the ARMS tool, please see (Kewley & Blandford, 2017). The ARMS tool aims to guide practitioners to develop meaningful risk management plans that support the safe reintegration and risk management of people convicted of sexual offending. To understand how effective these assessments are in practice, this study aims to examine the quality of a sample of ARMS assessments completed by the police in England and Wales. Other studies have examined the effectiveness of risk management planning from a probation perspective (Bonta & Wormith, 2007; Bosker, Witteman, & Hermanns, 2013; Kewley, Beech, Harkins, & Bonsall, 2015) however, no study to date has examined the quality of ARMS assessments completed by police practitioners.

ARMS Tool

ARMS is a structured risk assessment and management planning tool designed to assess factors known to be related to sexual recidivism in adult males only. In addition to risk, the tool considers protective factors that might support the desistance process. Practitioners are required to examine 12 factors. Risk factors include: Opportunity, Sexual Preoccupation, Offence Related Sexual Interests, Emotional Congruence with Children, Poor Self-Management, Hostile Orientation, and Social Influence. The five protective factors include: Social Influences, Commitment to Desist, Intimate Relationships, Employment and Positive Routine, and Social Investment. For details regarding the theoretical basis for these factors, please see Kewley and Blandford (2017). In order for police practitioners to make a complete assessment that informs the subsequent risk management plan, practitioners are required to gather information from a number of sources. Sources include in-depth interview(s) with the client and clients' partner/family members, previous offending history, case files, court documents, prison/probation records, offender behaviour reports, etc. Practitioners then draw on the information gathered and rate the priority of each factor (Very High, High, Medium, or Low), along with the required risk management action, in response to the presence of each factor. In addition, practitioners use the RM2000 category to help determine the overall ARMS priority assessment and general level of risk management. This level of general risk management informs the subsequent risk management plan. The plan should incorporate strategies to control, rehabilitate, and support the client while living in the community, along with appropriate timeframes in which actions ought to be achieved. Review periods are dependent on individual risk and need, but if the case is managed at Multi Agency Public Projection Arrangement Level 2 or 3 cases are reviewed within a maximum of 16 weeks (level 2), and eight weeks (level 3). Completed ARMS assessments are then reviewed and signed off by a supervisor and actions are carried out by the assessor or case manager. Should

a significant change occur in the client's life, such as a further offense, loss of a job, new relationship, etc, a new review is undertaken ahead of the planned review period.

This present study is part of a larger research project, examining the effectiveness of the implementation of the ARMS tool across police areas in England and Wales. The project consists of three studies, including a) The assessment of the quality of ARMS assessment tools (current study), b) An analysis of practitioners' experiences of the implementation of the roll out of the tool (Kewley, 2017), and c) The experiences of clients being assessed by the police (in progress). This present study aims to assess the quality of ARMS assessments by examining its four key areas: The 12 risk and protective factors; the use of RM2000; the Risk Management Plan; and the Supervisor Review. We, therefore, hypothesised the following:

1) All risk factors, across all assessments, will be detailed, have evidence to support the priority rating, and have appropriate actions to address the concerns raised
2) All protective factors, across all assessments will be detailed, have evidence to support the rating, and have appropriate actions to address the concerns raised
3) All assessments will detail an RM2000 category
4) All assessments will provide a rationale for the final general level of risk management
5) All assessments will include a review period that is realistic given the risk indicated
6) All assessments will have been reviewed and signed by a supervisor

Theory and practice suggest the greatest degree of resources ought to be allocated to cases of highest risk and need (Bonta & Andrews, 2007). We, therefore, examined in further detail, cases where practitioners assessed clients to be of high or very high ARMS priority and hypothesised where an assessment is given a high or very high ARMS priority assessment:

1) The subsequent risk management strategy will include a comprehensive case summary update, a list of source material, and a realistic risk prediction

2) The subsequent risk management strategy will have a clear narrative detailing how to address the client's risk
3) The subsequent risk management strategy will provide a range of actions that address the risk identified
4) The subsequent risk management strategy will provide actions that are specific, measurable, achievable, realistic, time-bound (SMART)

Method

Four police areas participated in this study. Convenience sampling was used so the sample could be as representative of the geographically diverse areas in England and Wales as possible. Therefore, police areas selected, include one from a large city, one smaller city, and two rural areas. Full access was granted by the National Police Chiefs' Council (NPCC) and full ethical approval gained by Birmingham City University's Business, Law and Social Science Ethics Committee.

Sample and Data Collection

Four police areas provided 104 ARMS assessments, however, after the removal of duplicates and incomplete assessments the final sample was *N=91*. Three police areas provided data in the form of original ARMS assessments, and one police area manually retrieved the data from the national police Violent and Sexual Offenders register (ViSOR). It is general practice for ARMS practitioners to first complete the ARMS assessment on a Word template, then manually copy, and paste the assessment into the appropriate sections on ViSOR. Police areas were free to self-select assessments but were asked to gather assessments completed within the previous 12 months and as representative of their assessment team as possible; this served to capture the practice of all practitioners. All practitioners had completed basic ARMS training. Table 1 provides an overview of the sample broken down by police area,

ARMS Priority Rating, RM2000 (sexual and violence), and the final General Level of Risk Management rating.

Table 1. Police Area by Priority Ratings and RM2000 Classification

Police Area (*N*=91)	Type of police Area	ARMS Priority Rating	RM2000 (s)	RM2000 (v)	General Level of Risk Management
Police Area 1 (*n*=23)	City	High *n*=4 Medium *n*=7 Low *n*=8 Blank *n*=4	High *n*=5 Medium *n*=6 Low *n*=5 Blank n=7	Very High *n*=0 High *n*=7 Medium *n*=4 Low *n*=5 Blank *n*=7	Very High *n*=0 High *n*=4 Medium *n*=8 Low *n*=6 Blank *n*=5
Police Area 2 (*n*=22)	Semi-Rural	Medium *n*=8 Low *n*=14	Very High *n*=2 High *n*=5 Medium *n*=8 Low *n*=7	Very High *n*=0 High *n*=1 Medium *n*=5 Low *n*=16	Very High *n*=0 High *n*=0 Medium *n*=7 Low *n*=15
Police Area 3 (*n*=25)	Rural	High *n*=3 Medium *n*=15 Low *n*=7	Very High *n*=2 High *n*=1 Medium *n*=11 Low *n*=3 Blank *n*=8	Very High *n*=1 High *n*=3 Medium *n*=4 Low *n*=8 Blank *n*=9	Very High *n*=0 High *n*=3 Medium *n*=15 Low *n*=7
Police Area 4 (*n*=21)	Small City	Very High *n*=1 High *n*=3 Medium *n*=8 Low *n*=6 Blank *n*=3	Very High *n*=3 High *n*=6 Medium *n*=4 Low *n*=8	Very High *n*=1 High *n*=1 Medium *n*=5 Low *n*=14	Very High *n*=1 High *n*=5 Medium *n*=8 Low *n*=7

To ensure anonymity, all identifying data was removed by each police area prior to assessments being shared with the lead researcher. Data in the form of Word documents were password protected and handed in person to the first author of this paper.

Data Analysis

Due to the subjective nature of the ARMS assessments, the first two authors of this paper developed a coding framework. The framework enabled researchers to assess the quality of assessments and adherence by practitioners to the principles of the tool as determined by the theories that underpin it. To develop the framework, both authors independently attended an ARMS training event, delivered by the police, and were provided copies of training guidance

and ARMS manual. To ensure the framework was designed as close to the specifications of the ARMS tool as possible language and terminology from the *College of Policing ARMS Practitioner Guidance Manual* (2016) was used. The framework was developed in an Excel spreadsheet. This aided both the collation and organisation of the data and transfer into SPSS. Discussions between researchers were undertaken to refine understanding of the framework questions and potential answers. To provide further legitimacy and external verification of the framework, , a national ARMS trainer, and one of the original developers of the ARMS tool was consulted. Following feedback, minor changes were made. It is worth noting, that although the framework somewhat reduced the level of subjectivity, it could not be fully eliminated, due to the subjective nature of 'quality'.

Coding Framework

The coding framework was formatted to examine the four key areas of each ARMS assessment, including 1) 12 domains of risk and protection (variables 1-52), 2) RM2000 scores (variables 53-58), 3) Risk Management Plan (variables 59-74), and 4) Supervisor Authorisation (variables 75-78). It aimed to measure the quality of these areas by considering a) The *detail* documented by the assessor within each item, b) the *evidence* provided by practitioners to support the ratings given, and c) the *rationale* used by practitioners to determine the risk management plan required to reduce the risk or strengthen protective factors.

The coding framework consists of 20 factors broken down into 78 variables in total. Each factor includes between two and five questions. Factors one to 12 measure each of the 12 risks and strengths in the ARMS tool. An example of the coding tool using *Intimate Relationships* is provided in Table 2.

Table 2. Sample of the Factor Intimate Relationships and its Four Questions (37-40)

Variables	Coder Selects One of the Following Answers
37. Has the assessor detailed the clients Intimate Relationships?	None or Minimal detail/Limited detail/Adequate detail/Comprehensive and clear account
38. Has the assessor provided evidence to support the rating of Intimate Relationships	None or Minimal evidence/Limited evidence/Adequate evidence/Comprehensive evidence gathered from a number of sources
39. Does the Priority Rating match the Assessment Summary?	No/Partially/Yes
40. Do the Actions detailed address the concerns raised?	No/Partially/Moderately/Fully

Across factors 13 to 20, questions aim to evaluate a) the *accuracy* of each section completed by the assessor, b) the *presence* of evidence, and c) the *quality* of the content of the data within each ARMS assessment. Table 3 details these factors and questions.

Table 3. Factors 13 to 20 Corresponding Questions and Possible Answers

Item	Variables	Coder Selects One of the Following Answers
13. ARMS Priority Assessment	49. Do all factors have a rating?	No/Yes
	50. Has a priority rationale been documented?	No/Yes
	51. Does the rationale provide enough detail and evidence to support the ARMS Priority Assessment made?	No/Partially/Yes
	52. What is the ARMS Priority Assessment?	Low/Medium/High/Very High
14. Risk Matrix 2000 Actuarial Assessment	53. Is the Risk Matrix 2000 S Scale/Category documented?	No/Yes
	54. What is the Category?	Low/Medium/High/Very High
	55. Is the Risk Matrix 2000 V Scale/Category documented?	No/Yes
	56. What is the Category?	Low/Medium/High/Very High
	57. Has a General Level of Risk Management rationale been documented?	No/Yes
	58. Does the rationale provide enough detail and evidence to support the General Level of Risk Management made?	No/Partially/Yes
15. Case Summary Update	59. Has a Case Summary been provided?	No/Yes
	60. Does the case summary outline the risks and strengths unique to the client?	No/Yes
	61. Is too much information provided that is historical and recorded elsewhere?	No/Yes
	62. Does the summary provide the reader with sufficient context on which the resultant strategy can be understood?	No/Yes
16. Source Material	63. Has a list of Source Material been provided?	No/Yes

	64. Is this list varied and diverse?	No/Yes
	65. Are details of sources included?	No /Yes
17. Realistic Risk Prediction	66. Has a prediction about the risk posed up to next review been provided?	No /Yes
	67. Is the prediction consistent with the findings of the assessment?	No/Partially/Yes
18. Risk Management Plan	68. Has a narrative been provided about how the plan aims to address the risk prediction?	No /Yes
	69. Has the plan been supported with a range of actions that address the risk prediction?	No/Partially/Yes
	70. Are the actions SMART?	No /Yes
19. Review	71. Has the assessor detailed the review period set?	No /Yes
	72. Is the review period realistic and acceptable given the risk and need indicated throughout the assessment?	No /Yes
	73. Has the assessor identified how the case is to be reviewed including MAPPA Level and Lead agency responsibility?	No /Yes
	74. Has the assessor identified how the case is to be reviewed including Lead agency responsibility?	No /Yes
	75. Has the assessment been reviewed by the supervisor?	No /Yes
	76. Has the assessment been signed by the supervisor?	No /Yes
	77. Has the supervisor recorded any discrepancies or problems in the assessment?	No /Yes
20. Contradictions or Errors	78. Are there any contradicting statements or pieces of evidence made throughout the assessment?	No /Yes

Coding

Quality standards were agreed and applied by the two coders (authors one and two) using the following process. First, each coder independently coded a sample of five ARMS assessments using the coding framework. Coders then met to compare and discuss ratings. Where there was disagreement, coders referred to the ARMS Guidance Manual and either a) the framework descriptors were strengthened, or b) an agreement was met regarding the discrepancy. This process was repeated with a further five assessments. At this point, agreement of quality standards between the coders was strong. Thus, coders analysed the remaining sample independent of each other. The third author, who was not involved in the coding process and therefore, blind to any coding discussions or development of the tool, analysed the data. She carried out an inter-rater reliability, using Intraclass Correlation Coefficient (ICC), and found 4 variables had moderate agreement, 36 variables had good agreement and 38 variables had excellent agreement between the two coders. The ICC benchmark indicates that <.05 = poor agreement, 0.5-0.75 = moderate agreement, 0.75-0.9 = good agreement and >0.9 = excellent agreement (Koo & Li, 2016).

Results

Findings of each of the four sections of the coding framework (12 risk and protective factors, RM2000, Risk Management Plan, and Supervisor Review) are presented here.

Risk and Protective Factors

It was hypothesised all risk and protective factors across all assessments would be detailed, have evidence to support the priority rating, and have appropriate actions to address the concerns raised. Analysis of the degree to which practitioners provided both the detail and the evidence to support their ratings of each factor was undertaken and is presented in Table 4.

Table 4. Detail and Evidence Provided by Practitioners Across all Risk and Protective Factors

		None or Minimal Detail	Limited Detail	Adequate Detail	Comprehensive and Clear Account
				N (%)	
Degree of Detail	Opportunity	18 (19.8)	46 (50.5)	26 (28.6)	1 (1.1)
	Sexual Preoccupation	26 (28.6)	36 (39.6)	27 (29.7)	2 (2.2)
	Offence Related Sexual Interests	37 (40.7)	32 (35.2)	21 (23.1)	1 (1.1)
	Emotional Congruence with Children	23 (25.3)	31 (34.1)	37 (40.7)	0 (0.0)
	Hostile Orientation	17 (18.7)	35 (38.5)	39 (42.9)	0 (0.0)
	Poor Self-Management	9 (9.9)	25 (27.5)	55 (60.4)	2 (2.2)
	Anti-Social Influences	55 (60.4)	28 (30.8)	8 (8.8)	0 (0.0)
	Pro-Social Network	21 (23.1)	48 (52.7)	22 (24.2)	0 (0.0)
	Commitment to Desist	7 (7.7)	43 (47.3)	41 (45.1)	0 (0.0)
	An Intimate Relationship	25 (27.5)	34 (37.4)	32 (35.2)	0 (0.0)
	Employment/Positive Routine	9 (9.9)	30 (33.0)	52 (57.1)	0 (0.0)
	Social Investment	18 (19.8)	29 (31.9)	44 (48.4)	0 (0.0)
Degree of Evidence	Opportunity	20 (22.0)	45 (49.5)	25 (27.5)	1 (1.1)
	Sexual Preoccupation	27 (29.7)	38 (41.8)	24 (26.4)	2 (2.2)
	Offence Related Sexual Interests	37 (40.7)	33 (36.3)	21 (23.1)	0 (0.0)
	Emotional Congruence with Children	29 (31.9)	26 (28.6)	36 (37.4)	0 (0.0)
	Hostile Orientation	20 (22.0)	34 (37.4)	37 (40.7)	0 (0.0)

Poor Self-Management	11 (12.1)	28 (30.8)	50 (54.9)	2 (2.2)
Anti-Social Influences	58 (63.7)	25 (27.5)	8 (8.8)	0 (0.0)
Pro-Social Network	26 (28.6)	48 (52.7)	17 (18.7)	0 (0.0)
Commitment to Desist	10 (11.0)	48 (52.7)	33 (36.3)	0 (0.0)
An Intimate Relationship	27 (29.7)	34 (37.4)	30 (33.0)	0 (0.0)
Employment/Positive Routine	10 (11.0)	33 (36.3)	48 (52.7)	0 (0.0)
Social Investment	19 (20.9)	28 (30.8)	44 (48.4)	0 (0.0)

The level of detail documented by the practitioners was found to be of a poor standard overall, with only Poor Self-Management (risk) and Employment/Positive Routine (protective) factors being coded as acceptable in more than 50% of the 91 cases. The weakest factor in terms of detail was Anti-Social Influences (risk) as it was only coded as being of an adequate level of detail in nearly 9% (n=8). The quality of evidence provided by practitioners was also to a poor standard overall with only the Poor Self-Management (risk) and Employment/Positive Routine (protective) factors being coded as adequate in more than 50%. The weakest factors were Anti-Social Influences (risk) and Pro-Social Network (protective) which were only coded as having adequate evidence in nearly 9% (n=8) and 19% (n=17) respectively.

An analysis of the degree to which coders agreed with the priority rating practitioners gave to each domain was also undertaken. Table 5 highlights areas of Poor Self-Management, Employment/Positive Routine, and Social Investment were the only areas coders able to support and agree with the practitioners' rating.

Table 5. Frequencies for 'Does the Priority Rating Match the Assessment Summary?'

Factor	No	Partially	Yes

	N (%)		
Opportunity	43 (47.3)	28 (30.8)	20 (22.0)
Sexual Preoccupation	52 (57.1)	24 (26.4)	15 (16.5)
Offence Related Sexual Interests	65 (73.6)	12 (13.2)	14 (15.4)
Emotional Congruence with Children	50 (54.9)	7 (7.7)	34 (37.4)
Hostile Orientation	45 (49.5)	11 (12.1)	35 (38.5)
Poor Self-Management	32 (35.2)	16 (17.6)	43 (47.3)
Anti-Social Influences	71 (78.0)	9 (9.9)	11 (12.1)
Pro-Social Network	52 (57.1)	25 (27.5)	14 (15.4)
Commitment to Desist	30 (33.0)	33 (36.3)	28 (30.8)
An Intimate Relationship	44 (48.4)	17 (18.7)	30 (33.0)
Employment/Positive Routine	26 (28.6)	18 (19.8)	47 (51.6)
Social Investment	35 (38.5)	12 (13.2)	44 (48.4)

Likewise, Table 6 details practitioners' inconsistencies and insufficient actions needed to address the risk and need of their clients.

Table 6. Frequencies for 'Do the Actions Detailed Address the Concerns Raised?'

Factor	No	Partially	Moderately	Fully
		N (%)		
Opportunity	57 (62.6)	21 (23.1)	7 (7.7)	6 (6.6)
Sexual Preoccupation	62 (68.1)	19 (20.9)	5 (5.5)	5 (5.5)
Offence Related Sexual Interests	67 (73.6)	15 (16.5)	4 (4.4)	5 (5.5)
Emotional Congruence with Children	56 (61.5)	11 (12.1)	5 (5.5)	19 (20.9)
Hostile Orientation	61 (67.0)	12 (13.2)	6 (6.6)	12 (13.2)
Poor Self-Management	47 (51.6)	16 (17.6)	16 (17.6)	12 (13.2)
Anti-Social Influences	81 (89.0)	6 (6.6)	3 (3.3)	1 (1.1)
Pro-Social Network	72 (79.1)	14 (15.4)	2 (2.2)	3 (3.3)
Commitment to Desist	57 (62.6)	20 (22.0)	6 (6.6)	8 (8.8)
An Intimate Relationship	62 (68.1)	18 (19.8)	3 (3.3)	8 (8.8)
Employment/Positive Routine	50 (54.9)	18 (19.8)	10 (11.0)	13 (14.3)
Social Investment	59 (64.8)	12 (13.2)	5 (5.5)	15 (16.5)

When combining these results, we arrived at an overall quality rating (see Appendix 1 for criteria used). Table 7 summarises the overall quality of assessments in terms of the degree of detail, evidence, priority rating, and appropriate actions documented across the 12 areas of risk and protective factors. It was found while the quality of detail and evidence for each of the factors was acceptable, the quality in relation to appropriate ratings and actions was however found to be low.

Table 7. Overall Quality of Assessments Across the Twelve Risk and Protective Factors

Item	Low	Acceptable	Acceptable with Caution	Excellent	Neither High nor Low
			N (%)		
Quality of detail provided	10 (11.0)	34 (37.4)	37 (40.7)	0 (0.0)	10 (11.0)
Quality of evidence provided	13 (14.3)	35 (38.5)	33 (36.3)	0 (0.0)	10 (11.0)
Does the rating match the summary	40 (44.0)	7 (7.7)	26 (28.6)	9 (9.9)	9 (9.9)
Do the actions address concerns*	72 (79.1)	11 (12.1)	0 (0.0)	2 (2.2)	-

*Six cases where not allocated a quality rating as a result of their inconsistences

Risk Matrix 2000 Sexual (s) and Violence (v) Rationale

As part of an ARMS assessment, practitioners are required to consider the rating as determined by the RM2000 tool. Practitioners must a) document both the RM2000 sexual and violence rating, and b) provide a rationale for their arrival at the overall priority assessment when considering the RM2000 category. These features, like the 12 risk and protective factor domains are a key part of the ARMS tool as they help to inform the development of the practitioners' subsequent risk management plan. We, therefore, hypothesised that all assessments would detail an RM2000 category. We found 16% (n=15) did not document an RM2000 (s) and 17% (n=16) did not document an RM2000 (v). Of the assessments without

an RM2000 (s), they also had no RM2000 (v) documented. We also expected all assessments would provide a rationale for the final general level of risk management. Yet, only 13% (*n*=12) of the 91 assessments provided enough detail and evidence to support the assessor's chosen General Level of Risk Management. It was found that 22% (*n*=20) of assessments failed to provide a General Level of Risk Management at all; with 26% (*n*=24) of assessments only partially justifying their rating and a further 38% (*n*=35) not providing enough detail or evidence to support the General Level of Risk Management made.

Review Period

All ARMS assessments are required to detail a review date; review periods are outlined in the training manual (College of Policing, 2016) however, if required and justified, cases can be reviewed outside of this guidance. It was hypothesised all assessments would include a review period that is realistic given the risk indicated. Findings did not support this hypothesis, 33% (*n*=30) of assessments either did not include a review period at all or provided one that did not match the given risk level indicated.

Supervisor Review

All ARMS assessments must be reviewed by an appropriate supervisor, as noted by the College of Policing (2016) this process ensures "the quality of the assessment has met the required standard" (p.23). For this hypothesis, the data from police Area 1 is excluded, as this police area did not send full and complete ARMS assessments. However, of the remaining 68 assessments, 31% (*n*=21) were not reviewed by a supervisor and 51% (*n*=35) were not signed by a supervisor.

High and Very High ARMS Priority Assessments

Of the sample, only 11% (*n*=10) were identified as high ARMS priority and 1% (*n*=1) as very high. It was expected, therefore, in these assessments, the subsequent risk management

strategy would include a comprehensive case summary update, a list of source material, and a realistic risk prediction, as required by guidance set out by the College of Policing (2014). Of the high ARMS priority assessments, one failed to provide a comprehensive case summary update, a list of source material, or a realistic risk prediction. The remaining nine all provided a case summary update, however, six of these did not provide the reader with enough context to understand the resultant strategy, two did not provide a list of source material and seven did not provide a realistic risk prediction. Consistent with the findings of the assessment, the single very high ARMS priority assessment did contain a case summary update, a list of source material, and a realistic prediction about the risk posed.

It was also expected that high or very high ARMS priority assessments' subsequent risk management strategies would have a clear narrative detailing how practitioners would address the client's risk and need by providing a range of actions to address the risk and need identified. Of the high ARMS priority assessments, two failed to provide a clear narrative, two cases partially detailed how they aimed to address the client's risk and six gave a clear narrative of their strategy. In terms of the actions, six failed to provide a range of actions, three provided a good range with one only partially detailing the range of actions. For the single very high ARMS priority assessment, this was found to provide the reader with insufficient context to understand the subsequent risk management strategy; likewise, it only partially detailed how the assessor would address the client's risk and did not provide a range of actions to address the risk prediction.

Finally, it was expected risk management strategies would provide actions that are specific, measurable, achievable, realistic, time-bound (SMART); as outlined in the practitioners training and training manual. It was found only one of the high ARMS priority assessments had SMART actions in which the risk prediction was addressed. The remaining

nine high and one very high ARMS priority assessments did not provide actions (if at all) that were SMART.

Discussion

Our study set out to explore the quality of police ARMS assessments. We sampled 91 assessments from four police areas across England and Wales and found for each of the four key areas (12 risk and protective factors, RM2000, Risk Management Plan, and Supervisor Review) the quality of assessment was less than satisfactory. This is a concern, as poor risk assessment impacts on resultant risk management strategies, which in turn, affects clients and potential victims. Poor risk assessment, as found in our sample, has the potential to cause either an over-prediction of risk, meaning those who present minimal risk receive unnecessary sanctions and treatment, which of course, is resource intensive, costly, and interferes with civil liberties. While on the other hand, an under-prediction of risk, and failure to apply resources to those who need them, can result in further harm through repeat offending; the economic costs are huge but the social and psychological costs to victims and their families are unimaginable.

Our findings support those of the early ARMS pilot study by Nicholls and Webster (2014). In their evaluation, they found practitioners both lacked consistency when rating different cases and struggled to identify what actions they could realistically set to help address the risk and needs of their client. Our findings also show considerable discrepancies across coder's ratings and those of practitioners, as well as practitioners failing to include meaningful actions to support risk and needs identified. In addition, this present study was able to reach beyond the scope of the Nicholls and Webster pilot study, as we also examined subsequent risk management plans. It is perhaps unsurprising we found problems in the quality of ARMS assessments, as in our earlier study, exploring police practitioners'

experiences of completing ARMS assessments (Kewley, 2017); practitioners reported several issues with both the tool, and their own capacity to complete the assessment. Practitioners felt their training and supervision was insufficient to such a degree they did not feel fully equipped with the appropriate skills and knowledge needed to carry out assessments. Nicholls and Webster (2014) called for greater training for practitioners, we echo this recommendation but perhaps go further in that the lack of supervision of officers and practitioners carrying out this task is likely to contribute to the quality of ARMS assessments.

Although the aim of this study was not to examine why assessments were deemed problematic, using insight gained from the previous assessor study by the first author, along with findings from Nicholls and Webster's (2014) pilot study, and feedback from national ARMS trainers, we are able to make suggestions for improving the quality of future ARMS assessments. Before we offer any advice, however, it is worth noting several limitations to our study. We sampled only four police areas, there are 43 police areas across England and Wales and as such, the sample may not be truly representative. Although we found no area-specific issues, this is still a small sample. Our sample also included assessments that had been completed prior to the roll out of any supervisor/supervisor ARMS training; therefore, we cannot be confident the gatekeeping process was as rigorous as it may be today. Since the collection of this data, supervisors have attended ARMS training. It is, therefore, possible the quality of ARMS assessments, have improved. A final limitation of this study is that police areas' self-selected their sample for analysis. As such, we cannot be confident each areas sample does not include assessments carried out by the same assessor, and is thus, not representative of the area, but rather a few assessors.

To help improve the quality of ARMS assessments, we feel the following policy and practice recommendations might help. First, it is important to establish why poor quality levels were found in this study. For example, this may be a result of assessor override

(Wormith, Hogg, & Guzzo, 2010), noncompliance as a result of unmotivated practitioners, high stress levels or resourcing issues (Schafer & Williamson, 2018), a lack of assessor knowledge or poor training (Luong & Wormith, 2011), or a poorly constructed risk assessment tool etc. It was reported in an earlier study (Kewley, 2017) practitioners felt unskilled and undertrained when completing these assessments; this is likely to impact their motivation and ability to complete quality assessments. However, further exploration is required.

Our second recommendation is for policy to mandate practitioners and ARMS managers to regular supervision. Formal supervision carried out by a trained specialist would provide practitioners and managers with advice and support, and help their professional development. Supervision would allow for discussion of complex cases, helping practitioners formulate risk assessment and management planning, that is more robust. A third recommendation is that a quality audit framework would help practitioners understand what a good quality ARMS assessment should look like; further, it would serve as a benchmark to guide practitioners and supervisors when discussing cases. Our fourth recommendation calls for clearer guidance in the ARMS manual and training strategy in its current form it may serve as a barrier (Hochstetler, Peters & Copes, 2017). This should include a clearer rationale of ratings for each of the 12 risk and protective factors. Plus, for the section considering social networks, separating out both prosocial and antisocial influences would assist practitioners to consider both sides of the client's social network, and thus, improve the detail and evidence for these domains; as would some detailed examples of what should be included as actions to address the risk identified.

In addition to policy and practice, we have some research recommendations we would like to make. Since the initial Nicolls and Webster (2014) pilot study examining the reliability and validity of the tool, ARMS has undergone some changes in terms of both the

tool and the training. A further validation and reliability study are needed to reflect these changes. Furthermore, the experiences of those subjected to an ARMS assessment would be of value for further improving the quality of the tool. It is believed those being assessed hold unique knowledge and insight into the ARMS process that only those subjected to such a process would have. As such, their unique experiences may hold value for improving the quality of assessment.

Conclusion

Criminal justice practitioners assessing the risk posed by people convicted of sexual offending, must draw upon empirically tested risk and need assessment tools and models, to help inform and develop robust risk management strategies. This approach both helps prevent further harm and helps reintegrate clients safely back into the community. Indeed, when risk is already known to criminal justice practitioners, they have a duty to mitigate and plan for any escalation or change in a client's risk. For police practitioners, at least, the development and use of ARMS is encouraging, as although actuarial tools (such as RM2000) outperform unstructured clinical judgement (Bengtson & Långström, 2007) when used in combination with a structured clinical judgement (such as ARMS), prediction is strengthened with the addition of a dynamic assessment (Craig, Beech, & Harkins, 2009). Nevertheless, if assessor input is faulty or the structured tool itself is poorly constructed, unvalidated and thus, problematic, subsequent action and risk management planning will without a doubt, be flawed.

References

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct*. (Fifth ed.). New Providence, NJ: Matthew Bender & Company.

Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology, 50*(1), 88-100.

Beech, A., Craig, L. A., & Browne, K. (2009). *Assessment and treatment of sex offenders: A handbook*. Chichester: John Wiley & Sons Ltd.

Bengtson, S., & Långström, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: A direct comparison with sex offenders in Denmark. *Sexual Abuse, 19*(2), 135-153.

Boer, D., Hart, S., Kropp, P., & Webster, C. (1997). Manual for the Sexual Violence Risk-20. Burnaby, British Columbia: The Mental Health. *Law, & Policy Institute, Simon Fraser University*.

Boer, D. P., Haaven, J., Lambrick, F., & Lindsay, W. R. (2006). *Assessment of risk manageability for individuals with developmental and intellectual limitations who offend (ARMIDILO).* Paper presented at the 9th Conference of the International Association for the Treatment of Sexual Offenders (IATSO), Hamburg, Germany.

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation.* Public Safety Canada. Accessed from: http://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/rsk-nd-rspnsvty/rsk-nd-rspnsvty-eng.pdf

Bonta, J., & Wormith, J. S. (2007). Risk and need assessment. In G. McIvor & P. Raynor (Eds.), *Developments in social work with offenders* (pp. 131-152). United Kingdom: Jessica Kingsley Publishers.

Bosker, J., Witteman, C., & Hermanns, J. (2013). Do intervention plans meet criteria for effective practice to reduce recidivism? How probation officers forget about social capital and basic needs. *European Journal of Probation, 5*(1), 65-85.

College of Policing. (2016). *ARMS Practitioner Guidance Manual v.2.0*.

Cortoni, F. (2009). Factors associated with sexual recidivism. In A. R. Beech, L. A. Craig & K. D. Browne (Eds.), *Assessment and treatment of sexual offenders: A handbook* (pp. 39-52). Chichester: Wiley and Sons Ltd.

Craig, L. A., Beech, A. R., & Harkins, L. (2009). The predictive accuracy of risk factors and frameworks. In A. R. Beech, L. A. Craig & K. D. Browne (Eds.), *Assessment and treatment of sex offenders: A handbook* (pp. 53-74). Chichester: Wiley and Sons.

Craig, L. A., Browne, K. D., & Beech, A. R. (2008). *Assessing risk in sex offenders: A practitioner's guide*. Chichester, England: John Wiley & Sons Inc.

de Vries Robbé, M., de Vogel, V., & Douglas, K. S. (2013). Risk factors and protective factors: A two-sided dynamic approach to violence risk assessment. *Journal of Forensic Psychiatry & Psychology, 24*(4), 440-457.

de Vries Robbé, M., de Vogel, V., Koster, K., & Bogaerts, S. (2015). Assessing Protective Factors for Sexually Violent Offending With the SAPROF. *Sexual Abuse: A Journal of Research and Treatment, 27*(1), 51-70.

de Vries Robbé, M., Mann, R. E., Maruna, S., & Thornton, D. (2015). An Exploration of Protective Factors Supporting Desistance from Sexual Offending. *Sexual Abuse: A Journal of Research and Treatment, 27*(1), 16-33.

Hanson, R. K. (2018). Long-Term Recidivism Studies Show That Desistance Is the Norm. *Criminal Justice and Behavior, 45*(9), 1340-1346.

Hanson, R. K, Harris, A., Scott, T., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project*. Public Safety Canada.

Hanson, R.K, & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*(6), 1154-1163.

Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism*. Solicitor General Canada.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: a comparison of three actuarial scales. *Law and Human Behavior, 24*(1), 119.

Heffernan, R., & Ward, T. (2017). A comprehensive theory of dynamic risk and protective factors. *Aggression and Violent Behavior, 37*, 129-141.

Helmus, L., Hanson, R. K., Babchishin, K. M., & Thornton, D. (2015). Sex offender risk assessment with the Risk Matrix 2000: validation and guidelines for combining with the STABLE-2007. *Journal of Sexual Aggression*, 21(2), 136-157

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse*, 24(1), 64-101.

Hochstetler, A., Peters, D. J., & Copes, H. (2017). Supervisory experience and data points: implementation barriers, risk assessment and correctional professionals. *Criminal Justice Studies, 30*(3), 240-256.

Ireland, C., & Craig, L. A. (2015). Adult Sexual Offender Assessment In D. P. Boer, R. Eher, L. A. Craig, M. H. Miner & F. Pfäfflin (Eds.), *International perspectives on the assessment and treatment of sexual offenders: Theory, practice and research* (pp. 13-33): John Wiley & Sons.

Kelley, S. M., Ambroziak, G., Thornton, D., & Barahal, R. M. (2018) How Do Professionals Assess Sexual Recidivism Risk? An Updated Survey of Practices. Sexual Abuse, 0(0), available at: https://journals.sagepub.com/doi/pdf/10.1177/1079063218800474

Kewley S, Beech A, Harkins L, Bonsall, H. 2015. Effective risk management planning for those convicted of sexual offending. *Journal of Aggression, Conflict and Peace Research*, 7(4), 237-257

Kewley, S., & Blandford, M. (2017). The development of the active risk management system *Journal of Criminal Psychology*, 7(3), 155-167

Kewley S. (2017). Policing people with sexual convictions using strengths-based approaches *Journal of Criminal Psychology*, 7(3), 168-182

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine, 15*(2), 155-163.

Laws, D. R., & Kropp, P. R. (2003). The promise and the peril of sex offender risk assessment. In T. Ward, D. R. Lawson & S. M. Hudson (Eds.), Sexual deviance: Issues and controversies pp. 207-225. Thousand Oaks, CA: Sage Publications

Luong, D., & Wormith, J. S. (2011). Applying risk/need assessment to probation practice and its impact on the recidivism of young offenders. *Criminal Justice and Behavior*, 38(12), 1177-1199.

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment, 22*(2), 191-217.

Nicholls, C.M. and Webster, S. (2014), "Sex offender management and dynamic risk: pilot evaluation of the Active Risk Management System (ARMS)", Ministry of Justice Analytical Series, available at: https://www.gov.uk/government/publications/sex-offender-management-and-dynamic-risk (accessed 30 April 2018).

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale-Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological assessment*, 19(3), 318.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent Offenders: Appraising and Managing Risk*. Washington DC: American Pscyhological Association

Schaefer, L., & Williamson, H. (2018). Probation and Parole Officers' Compliance With Case Management Tools: Professional Discretion and Override. *International Journal of Offender Therapy and Comparative Criminology*, 62(14), 4565-4584

Studer, H. L., Aylwin, S. A., Sribney, C., & Reddon, J. R. (2011). Uses, misues and abuses of risk assessment with sexual offenders. In D. P. Boer, R. Eher, L. A. Craig, M. H. Miner & F. Pfäfflin (Eds.), *International perspectives on the assessment and treatment of sexual offenders: Theory, practice and research* (pp. 193-212). Sest Sussex: John Wiley & Sons.

Thornton, D. (1997). *Structured anchored clinical judgement.* Paper presented at the NOTA Annual Conference, Southampton.

Thornton, D., Mann, R. E., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing and combining risks for sexual and violent recidivism. In R. Prentky, E. Janus, M. C. Seto & A. W. Burgess (Eds.), *Annals of the New York Academy of Sciences* (Vol. 989, pp. 225-235): Wiley Online Library.

Ward, T. (2016). Prediction and agency: The role of protective factors in correctional rehabilitation and desistance. *Aggression and Violent Behavior, 32*, 19-28.

Ward, T., & Beech, A. R. (2015). Dynamic risk factors: a theoretical dead-end? *Psychology, Crime & Law, 21*(2), 100-113.

Ward, T., & Fortune, C.-A. (2016). The role of dynamic risk factors in the explanation of offending. *Aggression and Violent Behavior, 29* 79-88.

Wood, J., & Kemshall, H. (2010). Effective multi-agency public protection: Learning from the research. In K. Harrison (Ed.), *Managing high-risk sex offenders in the community: risk management, treatment and social responsibility*. Cullompton, UK: Willan.

Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, 39(12), 1511-1538

Appendix 1. Criteria for coding over quality of risk and protective factors

For Items 1 and 2, if the coder coded 'no/minimal' for 5 or more domains this is deemed as low quality; if the coder coded ‘limited’ or ‘adequate' for 8 or more domains (separately) this was recorded as acceptable quality; if the coder coded ‘limited’ or ‘adequate' for 8 or more domains (cumulatively) this became ‘acceptable with caution’; and if the coder coded 'comprehensively' for 10 or more domains this was recorded as ‘excellent’ quality. For Item 3, if the coder coded 'no' for 5 or more domains this is ‘low’ quality; if the coder coded ‘partially’ for 6 or more domains this was ‘acceptable’ quality; if the coder coded ‘partially’ or ‘yes’ for 6 or more domains cumulatively this quality was deemed ‘acceptable with caution’ and if the coder coded 'yes' for 10 or more domains this was recorded as ‘excellent’ quality. For Item 4, if the coder coded 'no' for 5 or more domains this was ‘low’ quality; if the coder coded ‘partially’ or ‘moderately for 6 or more domains (either separately or cumulatively) this was ‘acceptable’ quality and if the coder coded ‘fully' for 10 or more domains this was recorded as ‘excellent’ quality. During the analysis of quality, several cases were found to be neither high nor low quality, in that they did not meet the requirements of any quality rating. Additionally, in Item 4 some cases were deemed to have too many inconsistencies to be given a quality rating at all.