



Selby, Sarah Elizabeth (2018) *Inter-rater reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) amongst mental health professionals*. D Clin Psy thesis.

<https://theses.gla.ac.uk/30869/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk



University
of Glasgow

**Inter-Rater Reliability of the Structured Assessment of Violence Risk in
Youth (SAVRY) Amongst Mental Health Professionals
and
Clinical Research Portfolio**

Sarah Elizabeth Selby
BA (Hons), MSc, MSc

*Submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical
Psychology*

Institute of Health & Wellbeing
College of Medical, Veterinary & Life Sciences
University of Glasgow

October 2018

Acknowledgements

I would first like to thank my supervisors, Dr Hamish McLeod and Dr Jennifer McDonald, your wisdom and guidance has enabled me to navigate this thesis adventure. Thank you to Hamish for all your advice and support in conducting my research. Special thanks to Jennifer and Dr Martin Gallagher for providing their time to deliver the training events, and Jennifer Mill for dealing with the admin. Thank you to all the professionals who took time out of their busy clinical work to attend the training, this would not have been possible without your participation.

I would not have got to this point without the kindred spirits who have made this journey with me, you know who you are – thank you for being there (in the library and beyond). To my friends in the world outside DClinPsy thank you for spurring me on and sticking around!

Finally, and most importantly, thank you to my family. To the Maverick who has facilitated the last 3 years, thank you from the bottom of my heart. And to my children, Emma, Tom and Laura, thank you for putting up with everything, I love you so much.

Table of Contents

Acknowledgements	3
List of Appendices.....	6
Chapter 1: Systematic Review	7
<i>Gender Differences in Protective Factors for Youth Violence: A Systematic Review.....</i>	<i>7</i>
Abstract	9
Background	11
Methods.....	14
Results	18
Discussion	30
References	34
Chapter 2: Major Research Project.....	39
<i>Inter-Rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) Amongst Mental Health Professionals.....</i>	<i>39</i>
Plain English Summary	41
Abstract	43
Introduction	45
Method	50
Results	56
Discussion	67
Conclusions	71
References	73

List of Appendices

Appendix 1 International Journal of Forensic Mental Health Instructions for Authors	78
Appendix 2 Database search terms	85
Appendix 3 Outcome Measures for Violence and Protective Factors	88
Appendix 4 SAVRY items.....	91
Appendix 5 Ethical Approval Letter.....	92
Appendix 6 NHS Research & Development Approval	93
Appendix 7 Vignette Example.....	94
Appendix 8 Expert Rater Study Information Sheet	97
Appendix 9 Expert Rater Consent Form.....	100
Appendix 10 Expert Rater Vignette Feedback Form.....	101
Appendix 11 Participant Information Sheet	107
Appendix 12 Participant Consent Form.....	110
Appendix 13 Participation Information Questionnaire.....	111
Appendix 14 Calculating Percentage Agreement	115
Appendix 15 Rounding to the Nearest Integer (Method and Impact)	117
Appendix 16 Results of Percentage Agreement with Expert.....	119
Appendix 17 Comparison of male and female vignettes	120
Appendix 18 MRP Proposal	122

Chapter 1: Systematic Review

Gender Differences in Protective Factors for Youth Violence: A Systematic Review

*Submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical
Psychology*

October 2018

Address for Correspondence:

Sarah Selby

Institute of Health & Wellbeing

College of Medical, Veterinary & Life Sciences

University of Glasgow

Administration Building

Gartnavel Royal Hospital

1055 Great Western Road

Glasgow G12 0XH

Email: s.selby.1@research.gla.ac.uk

Chapter word count (including references): 6561

Prepared in accordance with the requirements for submission to the *International Journal
of Forensic Mental Health* (see Appendix 1).

Abstract

Various individual, social, and environmental factors have been identified as protective against adolescent violence perpetration. However, less is known about how protective factors may differ between males and females. A previous review in 2012 identified some gender differences. This review updates the previous findings by exploring protective factors in more detail and providing a critical appraisal of the research. Across the studies there was a tendency for family factors, such as family connectedness and parental monitoring, to be more protective for females. However, this is a tentative conclusion due to variable quality and methodological limitations of the included studies.

Keywords: protective factors, violence, adolescence, gender, systematic review

Background

Youth violence is a “global public health problem” (WHO, 2016) with wide-ranging consequences including problems for longer-term health, social functioning and behaviour, and cognitive and academic performance of both perpetrators and victims (WHO, 2018). Direct and indirect exposure to violence during childhood and adolescence has been linked with poorer physical and emotional wellbeing (Salzinger, Feldman, Stockhammer, & Hood, 2002) and a reduced capacity to form healthy relationships (Herrenkohl, Kosterman, Mason, & Hawkins, 2007).

Risk and protective factors

Much research has focused on increasing understanding of the factors associated with youth violence to inform interventions. Through the 1990s the risk prevention paradigm was highly influential and many studies focused on the factors associated with an increased violence likelihood (Farrington, Ttofi, & Piquero, 2016). Two robust risks to emerge were previous victimization and exposure to violence. Calvert (1997) reported that violent adolescents experienced at least 50% more victimization than their peers. Gorman-Smith and Tolan (1998) showed that seeing someone being threatened or attacked increased the likelihood of the adolescent being violent themselves, even when controlling for previous levels of aggression. Exposure to violence is proposed to increase the likelihood of violence perpetration through social learning, by normalising violence and reducing inhibitions to act violently (Bandura, 1986).

As understanding of risk factors has developed, many researchers have argued that investigation of protective factors that mitigate violence risk is also necessary (Pollard, Hawkins, & Arthur, 1999). Taking a strengths based approach to intervention by focusing

on protective factors is argued to increase offenders' engagement motivation and also aids professionals by reducing burnout and increasing the likelihood of successful interventions (Powell, 2015). Despite agreement on the importance of protective factors, there is a lack of consistency in their definition (Lösel & Farrington, 2012). Some define them as variables that predict a low likelihood of offending, or the opposite of a risk factor. For example, the presence of a stable emotional bond to a caregiver may be directly protective, but the lack of such a bond is a risk factor (Werner & Smith, 2001). However, for some variables, this is not the case due to the nonlinear relationship between the variable and outcome measure. For example, disadvantaged communities are only a risk for delinquency. Loeber, Farrington, Stouthamer-Loeber and White (2008) argue that protective factors should be differentiated between variables that interact with risk factors to reduce the likelihood of offending ("interactive protective factors") and variables that predict a low probability of offending in those at risk of offending ("risk-based protective factors"). Variables that solely predict a low probability of offending are defined as "promotive factors" (Loeber et al., 2008). To be most informative, protective factors should be studied using a prospective, longitudinal design to enable information on predictors and outcomes to be more easily identified (Lösel & Farrington, 2012). Cross-sectional or retrospective longitudinal designs have utility for identifying potentially relevant factors, however they cannot make causal links due to their inability to account for the temporal ordering of protective and risk factors.

Gender

Being male has consistently been associated with an increased involvement in violence (e.g. Saner & Ellickson, 1996). The nature of violence also differs across gender. For example, although assaults against same-sex peers account for half of assaults perpetrated by girls and boys, girls are next most likely to assault family members whereas boys are

more likely to assault strangers (Franke, Huynh-Hohnbaum, & Chung, 2002). The impact of gender on protective factors for violence is less clear. Several studies of protective factors have included violence as part of a delinquency variable making it difficult to draw conclusions about violence specifically (e.g. Hartman, Turner, Daigle, Exum & Cullen, 2009). Losel and Farrington (2012) suggested that "...the patterns of protective factors in women/girls seem to be partially different (e.g. greater relevance of relationship issues)."

Rationale for review

A review of gender differences in risk and protective factors in 2012 highlighted the need for more research to understand the influence of protective factors and their possible differential impact at different stages of development (Baxendale, Cross & Johnston, 2012). Whilst the review discussed protective factors, there was a lack of detail on factors included such as their definition and measurement. No evaluation of research quality was included. The current review will build on Baxendale et al.'s review (2012) to explore protective factors in more detail, include recent publications, and critically appraise the research.

Research Questions

This systematic review will seek to answer the following questions:

1. What types of violence outcome are measured?
2. What types of protective factors are measured?
3. Do protective factors for adolescent violence differ by gender?
4. Do protective factors moderate the impact of risk factors for violence over time?

Methods

This review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2009).

Search Criteria

Inclusion and exclusion criteria were specified using the PICOS framework (Centre for Reviews and Dissemination, 2009), see Table 1. Violence is defined as “the intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community, that either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment or deprivation” (Krug, Mercy, Dahlberg, & Zwi, 2002). Sexual and dating violence were treated as discrete groups distinct from general violence and therefore excluded, consistent with other studies e.g. Jolliffe & Farrington, 2007.

Table 1
Inclusion and Exclusion Criteria Using PICOS Framework

PICOS	Inclusion Criteria	Exclusion Criteria
Population	Young people aged 0-18 Males and females	Adults 19 years or older Males or females
Intervention/Exposure	Protective factors (operationalised and measured)	Risk factors only
Comparator/Outcome	Gender comparison or engagement/non-engagement in violence (self-report or official records) AND presence/absence of protective factors	
Study Design	Cohort, longitudinal, prospective, cross-sectional	Intervention studies, reviews, narratives, commentaries, editorials, other types of opinion papers
Other Factors	English language publication Peer reviewed	

Search Terms

Computerised searches were completed on the 2nd April 2018 using the following databases: MEDLINE (Ovid), EMBASE (Ovid), PsycINFO (EBSCOhost), CINAHL (EBSCOhost), ASSIA (Proquest), Web of Science and Google Scholar. Searches included publications from the commencement of the database to the search date. Based on the PICOS framework, subject headings and keywords were identified and used to search for the following concepts: youth, gender difference, protective factors, violent offending. Search terms were combined using BOOLEAN operators 'AND' and 'OR'. See Appendix 2 for search terms.

Study Selection

The review process is shown in Figure 1. Articles were systematically screened by reading the title and abstract to determine eligibility. Records were excluded if they did not meet the inclusion criteria specified in Table 1. If the title and abstract provided insufficient information the full article was read and a final set of papers for review were identified.

Quality Assessment

Included studies were quality rated using the Crowe Critical Appraisal Tool (CCAT) (Crowe, 2013). The CCAT was developed following an evaluation 44 existing critical appraisal tools. It has established construct validity (Crowe & Sheppard, 2011). Reliability examined using intraclass correlation coefficients shows good levels of agreement (consistency=0.83, absolute agreement=0.74) (Crowe, Sheppard, & Campbell, 2012). The CCAT can be used for quantitative and qualitative studies and comprises eight criteria on reporting and methodological issues: Preliminaries, Introduction, Design, Sampling, Data Collection, Ethical Matters, Results, and Discussion. Risk of bias is assessed in line with

the requirements of the PRISMA guidance. Each criterion is scored separately out of 5, all criteria scores are summed for a total score (maximum=40) which is converted to a percentage as specified in the CCAT manual. CCAT does not specify qualitative descriptions of scores. Studies were rated by the author and an independent rater to check reliability. There was agreement on 75% of items, disagreements were resolved by discussion and consensus ratings were used.

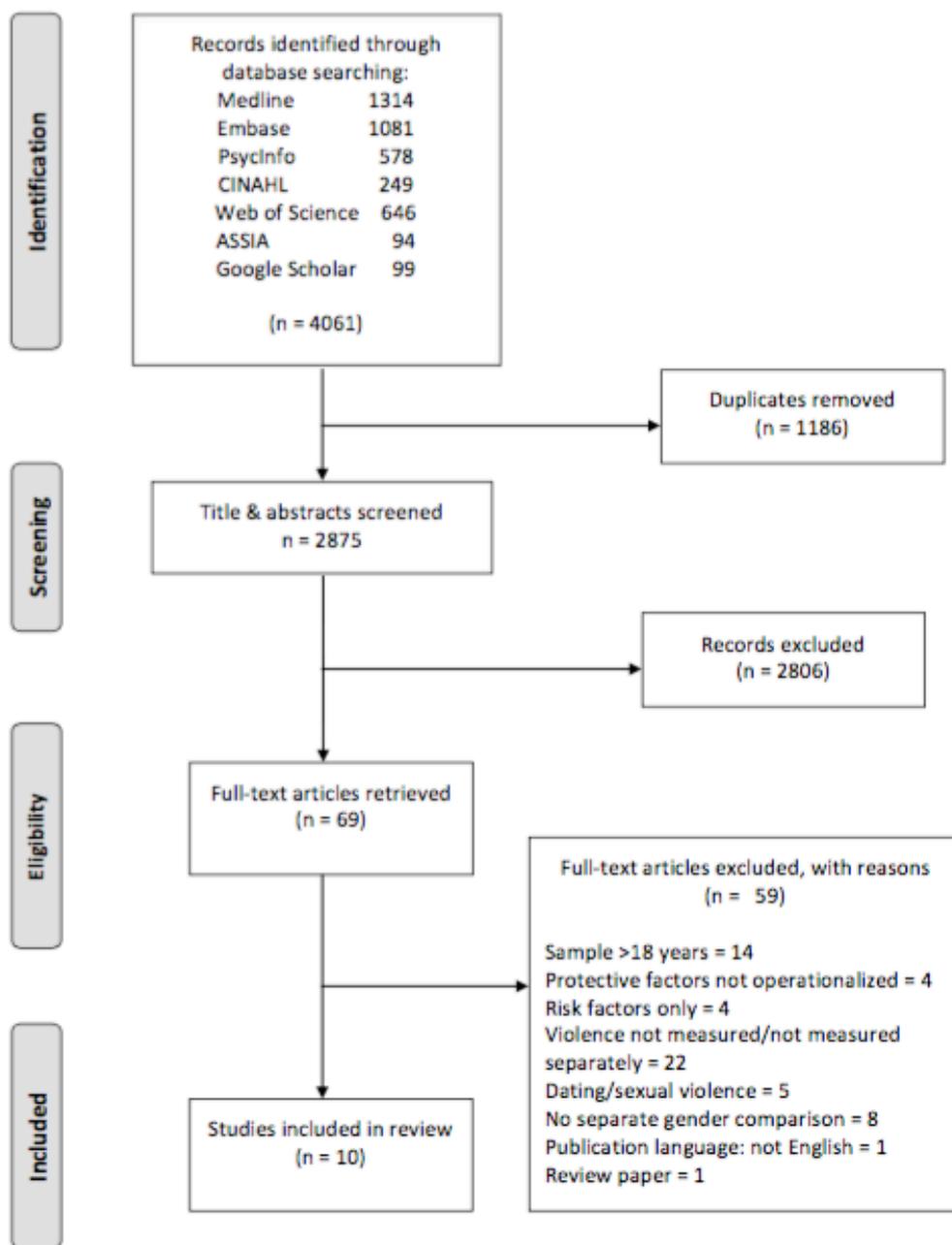


Figure 1. PRISMA flow chart for systematic review.

Results

Overview of studies

Characteristics of the 10 included studies are reported in Table 2. Eight studies were conducted in the USA, one in Hong Kong, and one in Bosnia and Herzegovina. They were published between 2000 and 2017.

Table 2
Characteristics of Included Studies

Reference	Country/ Sample	Study Design/ Data Collection	Violence Outcome	Protective Factors
Blum, J., Ireland, M., Blum, R.W. (2003)	USA 17,036: 8290 males, 8836 females High school students (12–18 years) Ethnicity not stated	Cross-sectional Add Health survey data from Wave 1 (1995) Self-report	Interpersonal violence including weapon use and gang fighting	<ul style="list-style-type: none"> • Individual (higher grade point average) • Environmental (school connectedness) • Family (family caring/ connectedness, parent expectations for school completion)
Boyas, J.F., Kim, Y.J., Sharpe, T.L., Moore, D.J., Prince-Stehley, K. (2017)	USA 2,328: 1152 males, 1176 females 12–17 years African-American	Cross-sectional Secondary data analysis of the 2012 National Survey of Drug Use and Health (data collected in 2011)	Interpersonal violence including gang fighting and carrying a handgun	<ul style="list-style-type: none"> • Parental involvement • Family composition • Religious beliefs • School connectedness • Community engagement Aggregate score calculated for each factor
Brookmeyer, K.A., Henrich, C.C., Schwab-Stone, M. (2005)	USA 1,599: 49% males, 51% females (exact numbers not stated) 6 th and 8 th grade students 61% African American, 26% Hispanic, 12% White	Longitudinal (2000-2001) Self-report Community violence exposure measured as a risk factor for future violence	Interpersonal community violence	<ul style="list-style-type: none"> • Parent support • Social cognitive processes
Chui, W.H.,	Hong Kong	Cross-sectional	Interpersonal	<ul style="list-style-type: none"> • Peer attachment

Reference	Country/ Sample	Study Design/ Data Collection	Violence Outcome	Protective Factors
Chan, H.C.O. (2012)	1,377: 666 males, 711 females Secondary school students (12 – 17 years) Ethnicity not stated	Self-report	violence	<ul style="list-style-type: none"> • Parental bonding • Parental dependence • School commitment • Belief in the legal system
Dornbusch, S.M., Erickson, K.G., Laird, J., Wong, C.A. (2001)	USA 13,568: numbers of males/females not stated 7 th to 12 th grade students Ethnicity not stated	Longitudinal: reporting period 1995 – 1996 Add Health survey data from Wave 1 (1995) and Wave 2 (1996) Self-report and parent interview	Interpersonal violence including weapon use and gang fighting	<ul style="list-style-type: none"> • Family attachment (parent-family connectedness, parents' closeness to the adolescent) • School connectedness
Griffin, K.W., Botvin, G.J., Scheier, L.M., Diaz, T., Miller, N.L. (2000)	USA 228: 50% males, 50% females (numbers not stated) 6 th grade students Black 88%, Hispanic 2%, Asian 2%, White 1%, Mixed/other 7%	Cross-sectional Self-report and parent interview	Interpersonal violence including gang fighting and destruction of property	<ul style="list-style-type: none"> • Parental monitoring • Parent-child communication • Parental involvement (checking homework, family eating together)
Nash, J.K., Mujanovic, E., Winfree Jr, L.T. (2011)	Bosnia and Herzegovina 2,134: 1,037 males, 1,097 females Mean age 14.97 (SD=1.32) Ethnicity not reported	Cross-sectional Self-report Community violence exposure measured as a risk factor for future violence	Interpersonal violence including gang fighting and destruction of property	<ul style="list-style-type: none"> • Parental monitoring
Park, S., Morash, M., Stevens, T. (2010)	USA 2,552: numbers of males/females not stated 12-13 years Ethnicity not reported	Longitudinal; reporting period 1997 and 2001/02 Secondary analysis of data from the National Longitudinal Survey of Youth 1997 cohort	Interpersonal violence (frequency of assaults/attacks)	<ul style="list-style-type: none"> • Parental support • Parental monitoring • Grades completed in school • Religious activity • Work activity

Reference	Country/ Sample	Study Design/ Data Collection	Violence Outcome	Protective Factors
Pu, J., Chewning, B., St Clair, I.D., Kokotailo, P.K., Lacourt, J., Wilson, D. (2013)	USA	Self-report Cross-sectional	Interpersonal violence and victimisation	<ul style="list-style-type: none"> • Perceived parental monitoring • Self-efficacy
	630: 335 males, 285 females	Self-report		
	6 th to 12 th grade school students			
	American Indians			
Resnick, M.D., Ireland, M., Borowsky, I. (2004)	USA	Longitudinal; reporting period 1995 and 1996	Interpersonal violence including weapon use and gang fighting	<ul style="list-style-type: none"> • Community factors (school connectedness, other adult connectedness) • Family factors (family connectedness, discuss problems with parents, parental school expectations, parental presence, activities with parents) • Personal factors (religiosity, grade point average)
	6,913 males, 7,419 females	Add Health survey data from Wave 1 (1995) and Wave 2 (1996)		
	12-18 years	Self-report		
	Ethnicity not reported			

Study design and data collection

Six studies were cross-sectional and four were longitudinal in design. Three studies used data from the same source, the National Longitudinal Study of Adolescent Health (Add Health). Add Health surveyed a nationally representative sample of adolescents from the USA in grades 7-12 during the 1994-5 school year with follow-up into adulthood (Add Health, 2018). Five waves of data have been collected and include information on individuals' social, economic, psychological and physical well-being. The Add Health studies in this review used data from Wave 1 (1995) or Wave 1 and 2 (1996). These three studies had the largest sample sizes ranging from 13,568 (Dornbusch et al., 2001) to 17,036 (Blum et al., 2003). Table 2 reports information as detailed in the research papers therefore a total sample size is not reported for Resnick et al. as the numbers reported in

the text are inconsistent with the number of participants reported in their data tables.

Sample sizes of the remaining studies ranged from 228 (Griffin et al., 2000) to 2,552 (Park et al., 2010), age range 11-18 years.

Violence

All studies used self-report data to measure violence outcome. In most studies a single composite measure was calculated from responses to multiple questionnaire items about violence or aggression. Whilst self-report data can be a relatively quick and easy way to gather individual's opinions it is very susceptible to response bias with people exaggerating or under-reporting the occurrence of behaviours. Also, there are limitations in the recall accuracy of self-report data when people are asked about what they did or experienced in the past 12 months. The over-reliance on such data in these studies raises questions over the validity of the results. One study used frequency data on interpersonal violence (assaults or attacks) (Park et al., 2010). In the Add Health studies, Blum (2003) and Resnick (2004) used an eight-item scale which included seven questions on violence perpetration and one on violent victimization. Dornbusch et al. (2001) only used the violence perpetration questions. See Appendix 3 for further detail of outcome measures.

Protective factors

All studies used family factors such as parental monitoring, involvement or communication within the home. Six studies included school factors, such as grade point average or connectedness with school. Seven studies included individual factors such as self-efficacy or religious beliefs.

Two longitudinal studies additionally measured community violence exposure at baseline and examined the mediating impact of protective factors on this exposure to violence perpetration at follow-up.

Study Results

The diversity in study design, participant ethnicity, and protective factors measured precluded a statistical synthesis of the included studies therefore a meta-analysis was not performed. A narrative synthesis of the results is provided.

Study outcomes are shown in Table 3. All studies used a violence outcome measure of interpersonal violence, some studies included gang fighting and weapon use in this measure (see Table 2). Eight studies report gender differences in the association between some protective factors and violence. Blum et al. (2003) found gender differences in protective factors with ‘school connectedness’ and ‘family caring/connectedness’ being significantly related to less violence for males and females respectively. Results for remaining protective factors were not significantly related to violence. Boyas et al. (2017) found gender differences in family composition and school grades. Having a father in the household was significantly associated with less violence for males, and higher school grades were significantly associated with less violence for females. Results for remaining protective factors were not significantly related to violence.

In their longitudinal study, Brookmeyer et al. (2005) identified gender differences in the buffering effect of parent support and social cognition processes on adolescents exposed to community violence. Social cognitions were measured using hypothetical vignettes in which the adolescent was asked about the cause of an ambiguous peer provocation situation. They chose from four responses indicating different causal attributions, e.g.

hostile, prosocial, or benign¹. For males, average or high levels of parent support were significantly related to lower levels of violence perpetration. For females, possessing more prosocial cognitions (relative to the rest of the sample) was significantly related to lower levels of violence perpetration. No significant relationships were found for parent support and violence perpetration for females, or prosocial cognitions and violence perpetration for males.

Griffin et al. (2000) identified a gender difference in more frequent parent checking of homework and levels of interpersonal aggression. For females, it was associated with less aggression, but for males it was associated with increased levels of aggression. No other significant gender differences were identified.

Nash et al. (2011) found that parental monitoring on its own was not significantly related to violent offending for girls and boys. They also measured victimisation with Likert-scale responses to questions such as how often the individual had “been hit by someone trying to hurt you” in the previous 12 months. They defined ‘high victimisation’ as “at or above the 90th centile” and ‘low victimisation’ as “at or below the 90th centile”. No rationale was given for the high/low criteria used. However, multivariate analysis of victimisation, parental monitoring and violent offending showed a significant interaction of victimisation and parental monitoring indicating differential effects of monitoring on violent offending for girls and boys depending on their level of victimisation (high=, low=at or below the 90th centile). Further analysis of this relationship showed a gender-related age difference in the interaction. Specifically, male and female adolescents under 15 years old who reported experiencing high levels of victimisation experienced a strong protective effect of high levels of parental monitoring i.e. they reported perpetrating lower levels of violence. For

¹ Exact details of vignettes and responses not stated in the paper.

girls aged 15 years or older, the protective effect of high levels of parental monitoring was evident, but at a more moderate level. However, for older boys, although high levels of monitoring were associated with lower levels of offending, the magnitude of difference between the high and low victimization groups was negligible. This suggests that older boys, previously exposed to violence, do not experience the same protective effect of parental monitoring unlike older girls.

Using longitudinal data Park et al. (2010) identified gender differences across the protective factors. Parental monitoring had a significant negative association with the perpetration of assault for females but not males. For males and females, grades completed in school were negatively associated with the frequency of violence. Religious activity was not significantly associated with violence for males or females.

Pu et al. (2013) investigated adolescent violence in an American Indian population. They identified gender differences with females showing a significant negative association between the protective factors of perceived parental monitoring and self-efficacy with violence perpetration. These associations were not significant for males.

In their analysis of longitudinal data, Resnick et al. (2004) identified level of parental expectations for school performance, the ability to discuss problems with parents, and a sense of connectedness to adults outside of the family as significant protective factors for boys. For girls, family connectedness, religiosity, and school connectedness were significant. Grade point average was significantly protective for boys and girls. All other associations between protective factors and violence were not significant.

Two studies did not report any gender differences. Dornbusch et al. (2001) identified parent closeness and school connectedness as having significant associations with violence for males and females, with stronger associations being shown for school connectedness.

No significant relationship was identified for connectedness to the family, and no gender differences were identified for any of the other factors. For males and females, Chui et al. (2012) found a significant positive association between parental bonding, school commitment, belief in the legal system and violence perpetration. Peer attachment and parental dependence was not statistically significant for males or females.

Summary

From these studies, there is an indication that family-related factors may be more protective for females, with six studies identifying significant associations between lower levels of perpetrated violence and factors such as family connectedness and parental monitoring (for example, parents being aware of their child's location and who their friends are). Yet studies also identified family-related factors that were significant for males, such as parental support and the presence of a father in the home. Half the studies used a wide age range (e.g. 12-18 years) but age related differences in protective factors were only examined by Nash et al. (2011). They examined how the impact of parental monitoring may vary across younger and older adolescents who have experienced violent victimisation. The results showed a similar relationship between parental monitoring and violence amongst younger adolescents, but for older adolescents, there was a stronger protective effect for girls than boys. Interestingly, Nash et al. only found a gender difference when analysing parental monitoring as an interactive risk factor, i.e. taking into account whether participants had been exposed to violence. Analysis of parental monitoring by itself was not significantly related to violence perpetration. Both Dornbush et al. and Chui et al. who reported no gender differences in protective factors used samples ranging from 12 to 17 or 18 years old, did not distinguish between age in their analysis, and did not include risk factors. These factors may have impacted on their non-significant results.

Table 3

Comparison Analysis and Outcomes of Included Studies

Reference	Comparison Analysis	Outcome (relationship between protective factors and violence)
Blum, J., Ireland, M., Blum, R.W. (2003)	Odds ratios of protective factors and violence	Statistically significant protective factors: <ul style="list-style-type: none"> • Males <ul style="list-style-type: none"> ○ school connectedness (OR = .70, $\chi^2=5.66$, $p<.05$) • Females <ul style="list-style-type: none"> ○ family caring (OR = .61, $\chi^2=11.76$, $p<.001$) ○
Boyas, J.F., Kim, Y.J., Sharpe, T.L., Moore, D.J., Prince-Stehley, K. (2016)	Path analysis of violence and protective factors by gender Pairwise comparison of gender path analysis by gender	Statistically significant protective factors: <ul style="list-style-type: none"> • Males <ul style="list-style-type: none"> ○ father in the household ($\beta=-.08$, $p<.05$) • Females <ul style="list-style-type: none"> ○ school grades ($\beta=-.13$, $p<.001$) <p>Critical ratio comparisons between the male and female models showed a significant moderating effect of gender between ‘father in household’ (critical ratio = 2.99, $p<.01$) and ‘family members in household’ (critical ratio = -2.18, $p<.05$), despite the latter not being significant in the individual male and female pathways. Results for all other protective factors were not significant.</p>
Brookmeyer, K.A., Henrich, C.C., Schwab-Stone, M. (2005)	Three-way interaction regression analyses	Statistically significant protective factors from multiple regression: <ul style="list-style-type: none"> • Males <ul style="list-style-type: none"> ○ Parent support – three-way interaction of witnessing violence x perceived support x gender, $\beta=-.12$, $p<.01$. Further analysis revealed that average and high levels of perceived support buffered the effects of violence exposure. • Females <ul style="list-style-type: none"> ○ Social cognitions – three-way interaction of witnessing violence x social cognitions x gender, $\beta=-.13$, $p<.01$. Adolescent females who witnessed violence appeared to be protected from committing acts of violence if they had prosocial cognitions relative to the rest of the sample. ○
Chui, W.H., Chan, H.C.O. (2012)	Ordinary least square (OLS) regression modelling	No gender difference in statistically significant protective factors (Parental bonding, School commitment, Belief in the legal system)
Dornbusch, S.M., Erickson, K.G., Laird, J., Wong, C.A. (2001)	Linear and logistic regression	No gender difference in statistically significant protective factors (parent closeness and school connectedness)
Griffin, K.W., Botvin, G.J., Scheier, L.M., Diaz, T., Miller, N.L. (2000)	Hierarchical multiple regression analyses	More frequent parent checking of homework was associated with less aggression in girls ($\beta =-.21$, $p<.10$) and more aggression in boys ($\beta=.15$, $p<.10$).
Nash, J.K., Mujanovic, E., Winfree	Multiple linear regression and general linear models	Parental monitoring on its own not significantly related to violent offending for girls or boys. Analysis of interaction between monitoring and victimization:

Reference	Comparison Analysis	Outcome (relationship between protective factors and violence)
Jr, L.T. (2011)		<ul style="list-style-type: none"> • girls and boys ≤ 15 years old: monitoring had a stronger protective effect for those reporting high, relative to low, levels of victimization • girls > 15 years old: similar but more modest association between monitoring and victimization • boys > 15 years old: effect of parental monitoring similar across low and high victimization groups
Park, S., Morash, M., Stevens, T. (2010)	Negative binomial regression	<ul style="list-style-type: none"> • Parental support <ul style="list-style-type: none"> ○ Males: significant positive association ($b=.59, p \leq .05$) • Parental monitoring <ul style="list-style-type: none"> ○ Females: significant negative association ($b=-.06, p \leq .01$) • Work activity <ul style="list-style-type: none"> ○ Females: significant positive association ($b=.06, p \leq .05$)
Pu, J., Chewning, B., St Clair, I.D., Kokotailo, P.K., Lacourt, J., Wilson, D. (2013)	Path analysis	<ul style="list-style-type: none"> • Perceived parental monitoring <ul style="list-style-type: none"> ○ Females: significant negative association (estimate $= -.06, p < .001$) • Self-efficacy <ul style="list-style-type: none"> ○ Females: significant negative association (estimate $= -.09, p < .001$)
Resnick, M.D., Ireland, M., Borowsky, I. (2004)	Multiple linear regression	<p>Community factors</p> <ul style="list-style-type: none"> • School connectedness <ul style="list-style-type: none"> ○ Males: significant positive association (estimate $= .0121, p = .043$) • Other adult connectedness <ul style="list-style-type: none"> ○ Females: significant negative association (estimate $= -.0255, p = .012$) <p>Family factors</p> <ul style="list-style-type: none"> • Family connectedness <ul style="list-style-type: none"> ○ Males: significant positive association (estimate $= .0161, p = .007$) • Discusses problems with parents <ul style="list-style-type: none"> ○ Females: significant negative association (estimate $= -.0175, p = .004$) • Parental school expectations <ul style="list-style-type: none"> ○ Females: significant negative association (estimate $= -.0326, p < .001$) <p>Personal factors</p> <ul style="list-style-type: none"> • Religiosity <ul style="list-style-type: none"> ○ Males: significant negative association (estimate $= -.0155, p = .008$)

Quality Assessment

Results from assessment of studies using the CCAT are presented in Table 4. Individual criterion scores, total score and total percentage are shown. A higher percentage may be

considered indicative of a higher quality study, however, consideration of individual criteria scores is also important to interpretation (Crowe & Shepherd, 2013).

The CCAT indicates variable quality across the studies with percentage totals ranging from 55% (Blum et al., 2003; Park et al., 2010) to 83% (Brookmeyer et al., 2005). Half the studies achieved a percentage total of less than 70%. In the highest quality study, Brookmeyer et al. clearly explained the purpose of the study and clarified the design and data collection processes. Their use of a longitudinal design allowed for consideration of the temporal ordering of risk and protective factors. In contrast, Blum et al., (2003) and Park et al., (2010) lacked clarity in their explanation of study design, sampling and data collection processes. Potential sources of bias were also often unclear. Across all studies the reporting of ethical matters was poor, for example few studies reported the use of informed consent or the confidentiality of data.

Table 4
CCAT quality assessment scoring

Reference	Prelim.	Intro.	Design	Sample	Data	Ethics	Results	Discuss.	Total	Total %
Blum et al. (2003)	4	4	2	2	3	1	3	3	22	55
Boyas et al. (2017)	3	3	3	4	4	2	4	4	27	68
Brookmeyer et al. (2005)	5	5	4	4	5	1	4	5	33	83
Chui et al. (2012)	3	4	3	3	4	3	4	4	28	70
Dornbusch et al. (2001)	4	5	4	2	3	2	4	4	28	70
Griffin et al. (2000)	3	5	4	3	4	3	2	4	28	70
Nash et al. (2011)	4	4	3	3	3	1	3	4	25	63
Park et al. (2010)	4	2	3	3	2	1	4	3	22	55
Pu et al. (2013)	5	4	4	3	3	2	4	4	29	73
Resnick et al. (2004)	4	4	3	3	3	3	3	3	26	65

Both the studies identifying no gender differences, Dornbusch et al. (2001) and Chui et al. (2012) achieved higher percentage totals, both 70%. The former study had a very large sample of 13,568 and referred to the use of a “nationally representative sample”, however the number of boys and girls is not reported. This is clearly relevant to the interpretation of their non-significant gender difference findings and limits the conclusions that can be drawn from their study.

Discussion

This review examined studies that investigated gender differences in protective factors for adolescent violence perpetration. It aimed to identify the type of violence outcomes and protective factors measured, whether protective factors differ by gender, and whether protective factors moderate the impact of risk factors for violence over time. In terms of violence outcome, all the studies measured interpersonal violence as an outcome and made no distinction between the nature of violence, or victim type (such as stranger or known, peer, or family member). The one study that did distinguish between violence prevalence, intensity and frequency found these variables to have differing relationships with protective factors (Dornbusch et al., 2001). Specifically, that prevalence of violence was associated with school connectedness and this factor could operate to “deter adolescents from engaging at all in a specific form of deviation, but, once deviance has occurred, the relative strength of school connectedness as an influence on deviant behaviours tended to be weaker”. Dornbusch et al. caution against the sole use of frequency data, such as used by Griffin et al. (2000). Whilst this was only explored in one study, the CCAT results indicate it to be of reasonable quality and it highlights a pertinent issue that suggests a general limitation of the literature. Given that gender differences have been identified in the nature of violence perpetrated (e.g. Franke et al., 2002) it seems relevant to distinguish between different types of violence when examining the relationship between protective factors and violence outcomes.

The type of protective factors examined in the literature were from individual, family and social-environmental domains. The type of protective factors identified as having gender differences across studies were mainly family or social-environmental factors, such as family connectedness, parental monitoring or school connectedness. The gender differences in these factors were not consistent across all studies, but there was a tendency

towards family factors being more protective for girls than boys. This supports the suggestion that relational factors may be more relevant for girls (Losel & Farrington, 2012). However, given the limited number of studies and poor quality of some, such as failing to examine non-responders and address possible bias in the sample, ~~the~~ this is a tentative finding.

Interestingly, only two studies examined individual factors. Pu et al. (2013) and Brookmeyer et al. (2005) identified social cognition and self-efficacy as differing between boys and girls, with high levels of both factors being identified as relating to lower levels of aggression for girls only. Brookmeyer et al. also examined this relationship in the context of adolescents exposed to community violence. These were the highest quality rated studies suggesting these findings can be considered with greater confidence. Pu et al.'s study was of an American Indian sample and therefore cultural influences may be relevant to the interpretation of the results. The results of these two studies may indicate the presence of a cross-cultural developmental process, such as neurodevelopmental-based gender differences in the development of social cognition processes. It should be noted that Pu et al had a sample of students from 6th to 12th grade but no distinction in age was made in their analysis.

The identification of a consistent relationship between an individual factor and violence is overly simplistic given that it is likely to be influenced by associations with multiple other risk and protective factors (Baxendale et al., 2012). Some studies did measure additional risk factors as part of the cross-sectional design, and two studies examined the potential for protective factors to modify the impact of risk factors. Nash et al. (2011) and Brookmeyer et al. (2005) identified 'at risk' youth who had been exposed to community violence, the latter study doing so within a longitudinal design. Nash et al. further distinguished between younger and older adolescents and identified a stronger protective effect of parental

monitoring for males and females age 15 years or younger. This age impact is interesting to consider given other studies failed to find an association with monitoring and violence, such as Griffin et al. (2000).

A strength of the studies in this review include the fact that the sample populations were taken from the general population, and eight of the studies had large sample sizes of over 1000. However, there are several limitations identified within the included studies. Most studies used a cross-sectional design. Measuring the co-occurrence of violence and protective factors does not allow for inferences to be drawn about a causal relationship between the protective factor and the violence outcome. Despite this, there was evidence of studies referring to an intention to use protective factors to “predict” violence (Boyas et al., 2017). The use of school-based surveys do not generalise to out-of-school youth, and the data from the large surveys such as the Add Health studies were several years old. Finally, and importantly, there was a clear over-reliance on self-report data of violence in all the studies, which may result in the under or over-reporting of behaviours. Whilst some studies acknowledged this, such as Boyas et al. (2017), not all did.

A strength of this review is the inclusion of a quality assessment to evaluate the methodological quality and risk of bias in the research. In terms of limitations, the focus on violence specifically means that some studies that included violence as part of more general delinquency measure were excluded if the violence measure was not explicitly operationalised. Also, the publications searched did not include grey literature.

In conclusion, the findings from the studies reviewed show that there are gender differences in some protective factors for violence. It appears that family factors are more protective for females, however this is a tentative conclusion based on the variable quality of the studies and methodological limitations discussed above. There is an indication that

the impact of protective factors may vary with age. Although this was only explored in one study it highlights an important area that should be explored in future studies as the use of a sample with a wide age-range (as was frequently the case in these studies) may be insensitive to the variation in how protective factors operate across the developing adolescent. There is a need for more high quality longitudinal studies to explore the moderating effect of gender on protective factors, particularly individual factors which were only addressed in two studies. There is also a need for greater specificity of how violence outcomes are defined and measured in order to aid understanding of their association with protective factors. In addition, it is important to explore how these differences may vary across different communities with different cultural influences. This knowledge would help to inform policy makes and practitioners in their support of young people engaging in violence.

References

- Add Health. (2018). Retrieved from <http://www.cpc.unc.edu/projects/addhealth>
- Bandura, A. (1986). *Social foundations of thought and action: A social-cognitive theory*. Upper Saddle River, NJ: Prentice Hall.
- Baxendale, S., Cross, D., & Johnston, R. (2012). A review of the evidence on the relationship between gender and adolescents' involvement in violent behavior. *Aggression and Violent Behavior, 17*(4), 297-310.
- Blum, J., Ireland, M., & Blum, R. W. (2003). Gender differences in juvenile violence: a report from Add Health. *Journal of Adolescent Health, 32*(3), 234-240.
- Boyas, J. F., Kim, Y. J., Sharpe, T. L., Moore, D. J., & Prince-Stehley, K. (2017). An ecological path model of use of violence among African American adolescents. *Child & Youth Services, 38*(1), 24-52.
- Brookmeyer, K. A., Henrich, C. C., & Schwab-Stone, M. (2005). Adolescents who witness community violence: can parent support and prosocial cognitions protect them from committing violence? *Child Development, 76*(4), 917-929.
- Calvert, W. J. (1997). Protective factors within the family, and their role in fostering resiliency in African American adolescents. *Journal of cultural diversity, 4*(4), 110.
- Centre for Reviews and Dissemination. (2009). *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. Retrieved from https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf on 29th June 2018.
- Chui, W. H., & Chan, H. C. O. (2012). An empirical investigation of social bonds and juvenile delinquency in Hong Kong. *Child & Youth Care Forum, 41*(4), 371-386.

- Crowe, M. (2013). *Crowe Critical Appraisal Tool (CCAT) User Guide: Version 1.4*.
- Crowe, M., & Sheppard, L. (2011). A general critical appraisal tool: An evaluation of construct validity. *International Journal of Nursing Studies, 48*(12), 1505-1516.
- Crowe, M., Sheppard, L., & Campbell, A. (2012). Reliability analysis for a proposed critical appraisal tool demonstrated value for diverse research designs. *Journal of Clinical Epidemiology, 65*(4), 375-383.
- Dornbusch, S. M., Erickson, K. G., Laird, J., & Wong, C. A. (2001). The relation of family and school attachment to adolescent deviance in diverse groups and communities. *Journal of Adolescent Research, 16*(4), 396-422.
- Farrington, D. P., Ttofi, M. M., & Piquero, A. R. (2016). Risk, promotive, and protective factors in youth offending: Results from the Cambridge study in delinquent development. *Journal of Criminal Justice, 45*, 63-70.
- Franke, T. M., Huynh-Hohnbaum, A.-L. T., & Chung, Y. (2002). Adolescent Violence: With Whom They Fight and Where. *Journal of Ethnic And Cultural Diversity in Social Work, 11*(3-4), 133-158.
- Gorman-Smith, D., & Tolan, P. (1998). The role of exposure to community violence and developmental problems among inner-city youth. *Development and Psychopathology, 10*(1), 101-116.
- Griffin, K. W., Botvin, G. J., Scheier, L. M., Diaz, T., & Miller, N. L. (2000). Parenting practices as predictors of substance use, delinquency, and aggression among urban minority youth: moderating effects of family structure and gender. *Psychology of addictive behaviors, 14*(2), 174-184.

- Hartman, J. L., Turner, M. G., Daigle, L. E., Exum, M. L., & Cullen, F. T. (2009). Exploring the Gender Differences in Protective Factors Implications for Understanding Resiliency. *International Journal of Offender Therapy and Comparative Criminology*, 53(3), 249-277.
- Herrenkohl, T. I., Kosterman, R., Mason, W. A., & Hawkins, J. D. (2007). Youth violence trajectories and proximal characteristics of intimate partner violence. *Violence and victims*, 22(3), 259-274.
- Jolliffe, D., & Farrington, D. P. (2007). *A systematic review of the national and international evidence on the effectiveness of interventions with violence offenders*. Ministry of Justice.
- Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *The Lancet*, 360(9339), 1083-1088.
- Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & White, H. R. (2008). *Violence and serious theft: Development and prediction from childhood to adulthood*. New York, NY, US: Routledge/Taylor & Francis Group.
- Lösel, F., & Farrington, D. P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American journal of preventive medicine*, 43(2), S8-S23.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The Prisma Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS MEDICINE*, 6(7), e1000097.

- Nash, J. K., Mujanovic, E., & Winfree Jr, L. T. (2011). Protective effects of parental monitoring an offending in victimized youth in Bosnia and Herzegovina. *Child & Youth Services, 32*(3), 224-242.
- Park, S., Morash, M., & Stevens, T. (2010). Gender Differences in Predictors of Assaultive Behavior in Late Adolescence. *Youth Violence and Juvenile Justice, 8*(4), 314-331.
- Pollard, J. A., Hawkins, J. D., & Arthur, M. W. (1999). Risk and protection: Are both necessary to understand diverse behavioral outcomes in adolescence? *Social Work Research, 23*(3), 145-158.
- Powell, K. M. (2015). *A Strengths-Based Approach for Intervention with At-Risk Youth*. Champaign IL: Research Press.
- Pu, J., Chewning, B., St. Clair, I. D., Kokotailo, P. K., Lacourt, J., & Wilson, D. (2013). Protective factors in American Indian communities and adolescent violence. *Maternal and Child Health Journal, 17*(7), 1199-1207.
- Resnick, M. D., Ireland, M., & Borowsky, I. (2004). Youth violence perpetration: What protects? What predicts? Findings from the National Longitudinal Study of Adolescent Health. *Journal of Adolescent Health, 35*(5), 424.e1-424.e10.
- Salzinger, S., Feldman, R. S., Stockhammer, T., & Hood, J. (2002). An ecological framework for understanding risk for exposure to community violence and the effects of exposure on children and adolescents. *Aggression and Violent Behavior, 7*(5), 423-451.
- Saner, H., & Ellickson, P. (1996). Concurrent risk factors for adolescent violence. *Journal of Adolescent Health, 19*(2), 94-103.

Werner, E.E., & Smith, R.S. (2001). *Journeys from childhood to midlife: risk, resilience, and recovery*. Ithaca NY: Cornell University Press.

WHO. (2016). World Health Organisation Youth Violence Factsheet. Retrieved from <http://www.who.int/en/news-room/fact-sheets/detail/youth-violence> on 11th January 2018.

WHO. (2018). Youth Violence. Retrieved from <http://apps.who.int/violence-info/youth-violence> on 11th January 2018.

Chapter 2: Major Research Project

Inter-Rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) Amongst Mental Health Professionals

Submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology

October 2018

Address for Correspondence:

Sarah Selby

Institute of Health & Wellbeing

College of Medical, Veterinary & Life Sciences

University of Glasgow

Administration Building

Gartnavel Royal Hospital

1055 Great Western Road

Glasgow G12 0XH

Email: s.selby.1@research.gla.ac.uk

Chapter word count (including references): 6847

Prepared in accordance with the requirements for submission to the *International Journal of Forensic Mental Health* (see Appendix 1).

Plain English Summary

Inter-Rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) Amongst Mental Health Professionals

Violence is a significant problem worldwide due to the consequences of being a victim, offender, or witness to violence. The World Health Organisation (WHO) describes it as a “global health problem”. Homicide is the fourth leading cause of death amongst young people aged 10-29 years old with an estimated 200,000 homicides globally each year (WHO, 2016). For every death, many more young people suffer non-fatal injuries due to violence that lead to lifelong physical, psychological and social consequences.

To limit these consequences, clinicians need to understand why young people are violent so that they can provide effective interventions. The Structured Assessment of Violence Risk in Youth (SAVRY) is a clinician guide for assessing violence risk in 13-18 year olds (Borum et al., 2010). It contains 24 factors that are associated with violence risk, as well as 6 protective factors that may reduce violence risk if present. Clinicians rate the presence of each factor for the young person they are assessing. To ensure the SAVRY is useful in assessing violence it is important to know how much clinicians agree in their risk ratings when rating the same person. This is known as inter-rater reliability (IRR).

Previous IRR studies of the SAVRY have used just two or three raters who were specially trained for the purpose. In clinical practice, clinicians with varying levels of experience complete risk assessments and it is unknown how much their experience level may impact on the reliability of their ratings.

This study assessed the IRR of the SAVRY by asking a group of professionals who work with adolescents (psychologists, nurses, social workers) to rate case studies of boys and girls. Their ratings were also compared with expert (professionals with experience in adolescent risk assessment). A larger group of raters than previous research was used and possible sources of bias were examined that may impact on risk assessment decision-making, such as rater experience and gender bias.

Participants achieved a good level of IRR amongst themselves, but a lower level with experts. The gender of vignettes and violence severity did not conclusively affect ratings. Participants who were more confident in their ratings thought they were more objective. Confidence did not relate to years of experience or increased reliability of ratings. Professionals completing risk assessments would benefit from ongoing training regardless of their level of experience.

References

Borum, Lodewijks, Bartel, Forth (2010) Structured Assessment of Violence Risk in Youth (SAVRY). In R.K. Otto & K.S. Douglas (Eds) Handbook of Violence Risk Assessment (pp63-79). New York: NY, US: Routledge/Taylor & Francis Group.

WHO (2016). World Health Organisation Youth Violence Factsheet.

<http://www.who.int/mediacentre/factsheets/fs356/en/> Date accessed 23.01.17.

Abstract

Inter-rater reliability (IRR) of the Structured Assessment of Violence Risk in Youth (SAVRY) amongst mental health professionals was explored. Participants ($n=22$) rated vignettes varying in violence severity and gender (case characteristics). Professional characteristics including perception of their confidence and objectivity in ratings were measured. Using ICC1 scores, the IRR was low for items when interpreted using the 95% confidence interval but much higher using individual ICC scores. Case characteristics did not conclusively affect ratings. Self-reported confidence in ratings was not associated with increased reliability. This suggests that professionals completing risk assessments are likely to benefit from ongoing training.

Keywords: SAVRY, adolescence, violence, risk assessment, inter-rater reliability

Introduction

Youth Violence

Violence can be defined as “the intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community, that either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment or deprivation” (Krug et al., 2002). This broad definition seeks to convey the complexity of violent behaviour, and youth violence is similarly described in broad terms “harmful behaviours that can start early and continue into young adulthood. The young person can be a victim, an offender, or a witness to the violence” (Centers for Disease Control and Prevention, 2016). Violence is a significant problem across the world and is described as a “global public health problem” (WHO, 2016). Homicide is the fourth leading cause of death amongst young people aged 10-29 years old with an estimated 200,000 occurring worldwide each year (WHO, 2016). For every death, more young people suffer non-fatal injuries that lead to lifelong physical, psychological and social consequences. Increasing understanding of why violence occurs will contribute to the development of violence reduction interventions for youth.

Violence Risk Assessment

Violence risk assessment has evolved from unstructured clinical judgements and actuarial methods to the Structured Professional Judgement (SPJ) approach. SPJ draws on the strengths of both approaches and uses empirically supported risk factors to assess individuals. SPJ and actuarial measures have been shown to have similar moderate predictive validity (Guy, Douglas, & Hart, 2015). However, the predictive validity of the

summary risk rating² (SRR) has been shown to be greater compared to the calculation of a numeric total, as happens using actuarial approaches. Contemporary use of the SPJ approach is more concerned with understanding the presence and relevance of risk factors for the individual to aid formulation, scenario planning and risk management, rather than solely predicting violence. The SPJ approach is widely accepted within mental health and criminal justice settings as a better approach to managing risk (Hart, Douglas, & Guy, 2016).

Structured Assessment of Violence Risk in Youth

The Structured Assessment of Violence Risk in Youth (SAVRY) (Borum, Bartel, & Forth, 2006) is an SPJ tool used internationally to assess violence risk in young people aged between 13 and 18 years old. It evaluates a set of empirically derived risk and protective factors for violence to inform risk formulation, scenario planning, and risk management.

The SAVRY defines violence as “an act of physical battery sufficiently severe to cause injury that would require medical attention, a threat with a weapon in hand, or any act of forcible sexual risk”. Borum and Verhaagen (2006) define risk as arising from the dynamic, reciprocal interaction between factors that influence the likelihood of offending. SAVRY items are listed in Appendix 4. There are 24 risk factors in three domains (Historical, Social/Contextual, and Individual/Clinical) rated using a descriptive code (Low, Moderate, High). Six protective factors are also included; these are rated as Present or Absent. Additional risk and protective factors are included at the clinician’s discretion. An overall risk rating, or SRR, is given of Low, Moderate or High. This requires clinicians

² An overall rating of risk based on consideration of all factors and any idiosyncratic case characteristics, rather than calculating a numeric total from individual item scores

to use their professional judgement to determine the nature and degree of risk and is not simply a sum of the risk and protective factors.

The authors are clear that the SAVRY should not be used to quantify risk in an absolute sense, but rather to structure risk assessment and formulation based on the risk and protective factors. They assert that identifying a factor as present for an individual (presence) is less important than how that factor is associated with violence (relevance). For example, someone has a history of self-harm but this may not be not relevant to their future violence risk.

The SAVRY is applicable to males and females. The authors report that many risk and protective factors operate similarly for both genders, although sensitivity and rates of exposure for each may differ (Borum, Lodewijks, Bartel, & Forth, 2010). See Chapter 1 for a review of the association between gender and protective factors.

Reliability and Validity of the SAVRY

The SAVRY has high levels of predictive validity, for example, the SRR is predictive of violent and general offending over a four-year period (Gammelgård et al., 2015). However, arguably that predictive validity may not be the most appropriate assessment of SPJ tool utility due to the limitations arising from applying probability judgements about a group to the individual (Hart, Michie, & Cooke, 2007). Hart et al. (2007) explain, “probability is defined in frequentist terms as the proportion of people who will commit violence...the margin of error is uncertainty regarding the proportion of people who commit violence.” Conversely, at the individual level the margin of error is uncertainty about whether or not a person will commit violence, therefore the application of group level predictions to an individual is imprecise. Hart et al. (2007) also highlight that the error margins are unacceptably large to predict the future with any certainty. It is argued that risk assessment

should focus on risk management, identifying what may go wrong for the individual and how it can be prevented (Sutherland et al., 2012). This approach emphasises the role of professional judgement in assessing risk.

Given this central role of professional judgement in the SPJ approach it is advised that inter-rater reliability (IRR) should be evaluated (Sutherland et al., 2012). Six studies have examined SAVRY IRR. Results show good to excellent agreement between raters with ICCs ranging from 0.81 to 0.97 for the SAVRY Risk Total, and 0.72 and 0.95 for the SAVRY SRR (Catchpole & Gretton, 2003; Dolan & Rennie, 2008; Lodewijks, Doreleijers, de Ruiter, & Borum 2008; McEachran, 2001; Meyers & Schmidt, 2008; Viljoen et al., 2008). However, these studies are limited in their applicability to clinical practice. Firstly, comparisons were made between two or three raters who are academics trained in the SAVRY specifically for the study, not clinicians. Research suggests the professional background of raters may be an important moderator of validity and reliability (Sutherland et al., 2012). Secondly, they mostly assessed total and domain scores not individual item scores (Borum et al., 2010). Given the purpose of the tool in risk management, understanding the presence/relevance of specific risk factors is essential to developing an idiosyncratic risk management plan. Thirdly, these studies have not addressed the impact of gender in violence risk assessment. Evidence suggests that mental health professionals underestimate future violence by females, and this is not due to gender-related differences in violence (Skeem et al., 2005). Fourthly, these studies did not examine the impact of case specific factors on IRR. The level of complexity and risk can affect the degree of IRR with cases at the high or low extremes achieving higher rates of IRR than those in between (Sutherland et al., 2012).

Aims/Hypotheses

This research aims to address the following questions:

1. What level of IRR is achieved by mental health professionals using the SAVRY to assess violence risk?
2. What is the level of agreement between ratings made by non-expert mental health professionals and experts (professionals with expertise in the use of the SAVRY)?
3. What is the association between the IRR and case characteristics (gender, severity of violence risk)?
4. What is the association between the IRR and rater characteristics (professional background, years of experience)?

Hypotheses:

1. Raters with less experience of risk assessment will return lower IRR scores compared to scores from more experienced raters in previous IRR studies.
2. IRR will be higher for cases with low or high levels of overall violence risk compared to cases with a moderate level of violence risk (risk level defined by the SRR).
3. Based on the observation that female violence risk is typically underestimated, the IRR will be rated lower for female than male case studies.
4. Raters with more SAVRY experience/training will demonstrate greater concordance with expert ratings and have higher rates of IRR.

Method

Ethical Approval

Ethical approval was granted by the Medical, Veterinary and Life Sciences College Ethics Committee for the University of Glasgow (Appendix 5). Research and Development approval was obtained from NHS Greater Glasgow and Clyde (NHS GG&C) (Appendix 6).

Justification of Sample Size

Previous IRR studies of SPJ tools using similar methodologies (Sutherland et al., 2012; Dickson, 2014) used samples of 28 and 19 respectively. Sample size was calculated using Walter, Eliasziw, and Donner's (1998) formula. Based on power being set at 0.8, a null hypothesis of ICC 0.3 ('fair agreement'), an alternative hypothesis of ICC 0.7 ('substantial' agreement), and a significance level of 0.05 a minimum of six vignettes and 22 raters are required. Descriptive criteria of 'fair' and 'substantial' agreement were based on ICC interpretation guidelines by Landis and Koch (1977).

Vignettes

Six fictitious vignettes were developed by the author and Dr McDonald based on clinical experience and theoretical knowledge (example provided in Appendix 7). These were replicated and the gender changed to create 12 vignettes (six pairs). Where elements of real cases were included, personally identifiable information was altered to ensure anonymity. Vignettes were cases with Low, Moderate and High SRRs. Participants were randomly allocated six vignettes (three males, three females). Vignette ordering was randomised for each participant to minimise bias from the vignettes being rated in the same order.

Expert Review Process

Six qualified professionals (three clinical psychologists, two forensic psychologists, one social worker) with extensive experience and training in using the SAVRY were emailed to request their participation as expert raters. An information sheet and consent form were attached (Appendices 8 and 9). All consented, however one was unable to provide ratings in time for the training event. Experts were sent four vignettes so that each vignette would be rated by two experts. Vignettes were randomly allocated. The randomisation process ensured that each expert received two cases of each gender. Due to the non-participation of the sixth expert four of the vignettes were rated by one expert only. In addition to SAVRY ratings, experts provided feedback on the authenticity of the vignette (see Appendix 10).

All experts agreed that the vignettes appeared authentic. Of the eight vignettes rated by two experts the average agreement between experts on all ratings (individual items, SRR) was 61% (48-77%). Low-risk vignettes showed greater variation in ratings. All experts agreed with the SRR for the high-risk cases. Overall, experts rated one vignette pair as Low, two pairs as Moderate, and three pairs as High risk based on the SRR. Feedback was requested if experts experienced difficulties rating items. Discrepancies in ratings were mainly attributed to insufficient information. In these cases, additional clarifying information was added to the vignette before the training event, for example explicitly stating the absence of a self-harm history. Following this feedback rating disagreements were resolved through a final discussion between the author and Dr McDonald. Dr McDonald has extensive experience of forensic risk assessment and adolescents having worked in forensic services for eight years and being the Consultant Clinical Psychologist at NHS GG&C Forensic Child and Adolescent Mental Health Service.

Recruitment

A recruitment email was sent to all staff in NHS GG&C Specialist Children's Services. These services were selected due to the relevance of the SAVRY to their work. Staff signed up for the study by email.

Inclusion Criteria

Participants had to meet the user criteria specified in the SAVRY manual, "professionals...having expertise (i.e. knowledge, training and experience) in child/adolescent development, youth violence and delinquency, and conducting individual assessments" (Borum et al., 2006).

Training Event

Participants attended a one-day training event on advanced risk assessment using the SAVRY. This involved a half-day didactic teaching followed by rating vignettes, the latter being a planned part of the training. Participants were aware that study participation was voluntary and they could withdraw their consent at any time. They were advised that their data would be anonymised and confidential (see Appendices 11 and 12).

Participants

Twenty-seven mental health professionals from NHS GG&C volunteered for the training. Four failed to attend on the day resulting in a total of twenty-three (21 female, 2 male) participants.

Participants included qualified (n=18) and unqualified (n=5) staff. Unqualified staff included trainee clinical psychologists (n=3), assistant psychologist (n=1) and clinical support worker (n=1). Based on the SAVRY user requirements the support worker data

was excluded from the analysis due to their lack of adequate knowledge and experience. Trainee and assistant psychologists were included in analyses as they were assessed as having adequate knowledge and experience. Trainees were in their final year of clinical training and the assistant was assessed by their supervisor to have the relevant expertise.

Participant data was gathered using a questionnaire developed for this study (see Appendix 13). Participants' professions are shown in Table 1. The 'Other' category included an Allied Health Professional and Speech and Language Therapist. Participants reported being qualified for around 12 years on average (mean=12.14, S.D.=10.93). Seven participants had worked in youth justice for around one year on average (mean=1.15, S.D.=2.97).

Statistical Analysis

Analyses were conducted using SPSS (Statistical Package for the Social Sciences) Version 24 for Mac and Microsoft Excel 2017. Numerical scores were generated for SAVRY descriptive ratings for research purposes. 'Low', 'Moderate' and 'High' were transformed to numerical values of 1, 2, and 3 respectively. 'Present' or 'Absent' for protective factors were transformed to 1 or 0. For hypotheses 1, 2 and 3, Case 1 ICCs (ICC1) and percentage agreement statistics were calculated. For hypothesis 4, percentage agreement statistics were calculated.

Missing Data

Each participant (n=22) rated six vignettes and each vignette had 31 ratings (24 risk factors, 6 protective factors, and 1 SRR), a total of 4092 ratings. 48 ratings were missing (16 risk factors, 6 protective factors, 4 SRRs), 1.17% of possible ratings. Generally, ≤ 2

ratings were missing for an item except for H3, H6, H7, I21 where three or four items were missing.

Table 1
Summary data of participants

Professional Characteristics	n	% of sample
PROFESSION		
Social Work	2	9
Clinical Psychology	9 ³	41
Psychiatry	3	14
Nursing	6	27
Other	2	9
CURRENT SETTING(S)⁴		
Area/Community Team	17	77
Residential	1	5
Secure	3	14
Inpatient	3	14
Other	5	23
PREVIOUS SETTINGS⁵		
Area/Community Team	13	59
Residential	3	14
Secure	4	18
Inpatient	12	55
Other	5	23
NUMBER of YEARS QUALIFIED		
0	4	18
1-5	4	18
6-10	2	9
11-15	4	18
16-20	2	9
21-25	1	5
26+	4	18
Missing data	1	5

When computing ICCs, a missing rating would result in the exclusion of the entire vignette from which the rating was missing. Alternative methods were considered to manage missing data. This included excluding the rater with incomplete data on an item by item

³ 1 participant was dual qualified as a clinical and forensic psychologist

⁴ 7 participants worked concurrently in two settings; % of the 22 participants in each setting was calculated so total % is >100%

basis or replacing missing data with the mean score for that item across the whole sample. The removal of a rater has a lower impact on the power of the ICC calculation than the removal of the data series for an entire vignette, therefore this method was used.

Intraclass Correlations

The ICC is an index of the correlation and agreement between measurements by two or more raters. The ICC is the recommended statistic when there are multiple raters and the data is ordinal (Uebersax, 2015). There are different forms of ICC and the one chosen will depend on the nature of the data, such as whether all cases are assessed by the same or different raters (Koo & Li, 2016). The one-way random effects model (ICC1) was chosen as the most appropriate as each rater assessed a different subset of vignettes.

Benchmarks were identified to provide a qualitative evaluation of ICC agreement consistent with previous IRR studies that reported individual item ICCs. Cicchetti and Sparrow (1981) define ICC reliability criteria as: < 0.4 = “poor”, $0.4 - 0.59$ = “fair”, $0.60 - 0.74$ = “good”, $0.75 - 1.00$ = “excellent”. Koo and Li (2016) argue that the 95% confidence interval should be used for a more robust interpretation of ICC scores. They also advise different criteria: < 0.5 = “poor”, $0.5 - 0.75$ = “moderate”, $0.75 - 0.9$ = “good” reliability, > 0.90 = “excellent”.

Percentage Agreement

Percentage agreement statistics were calculated to identify the proportion of ratings in agreement with the mean, modal and expert ratings (see Appendix 14 for method). Such statistics do not control for chance agreement between ratings and for this reason should not be used as the sole evaluation of IRR. However, Uebersax (2015) asserts they have

⁵ 11 participants worked in >1 previous setting; % calculated as above so total is >100%

utility due to their “unique common-sense value” as they allow values to be considered relative to one another in terms of individual items, vignettes and raters. Furthermore, the use of mean and modal rating could be considered indicative of different assessment methods used in clinical practice. The mean could represent the rating achieved through discussion within a team in a team formulation situation. The mode could represent the most common rating assigned by team members.

Results

Research Question 1: What level of IRR is achieved by mental health professionals using the SAVRY to assess violence risk?

Percentage agreement amongst participants is shown in Figure 1. Percentage agreement across items with the mean and mode was 77% and 69% respectively. Calculation of percentage agreement statistics required the mean and modal ratings to be rounded to the nearest integer as item ratings are whole numbers (1, 2, or 3). This process results in a potential loss of variance in the data therefore percentage agreement should be interpreted with caution. This may be a factor in the discrepancies evident between the expert and mean and modal agreement (see Appendix 15).

ICC1 scores are shown in Figure 2. Individual risk items achieved reliability levels ranging from ‘poor’ (ICC1=0.25) to ‘excellent’ (ICC1=0.90), using Cicchetti and Sparrow’s (1981) criteria. ICC1 were not calculated for protective factors as these have binary ratings (Present or Absent) and ICC1 requires ordinal data. Sixteen risk factor items achieved ‘good’ or ‘excellent’ levels of agreement. However, the 95% confident intervals for many of these items was wide, for example H10 ICC1=0.67 (‘good’ agreement), 95% confidence interval 0.48-0.86 (‘fair’ to ‘excellent’). Five items achieved a ‘poor’ level of agreement: H1, H4, SC13, I18, I22. Using Koo and Li’s criteria, only three items score in the ‘good’ to

‘excellent’ range, H5, SC16 and SC21. The mean and mode percentage agreement scores on these items showed higher agreement in contrast to the low ICC1 scores, although the percentage agreement scores should be interpreted with caution as previously stated.

Qualitative feedback from participants suggested some uncertainty about whether to include the current offence within H1. Others reported difficulty in distinguishing between anger management problems (I20) and risk taking/impulsivity (I18) and at times felt that there was overlapping evidence for these items.

Scores for the three risk domains were calculated by summing the numerical scores for individual items within each domain. A Total Risk score was calculated by summing the numerical scores for all risk items. The three domains achieved ‘good’ or ‘excellent’ levels of agreement, the Total Risk Score achieved an ‘excellent’ level of agreement based on Cicchetti and Sparrow’s criteria (see Table 2).

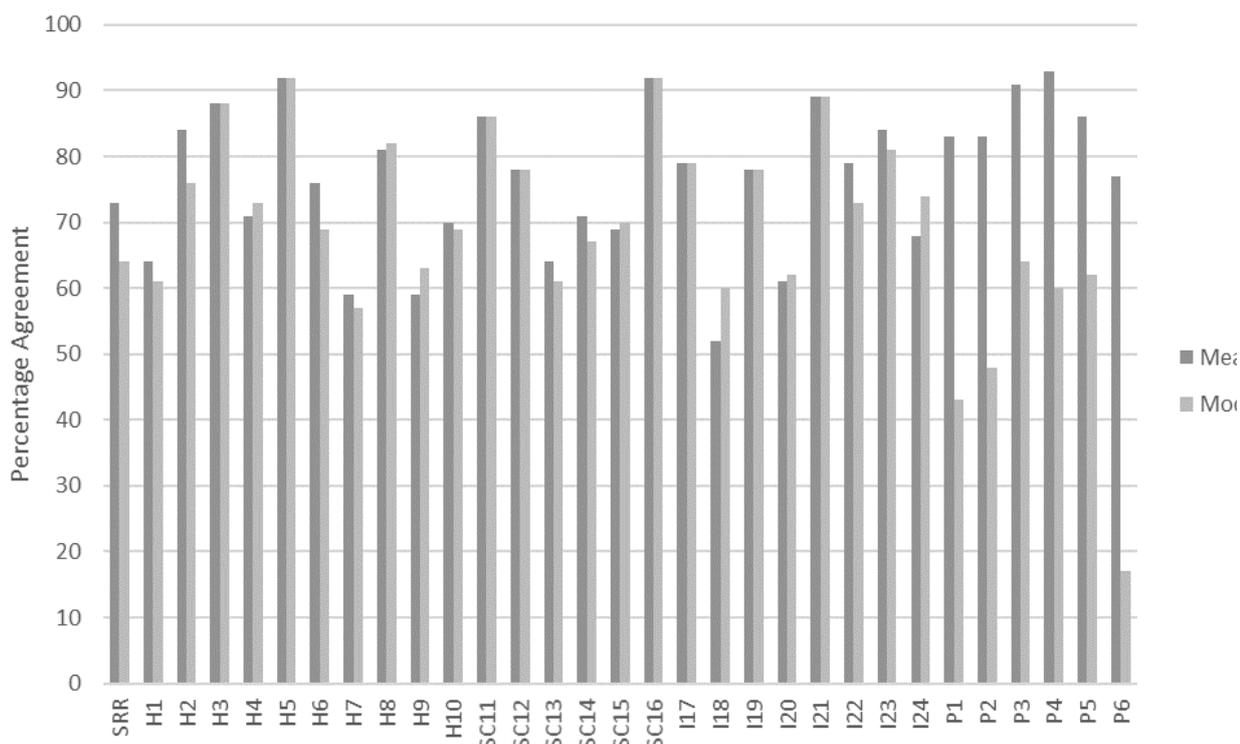


Figure 1. Percentage Agreement with Mean and Modal Ratings for SRR and All Items.

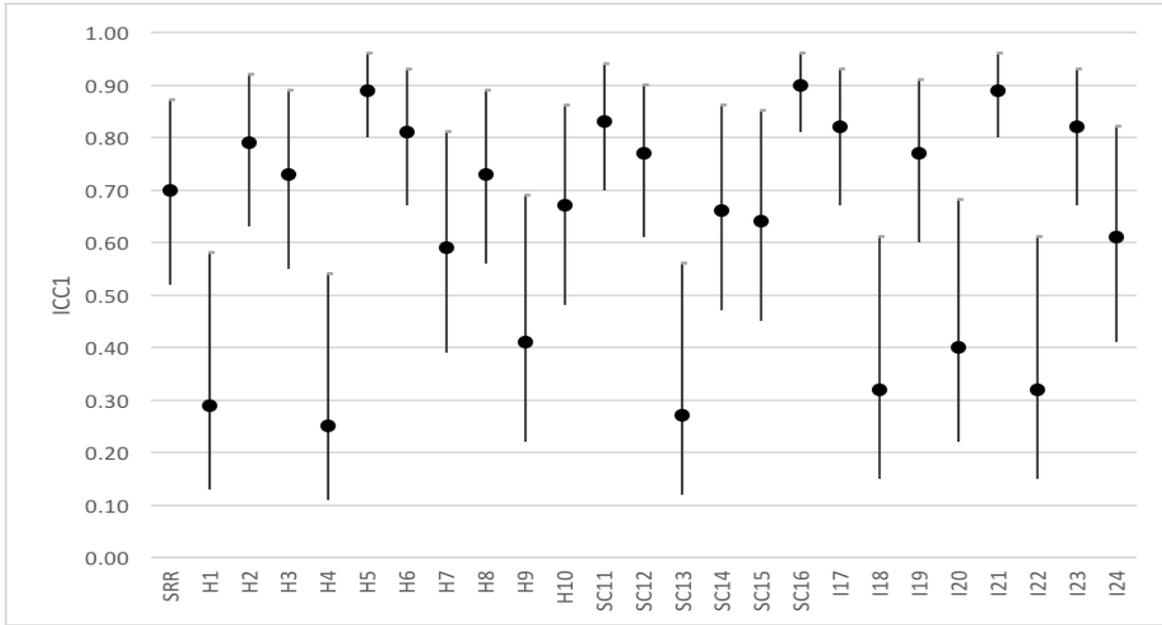


Figure 2. ICC1 and 95% confidence intervals for SRR and risk items.

Table 2
ICC1 for Domain and Total Risk Scores

Domain/Total Risk	ICC1	95% CI	
Historical	0.77	0.61	0.91
Social/ Contextual	0.73	0.56	0.89
Individual	0.83	0.70	0.94
Total Risk	0.86	0.74	0.95

Research Question 2 – *What is the level of agreement between ratings made by mental health professionals and experts?*

Percentage agreement between participants’ and expert SRR was 48%. Across individual items the average percentage agreement was 69% (41–92%). Half the items achieved a percentage agreement of $\geq 70\%$. This is comparable to the agreement level with the mode, but not the mean where 23 items achieved agreement of $\geq 70\%$ (Figure 1). Items with the highest level of agreement with experts were H5 (92%), SC16 (92%) and P3 (91%). Items with the lowest agreement were I18 (41%), I24 (45%), and SC13 (54%). Overall, comparing percentage agreement with the mean and mode and expert, participants tend to

have a lower level of agreement with experts compared with each other. Full results are shown in Appendix 16.

Figure 3 shows the percentage of the 12 vignettes rated Low, Moderate, or High on the SRR by participants and experts. Participants rated around twice as many vignettes as Low and around half as many as high compared to experts. Participant and expert SRR was similar for Moderate vignettes.

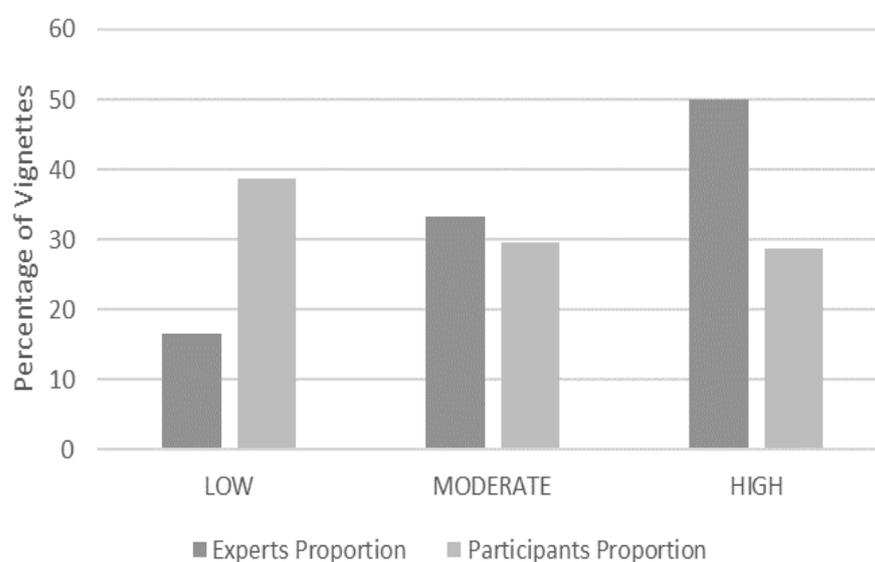


Figure 3. Comparison of expert and participants' SRR.

Research Question 3 – What is the association between the IRR and case characteristics?

Associations between vignette gender, severity of violence risk (SRR) and IRR were explored. Percentage agreement (mean, mode, expert) for the SRR and individual items were calculated for female and male vignettes. Participants showed a high level of agreement with female SRRs based on mean and modal scores (both 77%), but lower agreement with experts (47%). For males, participants showed lower agreement with each other 68% (mean) and 52% (mode), and 50% agreement with experts. Across all items the

average percentage agreements with mean, mode and expert ratings were 78%, 70% and 69% respectively for females, and 76%, 68% and 69% respectively for males. This indicates similarity in how participants rated the two groups of vignettes. See Appendix 17 for data and additional descriptive statistics. Figure 4 shows the proportion of male and female vignettes rated as Low, Moderate or High (SRR) by participants. A post-hoc analysis of the moderate rated vignettes showed that the proportion of female vignettes was not statistically significant ($z=-0.56$, $p=0.34$).

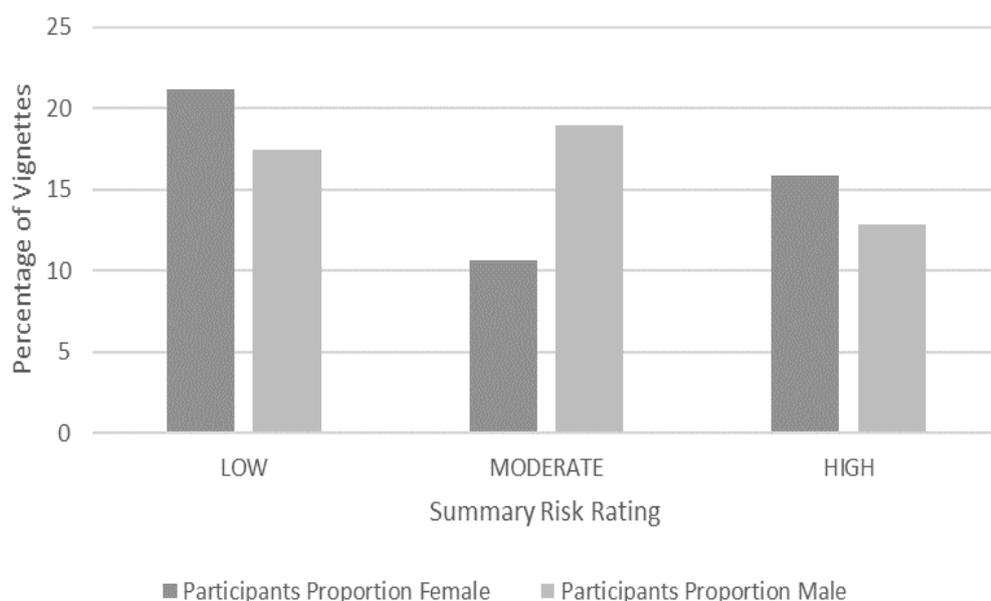


Figure 4. Comparison of Participants’ SRR by Vignette Gender.

ICC1 scores for individual risk item ratings ranged from 0.20 to 0.91 for female vignettes (Figure 5), and 0.23 to 0.91 for male vignettes (Figure 6). For female vignettes two items achieved ‘good’ and 13 items achieved ‘excellent’ levels of agreement. For the male vignettes eight items achieved ‘good’ and 9 items achieved ‘excellent’ levels of agreement. Table 3 shows ICCs for the three risk domains and Total Risk score. The agreement level and 95% confidence intervals are similar with male and female vignettes achieving ‘good’ or ‘excellent’ reliability.

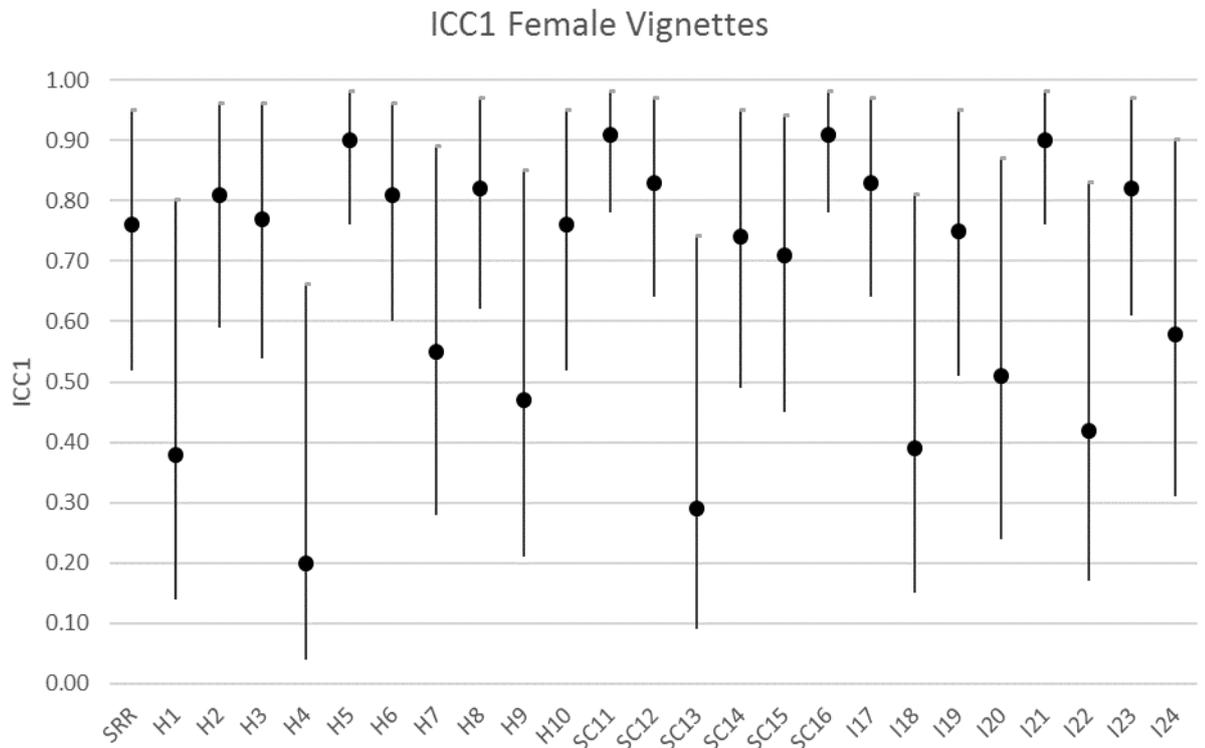


Figure 5. ICC1 and 95% confidence interval for SRR and risk items for female vignettes.

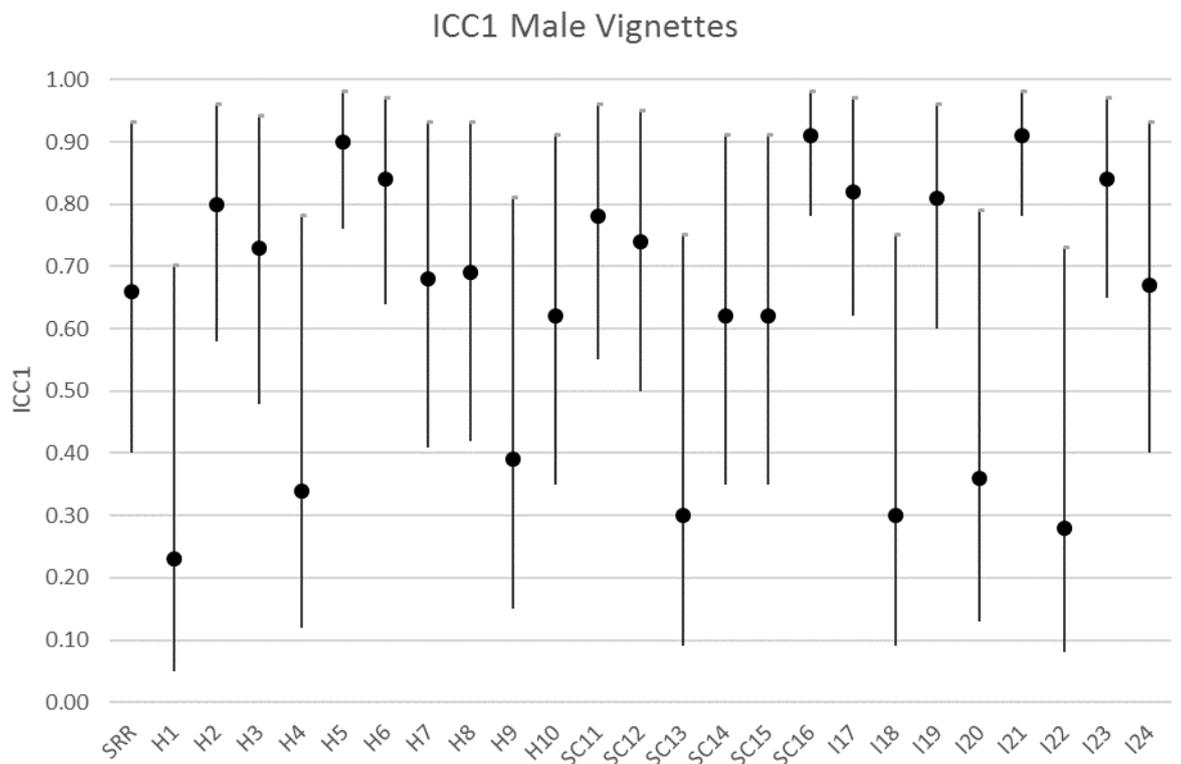


Figure 6. ICC1 and 95% confidence interval for SRR and risk items for male vignettes.

Table 3

ICC1 for Total Risk Score and Domains for Female and Male Vignettes

	Female			Males		
	ICC1	95% CI		ICC1	95% CI	
Historical	0.77	0.54	0.96	0.81	0.59	0.96
Social/ Contextual	0.79	0.56	0.96	0.71	0.46	0.94
Individual	0.86	0.68	0.97	0.83	0.64	0.97
Total Risk	0.87	0.71	0.98	0.86	0.69	0.98

Figure 7 shows the percentage agreement with the mean, mode and expert ratings across all vignettes. Agreement level was consistent across the gender pairs and those vignettes in the ‘Moderate’ group did not show lower levels of agreement than those in the ‘Low’ or ‘High’ groups. Participants agreed more with each other (mean and mode) than with experts. This discrepancy may reflect the process of amending vignettes following initial feedback from the experts. If experts highlighted insufficient information to rate a factor and further detail was added to the vignette then participants could be expected to show greater agreement as there was less ambiguity. Clearly having uneven numbers of vignettes in each risk group makes it more difficult to identify patterns in the data.

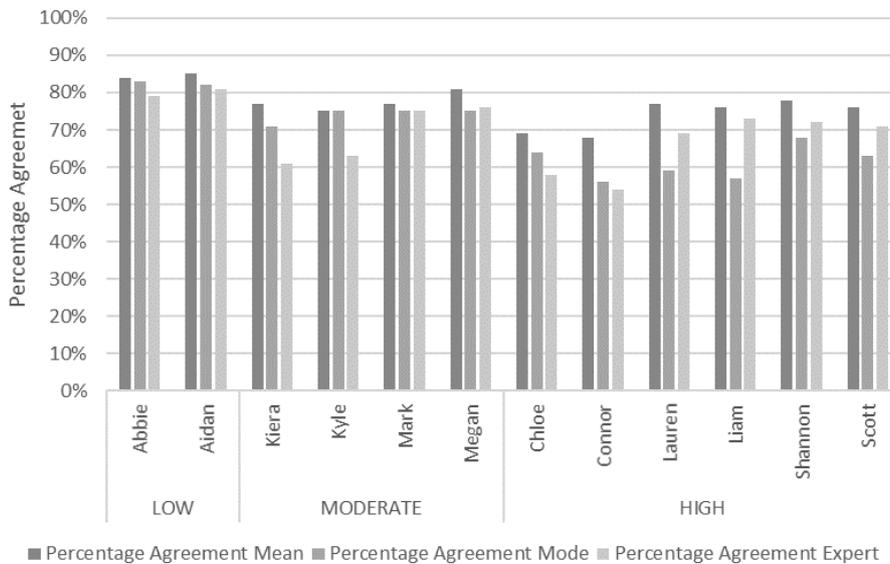


Figure 7. Percentage agreement with mean, mode and expert ratings across vignettes (all items combined).

Research Question 4 – What is the association between the IRR and rater characteristics (i.e. professional background, years of experience)?

Three participants had previously attended training using other risk assessment tools. No participants had attended prior training on the SAVRY, although six participants reported having used the SAVRY on between two and 30 occasions (mean=9; missing data n=1), reporting that it was “useful” (n=2) or “very useful” (n=3). Percentage agreement with experts across all items was calculated for each participant and ranged between 60% and 79%. Figure 8 shows the percentage agreement with experts by professional group. Most participants (n=16) achieved percentage agreement of $\geq 70\%$.

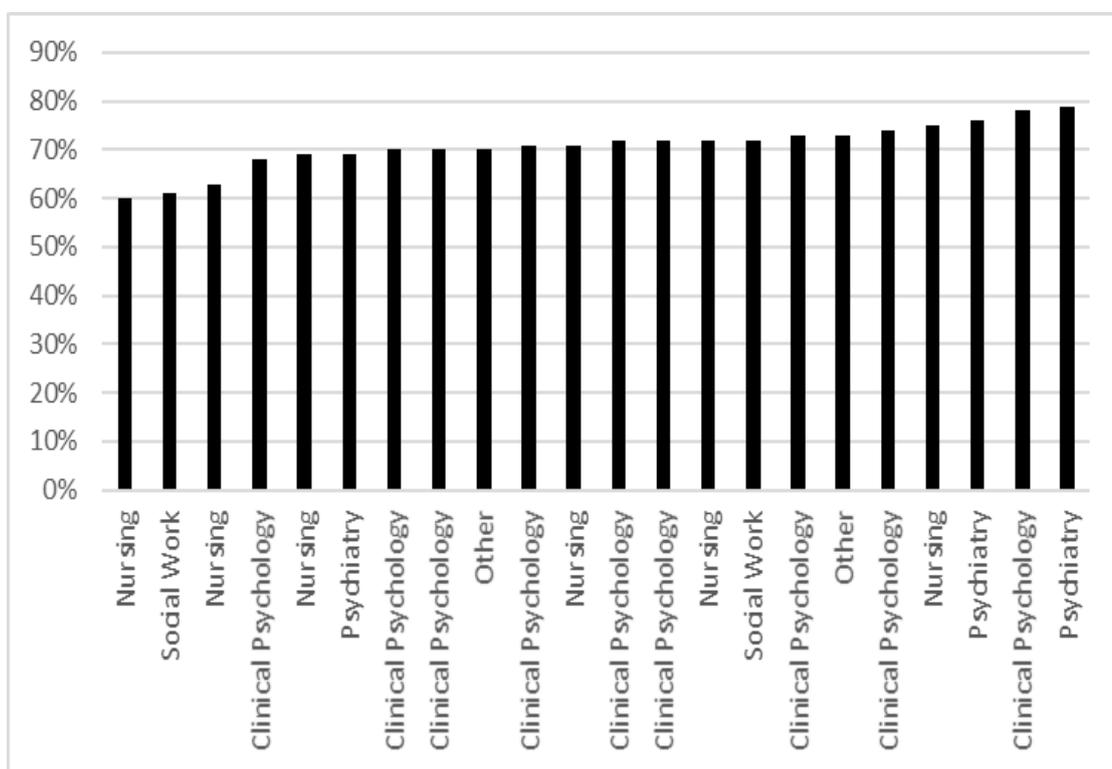


Figure 8. Percentage agreement with expert raters across all items calculated for each participant and grouped across profession.

Figure 9 shows the relationship between participants' experience in years and their percentage agreement with experts. Experience does not relate to higher level of agreement with experts. SRRs show greater variation compared to individual item ratings.

Participants rated their confidence in rating the SAVRY and how much they considered ratings to be influenced by their subjective feelings towards the vignette. Confidence and subjectivity were rated on 10-point Likert scales: 1 = 'Not at all confident' to 10 = 'Completely confident', and 1 = 'Not at all' and 10 = 'All the time' respectively. Mean rating for confidence was 5.9 (SD = 1.41), and 4.43 (SD = 1.57) for subjectivity. These variables have an inverse relationship with participants reporting greater confidence in their ratings when they considered themselves to be less subjective (Figure 10). The number of years qualified did not relate to higher levels of confidence or lower subjectivity in ratings (see Figures 11 and 12).

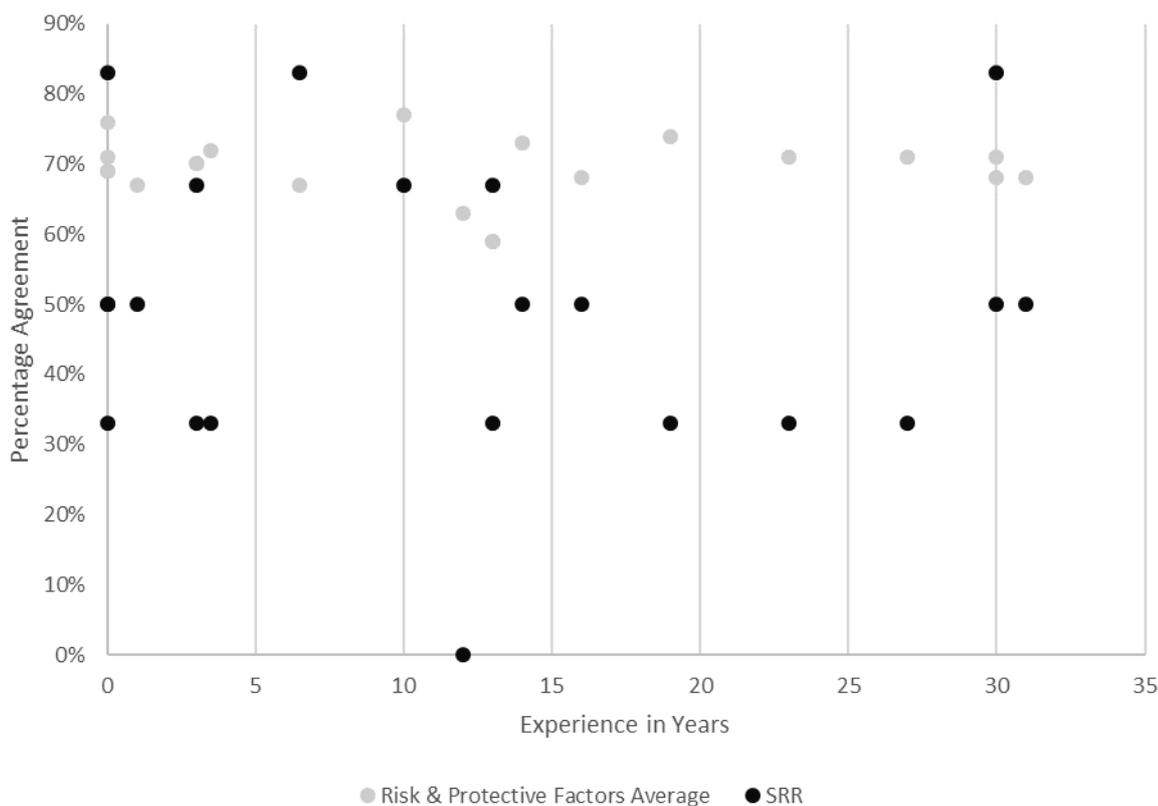


Figure 9. Percentage agreement with expert ratings comparing SRR and all individual factors combined.

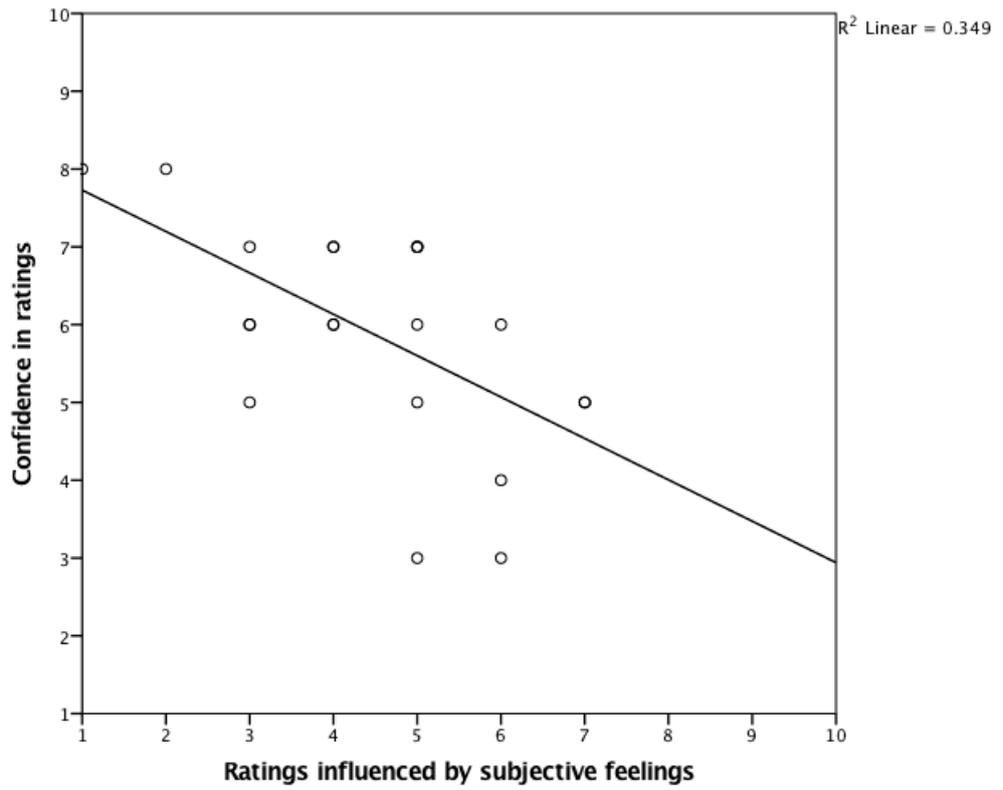


Figure 10. Rating confidence and subjective feelings about the vignette.

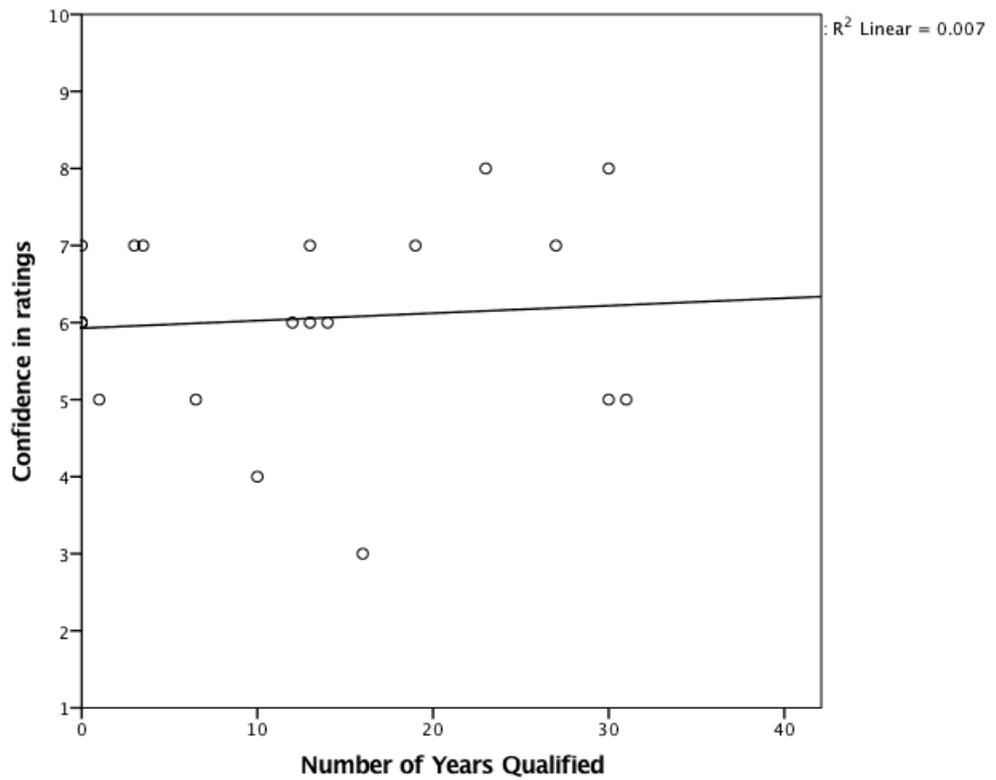


Figure 11. Number of years qualified and confidence in ratings.

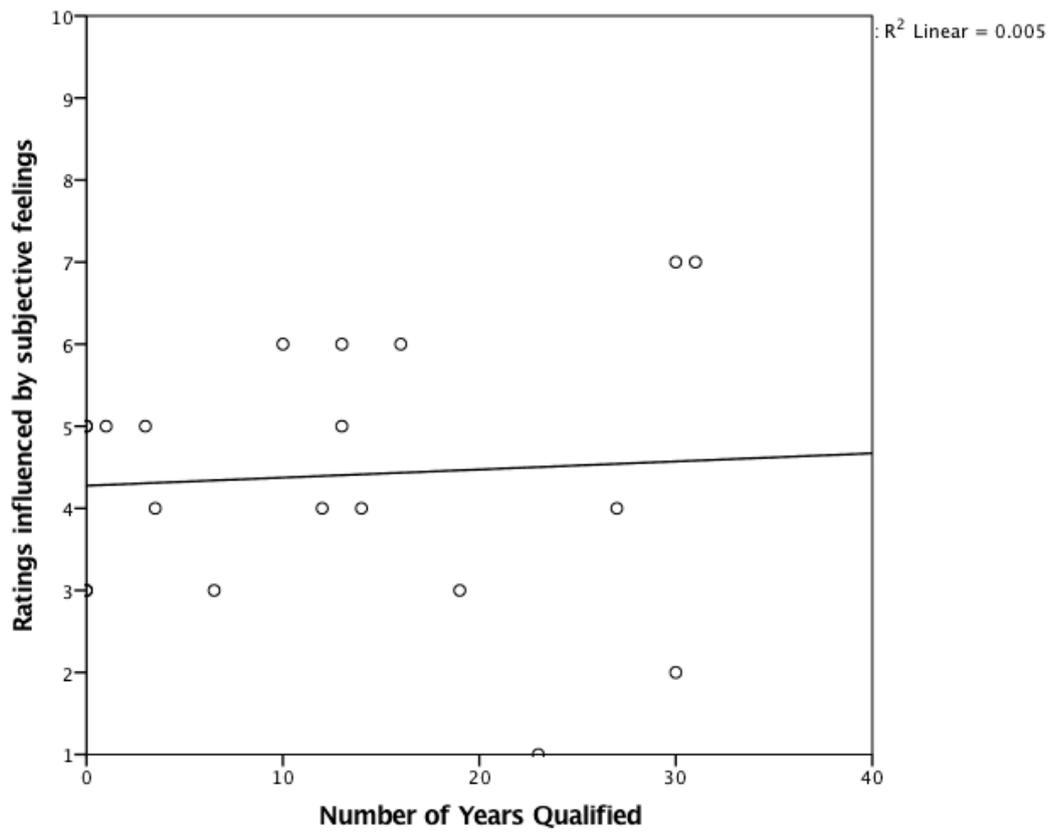


Figure 12. Number of years qualified and subjective feelings about vignette.

Discussion

The level of IRR achieved by mental health professionals in this study varies depending on the interpretation criteria used for the ICC1. Using the 95% confidence interval, as recommended by Koo and Li (2016), only three SAVRY items scored in the ‘good’ to ‘excellent’ domain. In contrast, using individual ICCs scores, most items achieved ‘good’ or ‘excellent’ reliability, and ‘good’ reliability for SRRs. These ICC scores are consistent with previous SAVRY IRR studies despite the participants having little or no prior experience of the SAVRY compared to the more experienced raters in the published literature. For percentage agreement statistics, level of agreement was higher amongst participants than with experts, which is perhaps unsurprising given that participants rated vignettes that had been refined based on feedback from the expert rater review process. This study had 22 participants, however, due to the study design 10 participants rated one set of vignettes and 12 rated the other set, therefore the total number of raters for each vignette was relatively small. The small number of participants and the number of vignettes rated may have impacted on the results, including the larger confidence intervals for the ICCs.

Case characteristics (gender, violence severity of the vignettes) did not significantly influence rating reliability. Previous research indicated the tendency to underestimate female violence (Skeem et al., 2005). However, there are notable differences with the current study. Skeem et al. assessed the accuracy of clinical judgement in predicting violence rather than using an SPJ tool. Consideration of risk and protective factors by using the SAVRY may aid clinicians in making more objective ratings rather than relying on heuristics in decision-making (see below for further discussion). This supports the findings of Child, Frick and Gottlieb (2016) who examined gender differences in the measurement invariance of the SAVRY. They found that the internal structure of risk was

invariant across sex. It should be noted that investigation of gender was a secondary aim of this study and the small sample limits the conclusions that can be drawn.

Participants with more confidence in their ratings also perceived greater objectivity in their decision-making. Interestingly, neither confidence nor subjectivity were significantly related to years of experience. The measure of experience was based on the number of years qualified rather than experience of completing SPJ risk assessments. A nurse with 15 years of post-qualification experience is unlikely to have had the same opportunities to complete SPJ risk assessments compared to psychologists or psychiatrists with a similar length of experience. This may explain in part why confidence did not relate to years of experience, although it is interesting to note that the percentage agreement across professional groups did not reveal notable differences.

In human decision-making, a new situation or individual is evaluated based on its similarity with previous situations or individuals a person has experienced. This is the 'representativeness heuristic' which is argued to influence the confidence people have in their decisions (Kahneman & Tversky, 1973). In risk assessment, this means a clinician judging one case based on its similarity to other cases they have encountered. Human decision-making research suggests that confidence is not a good indication of judgement accuracy. From the more limited research on confidence and accuracy in violence risk assessment results are equivocal (Desmarais, Nicholls, Read, & Brink, 2010). Whilst some show that confidence increases predictive accuracy (Douglas & Ogloff, 2003), others show few differences in accuracy as a function of confidence, and any significant differences found indicated that high confidence was associated with lower predictive accuracy (Desmarais et al., 2010). There do not appear to have been other studies on confidence and risk ratings using the SAVRY in the extant literature.

Limitations of the Current Work

Rating six vignettes was tiring for participants based on their self-report and may have impacted on their motivation and rating accuracy. When designing the study, consideration was given to allowing participants to complete the vignettes after training in their own time. However, it was decided that this could result in high levels of lost data so participants were required to complete the ratings at the training event.

Most participants were clinical psychologists or nurses. There were only two social workers. This sample population is not fully representative of the typical mental health professionals completing SPJ risk assessments, for example SPJ risk assessment workshops are currently only core training for psychologists and psychiatrists. Also, the number of participants for the analysis of vignette case characteristics was small and limits the conclusions that can be drawn due to the lack of statistical power.

The expert panel review of vignettes raised some challenges due to the degree of variation in initial ratings between experts. Whilst these variations were addressed through a process of discussion and addition of clarifying information as required, the amended vignettes were not re-rated by the experts. This introduced an additional source of potential variation when assessing the level of agreement between mental health professionals and experts due to the differences in vignettes. Using percentage agreement with expert ratings as a measure of reliability is therefore limited in its utility. It is relevant to note that this study placed greater demands on the experts than previous studies using similar methodology (Dickson, 2014; Sutherland et al., 2012). These studies provided pre-rated vignettes and experts provided feedback on their agreement or disagreement with the ratings. This alternative approach may have reduced some of the variability amongst experts compared to the current study.

Whilst the expert panel review process may have impacted on the lower percentage agreement found with expert ratings, there is also the issue of why agreement levels between participants were so much higher. Possible reasons include attendance at the training event increasing concordance in ratings due to improved understanding of the risk assessment process, or perhaps participants conferred on their ratings, despite being instructed not to do so. Nonetheless, percentage agreement statistics are limited in their utility as a measure of reliability due to the failure to account for chance when calculating agreement and therefore the use of ICC is important to ensure a more accurate assessment of IRR (see below).

Using vignettes limited the amount of information available for participants. There were time constraints in the expert review process and vignettes could have benefitted from further refinement. Nonetheless, no systematic difficulties were identified for participants in rating particular factors.

Future Directions

The set of 12 vignettes were developed specifically for this study. This was a time-consuming process and these materials could benefit future research on the SAVRY both extending the current study to increase participant numbers, and developing them for use in risk management planning. They could also be used in future training sessions for clinicians.

Previous IRR studies have used ICC scores to measure reliability. This study also reported the 95% confidence interval in addition to the ICC scores. Results from this study demonstrate that use of the ICC score alone can be misleading in measuring reliability. An ICC score may indicate an 'excellent' level of reliability, yet the 95% confidence interval shows that the true ICC value has a wide range from 'fair' to 'excellent'. The 95%

confidence interval is therefore more informative in reporting levels of reliability. Future IRR studies should take this into account when calculating statistical power and use the 95% confidence interval rather than individual ICC score to achieve a more accurate estimate of power.

Given the small sample size for aspects of the study and the limited range of multidisciplinary professionals it is recommended that the study is extended to increase participant numbers and professional diversity. Importantly, this would allow the potential impact of vignette gender to be explored in more detail. It would also be advisable to reduce the number of vignettes participants needed to rate or possibly give the option of posting responses back to allow them more time to complete. It would be interesting to explore how the SAVRY is used for risk management planning, for example, how factors ratings are linked to interventions with young people. In addition, there is a need for more research on the role of heuristics and biases in clinical judgements of violence risk assessment (Murray & Thomson, 2010) which could be explored with the SAVRY.

Conclusions

This study has contributed to the literature on the use of SPJ tools and the SAVRY specifically by exploring the reliability of mental health professionals with varying levels of experience in assessing risk. Professionals with varying levels of experience working with children and young people across health and social care are increasingly required to have the knowledge and skills to assess violence risk. Having greater understanding of their reliability in rating risk will assist in identifying future training needs. Using percentage agreement statistics this study found a generally good level of agreement amongst professionals, but lower levels of agreement with experts. Using ICC1 scores, the IRR was lower for items when interpreted using the 95% confidence interval but much

higher using individual ICC scores. Vignette gender did not conclusively impact on the rating of violence risk, however this finding should be explored further in future research due to the small numbers in this study. Professional characteristics did not appear to influence ratings. The results support previous findings from other SPJ tools that greater self-reported confidence in ratings does not indicate greater reliability. This indicates that ongoing training for professionals is necessary, regardless of their level of experience, to build awareness of relationship between confidence and accuracy.

References

- Bandura, A. (1986). *Social foundations of thought and action: A social-cognitive theory*. Upper Saddle River, NJ: Prentice Hall.
- Borum, R., Bartel, P., & Forth, A. (2006). *Structured Assessment of Violence Risk in Youth (SAVRY)*. Odessa, FL: Psychological Assessment Resources.
- Borum, R., Lodewijks, H., Bartel, P. A., & Forth, A. E. (2010). Structured Assessment of Violence Risk in Youth (SAVRY). In R. K. Otto & K. S. Douglas, K (Eds.), *Handbook of Violence Risk Assessment* (pp. 63-79). New York, NY, US: Routledge/Taylor & Francis Group.
- Borum, R., & Verhaagen, D. (2006). *Assessing and managing violence risk in juveniles*. New York: Guildford Press.
- Catchpole, R. E. H., & Gretton, H. M. (2003). The Predictive Validity of Risk Assessment with Violent Young Offenders: A 1-Year Examination of Criminal Outcome. *Criminal Justice and Behavior*, 30(6), 688-708.
- Centers for Disease Control and Prevention. (2016). Retrieved from <https://www.cdc.gov/violenceprevention/youthviolence/> on 11th January 2018.
- Childs, K. K., Frick, P. J., & Gottlieb, K. (2016). Sex Differences in the Measurement Invariance and Factors That Influence Structured Judgments of Risk Using the Structured Assessment of Violence Risk in Youth (SAVRY). *Youth Violence and Juvenile Justice*, 14(1), 76-92.

- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency, 86*(2).
- Desmarais, S. L., Nicholls, T. L., Read, J. D., & Brink, J. (2010). Confidence and accuracy in assessments of short-term risks presented by forensic psychiatric patients. *Journal of Forensic Psychiatry & Psychology, 21*(1), 1-22.
- Dickson, S. (2014). *The Psychopathy Checklist Youth Version (PCL:YV): an investigation into its inter-rater reliability*. Unpublished thesis.
- Dolan, M. C., & Rennie, C. E. (2008). The Structured Assessment of Violence Risk in Youth as a Predictor of Recidivism in a United Kingdom Cohort of Adolescent Offenders With Conduct Disorder. *Psychological Assessment, 20*(1), 35-46.
- Douglas, K. S., & Ogloff, J. R. P. (2003). The Impact of Confidence on the Accuracy of Structured Professional and Actuarial Violence Risk Judgments in a Sample of Forensic Psychiatric Patients. *Law and Human Behavior, 27*(6), 573-587.
- Gammelgård, M., Koivisto, A. M., Eronen, M., & Kaltiala-Heino, R. (2015). Predictive validity of the structured assessment of violence risk in youth: A 4-year follow-up. *Criminal Behaviour and Mental Health, 25*(3), 192-206.
- Guy, L. S., Douglas, K. S., & Hart, S. D. (2015). Risk assessment and communication. In B. Cutler & P. Zapf (Eds.), *APA handbook of forensic psychology: Vol.1. Individual and situational influences in criminal and civil contexts* (pp. 35-86). Washington, DC: American Psychological Association.

- Hart, S. D., Douglas, K. S., & Guy, L. S. (2016). The structured professional judgment approach to violence risk assessment: Origins, nature, and advances. In L. Craig, M. Rettenberger (Eds.), *The Wiley Handbook on the Theories, Assessment, Treatment of Sexual Offending: Volume II. Assessment* (pp643-666). Chichester: Wiley.
- Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the 'margins of error' of group v. individual predictions of violence. *The British Journal of Psychiatry*, 190(49), s60-S65.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237-251.
- 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.
- Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *The Lancet*, 360(9339), 1083-1088.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.
- Lodewijks, H., P.B., Doreleijers, T., A.H., de Ruiter, C., & Borum, R. (2008). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry*, 31(3), 263-271.
- McEachran, A. (2001). *The Predictive Validity of the PCL:YV and the SAVRY in a Population of Adolescent Offenders*. Unpublished thesis.

- Meyers, J. R., & Schmidt, F. (2008). Predictive Validity of the Structured Assessment for Violence Risk in Youth (SAVRY) With Juvenile Offenders. *Criminal Justice and Behavior, 35*(3), 344-355.
- Murray, J., & Thomson, M., E. (2010). Applying decision making theory to clinical judgements in violence risk assessment. *Europe's Journal of Psychology, 6*(2), 150.
- Skeem, J., Schubert, C., Stowman, S., Beeson, S., Mulvey, E., Gardner, W., & Lidz, C. (2005). Gender and Risk Assessment Accuracy: Underestimating Women's Violence Potential. *Law and Human Behavior, 29*(2), 173-186.
- Sutherland, A. A., Johnstone, L., Davidson, K. M., Hart, S. D., Cooke, D. J., Kropp, P. R., Logan, C., Michies, C., Stocks, R. (2012). Sexual violence risk assessment: an investigation of the interrater reliability of professional judgments made using the risk for sexual violence protocol. *International Journal of Forensic Mental Health, 11*(2), 119-133.
- Uebersax, J. (2015). Statistical Methods for Diagnostic Agreement. Retrieved from <http://www.john-uebersax.com/stat/agree.htm> on 11th January 2018.
- Viljoen, J. L., Scalora, M., Cuadra, L., Bader, S., Chávez, V., Ullman, D., & Lawrence, L. (2008). Assessing Risk for Violence in Adolescents Who Have Sexually Offended: A Comparison of the J-SOAP-II, J-SORRAT-II, and SAVRY. *Criminal Justice and Behavior, 35*(1), 5-23.
- Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine, 17*(1), 101-110.

WHO. (2016). World Health Organisation Youth Violence Factsheet. Retrieved from <http://www.who.int/en/news-room/fact-sheets/detail/youth-violence> on 11th January 2018.

International Journal of Forensic Mental Health

Instructions for authors

Thank you for choosing to submit your paper to us. These instructions will ensure we have everything required so your paper can move through peer review, production and publication smoothly. Please take the time to read and follow them as closely as possible, as doing so will ensure your paper matches the journal's requirements. For general guidance on the publication process at Taylor & Francis please visit our Author Services website.

SCHOLARONE MANUSCRIPTS™

This journal uses ScholarOne Manuscripts (previously Manuscript Central) to peer review manuscript submissions. Please read the guide for ScholarOne authors before making a submission. Complete guidelines for preparing and submitting your manuscript to this journal are provided below.

About the journal

International Journal of Forensic Mental Health is an international, peer-reviewed journal publishing high-quality, original research. Please see the journal's Aims & Scope for information about its focus and peer-review policy.

Peer review

Taylor & Francis is committed to peer-review integrity and upholding the highest standards of review. Once your paper has been assessed for suitability by the editor, it will then be double blind peer-reviewed by expert referees. Find out more about what to expect during peer review and read our guidance on publishing ethics.

Preparing your paper

Formatting and templates

Papers may be submitted in any standard file format, including Word and LaTeX. Figures should be saved separately from the text. The main document should be double-spaced,

with one-inch margins on all sides, and all pages should be numbered consecutively. Text should appear in 12-point Times New Roman or other common 12-point font.

Style guidelines

Submissions to *International Journal of Forensic Mental Health* should follow the style guidelines described in the *APA Publication Manual* (6th ed.). *Merriam-Webster's Collegiate Dictionary* (11th ed.) should be consulted for spelling.

References

References should be cited parenthetically in the text by author surname(s) and year, in accordance with *APA Publication Manual* guidelines:

<i>1 author</i>	(Smith, 2010)
<i>2 authors</i>	(Smith & Jones, 2010)
<i>3 to 5 authors</i>	(Smith, Jones, & Smythe, 2010) first mention; (Smith et al., 2010) thereafter
<i>6 or more authors</i>	(Smith et al., 2010)

When available, page numbers should be included in citations of direct quotations (e.g., (Smith, 2010, p. 25)).

References should be listed in a separate section at the end of the main text. All references in the list should be ordered alphabetically by the first author's surname. Examples of common reference types appear below

<i>Journal article</i>	Taylor, J., & Ogilvie, B. C. (1994). A conceptual model of adaptation to retirement among athletes: A meta-analysis. <i>Journal of Applied Sport Psychology</i> , 6(1), 1–20. doi:10.1080/10413209408406462
<i>Book</i>	Duke, J. A. (2001). <i>Handbook of phytochemical constituents of GRAS herbs and other economic plants</i> . Boca Raton, FL: CRC Press.
<i>Edited book chapter</i>	Gordon, S. (1995). Career transitions in competitive sport. In T. Morris & J. Summers (Eds.), <i>Sport psychology: Theory, applications and issues</i> (pp. 474–493). Milton, Australia: Wiley.
<i>Online/Website</i>	United States Census Bureau. (2014). <i>American housing survey: 2013 detailed tables</i> . Retrieved from http://www.census.gov/newsroom/press-releases/2014/cb14-tps78.html
<i>Dissertation/Thesis</i>	Allison, N. (1981). <i>Bacterial degradation of halogenated aliphatic acids</i> (Doctoral dissertation). Trent Polytechnic, Nottingham, UK.
<i>Conference</i>	Alfermann, D., & Gross, A. (1997, January). <i>Coping with career</i>

<i>presentation</i>	<i>termination: It all depends on freedom of choice.</i> Paper presented at the 9th Annual World Congress on Sport Psychology, Netanya, Israel.
<i>Paper/Report</i>	Grigg, W., Moran, R., & Kuang, M. (2010). <i>National Indian education study</i> (NCES 2010-462). Washington DC: National Center for Education Statistics.

Checklist: what to include

1. **Author details.** All authors of a manuscript should include their full name and affiliation on the cover page of the manuscript. Where appropriate, please also include ORCiDs and social media handles (Facebook, Twitter or LinkedIn). One author will need to be identified as the corresponding author, with their email address normally displayed in the published article. Authors' affiliations are the affiliations where the research was conducted. If any of the named co-authors moves affiliation during the peer-review process, the new affiliation can be given as a footnote. Please note that authorship may not be changed after acceptance. Also, no changes to affiliation can be made after your paper is accepted. Read more on authorship here.
2. **Abstract.** This summary of your article is normally no longer than 100 words. Read tips on writing your abstract.
3. **Keywords.** Keywords are the terms that are most important to the article and should be terms readers may use to search. Authors should provide 3 to 5 keywords. Please read our page about making your article more discoverable for recommendations on title choice and search engine optimization.
4. **Funding details.** Please supply all details required by your funding and grant-awarding bodies as follows:

For single agency grants

This work was supported by the <Funding Agency> under Grant <number xxxx>.

For multiple agency grants

This work was supported by the <Funding Agency #1> under Grant <number xxxx>;< Funding Agency #2> under Grant <number xxxx>; and <Funding Agency #3> under Grant <number xxxx>.

5. **Disclosure statement.** With a disclosure statement you acknowledge any financial interest or benefit that has arisen from the direct applications of your research. Further guidance, please see our page on what is a conflict of interest and how to disclose it.
6. **Supplemental online material.** Supplemental material can be a video, dataset, fileset, sound file, or anything else which supports (and is pertinent to) your paper. Supplemental material must be submitted for review upon paper submission. Additional text sections are normally not considered supplemental material. We publish supplemental material online via Figshare.
7. **Figures.** Figures should be high quality (600 dpi for black& white art and 300 dpi for color). Figures should be saved as TIFF, PostScript or EPS files. Figures embedded in your text may not be able to be used in final production.
8. **Tables.** Please supply editable table files. We recommend including simple tables at the end of your manuscript, or submitting a separate file with tables.
9. **Equations.** If you are submitting your manuscript as a Word document, please ensure that equations are editable. Please see our page on mathematical symbols and equations [for more information](#).

Author agreement / Use of third-party material

Authors are responsible for obtaining permission to reproduce copyrighted material from other sources and are required to sign an agreement for the transfer of copyright to the publisher. As an author you are required to secure permission if you want to reproduce any figure, table or extract text from any other source. This applies to direct reproduction as well as "derivative reproduction" (for which you have created a new figure or table which derives substantially from a copyrighted source). Please see our page on requesting permission to reproduce work(s) under copyright for more guidance. Authors are required

to sign an agreement for the transfer of copyright to the publisher. All accepted manuscripts, artwork, and photographs become property of the publisher.

Submitting your paper

International Journal of Forensic Mental Health uses ScholarOne Manuscripts to manage the peer-review process. If you have not submitted a paper to this journal before, you will need to create an account in ScholarOne Manuscripts. Please read the guidelines above and then submit your paper in the relevant Author Center, where you will find user guides and a helpdesk.

If you are submitting in LaTeX, please convert the files to PDF beforehand (you will also need to upload your LaTeX source files with the PDF). Your manuscript must be accompanied by a statement that it has not been published elsewhere and that it has not been submitted simultaneously for publication elsewhere.

We recommend that if your manuscript is accepted for publication, you keep a copy of your accepted manuscript. For possible uses of your accepted manuscript, please see our page on sharing your work.

CrossRef Similarity Check

Please note that *International Journal of Forensic Mental Health* uses CrossRef Similarity Check™ (Powered by iThenticate) to screen papers for unoriginal material. By submitting your paper to the journal you are agreeing to originality checks during the peer-review and production processes.

Color Reproduction

Color art will be reproduced in color in the online publication at no additional cost to the author. Color illustrations will also be considered for print publication; however, the author will be required to bear the full cost involved in color art reproduction. Please note that color reprints can only be ordered if print reproduction costs are paid. Print Rates: \$400 per figure for the first four figures; \$75 per figure for five or more figures. Art not supplied at a minimum of 300 dpi will not be considered for print. Please ensure that color figures and images submitted for publication will render clearly in a black and white conversion for print

Complying with funding agencies

We will deposit all National Institutes of Health or Wellcome Trust-funded papers into PubMedCentral on behalf of authors, meeting the requirements of their respective open access (OA) policies. If this applies to you, please ensure that you have included the appropriate funding bodies in your submission's funding details section. You can check various funders' OA policy mandates [here](#) and find out more about sharing your work [here](#).

Open access

This journal gives authors the option to publish open access via our Open Select publishing program, making it free to access online immediately on publication. Many funders mandate publishing your research open access; you can check open access funder policies and mandates [here](#).

Taylor & Francis Open Select gives you, your institution or funder the option of paying an article publishing charge (APC) to make an article open access. Please contact openaccess@tandf.co.uk if you would like to find out more, or go to our Author Services website.

For more information on license options, embargo periods and APCs for this journal please go [here](#).

Proofs

Page proofs are sent to the corresponding author using Taylor & Francis' Central Article Tracking System (CATS). They should be carefully checked and returned within 48 hours.

Reprints

Each corresponding author will receive 1 copy of the issue in which the article appears.

Authors for whom we receive a valid e-mail address will be provided an opportunity to purchase reprints of individual articles, or copies of the complete print issue. These authors will also be given complimentary access to their final article on Taylor & Francis Online.

For enquiries about reprints, please contact the Taylor & Francis Author Services team at reprints@tandf.co.uk. To order a copy of the issue containing your article, please contact our Customer Services team at Customer.Service@taylorandfrancis.com.

My Authored Works

On publication, you will be able to view, download and check your article's metrics (downloads, citations and Altmetric data) via My Authored Works on Taylor & Francis Online. We are committed to promoting and increasing the visibility of your article. Here are some tips and ideas on how you can work with us to promote your research.

Appendix 2 Database search terms

Population	child or juvenile* or youth* or adoles* or young people or young person or teen* or minors or boy* or girl* or male* or female* gender or ((sex or male or female) n2 (difference*))
Intervention/ Exposure	protect* or promot* or resilien* or strength* or asset* or resource* or moderat*
Comparator/ Outcome	offend* or delinquen* or crim* or convict* or detention* or prison* or incarcerat* viol* or conduct* or antisocial behavio* or criminal behavio* or correctional institute* or reformatories

PsycINFO (EBSCOhost)

Population (i)	Keywords (ti,ab,kw fields) Subject Headings	child* or juvenile* or youth* or adoles* or young* or teen* AG childhood or adolescence
Population (ii)	Keywords (ti,ab,kw fields) Subject Headings	((gender or sex or male or female) N3 (differen* or compar*)) DE "Human Sex Differences"
Comparator/ Exposure	Keywords (ti,ab,kw fields) Subject Headings	protect* or promot* or resilien* DE "Protective Factors" OR DE "Resilience (Psychological)" OR DE "Risk Management"
Outcome	Keywords (ti,ab,kw fields) Subject Headings	offend* or delinquen* or crim* or convict* or viol* or antisocial behavio* or conduct DE "Antisocial Behavior" OR DE "Behavior Disorders" OR DE "Female Delinquency" OR DE "Male Delinquency" OR DE "Criminal Behavior" OR DE "Juvenile Justice" OR DE "Juvenile Delinquency" OR DE "Juvenile Gangs"

Medline (Ovid) - Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present

Population (i)	Keywords Subject Headings	(child* or juvenile* or youth* or adoles* or young* or teen*).ti,ab,kw adolescent/ or child/ or child, preschool/
Population (ii)	Keywords Subject Headings	((gender or sex or male or female) ADJ3 (differen* or compar*)).ti,ab,kw Sex factors/ OR sex characteristics/
Comparator/ Exposure	Keywords Subject Headings	(protect* or promot or resilien*).ti,ab,kw Protective Factors/ OR Risk Factors/ OR Resilience, Psychological/
Outcome	Keywords Subject Headings	(offend* OR delinquen* OR crim* OR convict* OR viol* OR antisocial behavio* OR conduct).ti,ab,kw crime/ OR juvenile delinquency/ OR social

behavior disorders/ OR violence/ OR criminals/

Embase (Ovid) - Embase 1947-Present, updated daily

Population (i)	Keywords	(child* or juvenile* or youth* or adoles* or young* or teen*).ti,ab,kw
Population (ii)	Subject Headings	adolescent/ OR child/ OR child, preschool/
	Keywords	((gender or sex or male or female) ADJ3 (differen* or compar*)).ti,ab,kw
Comparator/ Exposure	Subject Headings	Sex Factors/ or Sex Characteristics/ or sex difference/ or "gender and sex"/
	Keywords	(protect* or promot* or resilien*).ti,ab,kw
Outcome	Subject Headings	Protective Factors/ or risk factors/ or Resilience, Psychological/
	Keywords	(offend* or delinquen* or crim* or convict* or viol* or antisocial behavio* or conduct).ti,ab,kw
	Subject Headings	crime/ or juvenile delinquency/ or social behavior disorders/ or violence/ or criminals/

Limit to English language

Limit to exclude medline journals

Limit to (conference abstract or conference paper or "conference review" or editorial)

CINAHL (EBSCOhost)

Population (i)	Keywords (ti,ab fields)	child* or juvenile* or youth* or adoles* or young* or teen*
Population (ii)	Subject Headings	(MH "Adolescence") OR (MH "Child")
	Keywords (ti,ab fields)	((gender or sex or male or female) n3 (differen* or compar*))
Comparator/ Exposure	Subject Headings	MH Sex Factors
	Keywords (ti,ab fields)	protect* or promot* or resilien*
Outcome	Subject Headings	(MH "Behavior and Behaviour Mechanisms") OR (MH "Psychological Well-Being")
	Keywords (ti,ab fields)	offend* or delinquen* or crim* or convict* or viol* or antisocial behavio* or conduct
	Subject Headings	(MH "Crime") OR (MH "Violence") OR (MH "Juvenile Delinquency")

ASSIA (Proquest)

Population (i)	Keywords (ti,ab fields only)	child* OR juvenile* OR youth* OR adoles* OR young* OR teen*
Population (ii)	Keywords (ti,ab fields only)	(gender or sex or male or female) near/3 (differen* or compar*)
Comparator/ Exposure	Keywords (ti,ab fields only)	protect* OR promot* OR resilien*
Outcome	Keywords (ti,ab fields only)	offend* OR delinquen* OR crim* OR convict* OR viol* OR antisocial behavio* OR conduct

Web of Science - Core Collection

Population (i)	Keywords (ti,ab,kw)	TS=(child* or juvenile* or youth* or adoles* or
----------------	---------------------	---

	using TOPIC (TS field)	young* or teen*)
Population (ii)	Keywords (ti,ab,kw using TOPIC (TS field)	TS=((gender or sex or male or female) near/3 (differen* or compar*))
Comparator/ Exposure	Keywords (ti,ab,kw using TOPIC (TS field)	TS=(protect* OR promot* resilien*)
Outcome	Keywords (ti,ab,kw using TOPIC (TS field)	TS=(offend* OR delinquen* OR crim* OR convict* OR viol* OR antisocial behavio* OR conduct)

Google Scholar

gender differences protective factors violence

Appendix 3 Outcome Measures for Violence and Protective Factors

Reference	Violence Outcome Scales/Measures	Protective Factors Scale/Measures
Blum, J, Ireland, M., Blum, R.W. (2003)	8 items scale: “In the past 12 months how often did you: use or think to use a weapon to get something from someone; take part in a group fight; pull a knife/gun on someone; shoot/stab someone; get into a serious physical fight; use a weapon in a fight; get into a fight where you had to be treated by a doctor or nurse; hurt someone badly enough to need bandages or care from a doctor or nurse?” The continuous scale was dichotomized at the 80 th percentile (reason for this not stated).	
Boyas, J.F., Kim, Y.J., Sharpe, T.L., Moore, D.J., Prince-Stehley, K. (2017)	Aggregate score based on four Likert-scale questions: “How many times have you gotten into a serious fight at school or work? How many times have you taken part in a fight where a group of your friends fought against another group? How many times have you attacked someone with the intent to seriously hurt them? How many times have you carried a handgun?”	<ul style="list-style-type: none"> ▪ Parental involvement - seven Likert-scale questions: In the past 12 months how often did your parents check if you had done your homework? How often did your parents provide help with your homework when you needed it? How often did your parents make you do chores around the house? How often did your parents limit the amount of time you watched TV? How often did your parents limit the amount of time you went out with your friends on school nights? How often did your parents let you know when you had done a good job? How often did your parents tell you they were proud of you for something you had done? ▪ Family composition – three questions about who lived in household (father, mother, total family members) ▪ Religious beliefs – are religious beliefs important to them, influence their decisions, is it important if friends share their beliefs? ▪ School connectedness – how often did the respondent feel school work was meaningful, how important were the things learned going to be, how interesting were their courses at school, and how often did their teachers tell them they were doing a good job? <p>Community engagement – “In how many different kinds of community-based activities, such as volunteer activities, sports, clubs, or groups have you participated?” Each factor was given a single aggregate score based on the multiple responses</p>
Brookmeyer, K.A., Henrich, C.C., Schwab-Stone, M. (2005)	Community violence perpetration measure modified from Richters and Martinez’s (1993) Survey of Children’s Exposure to Community Violence. Items included “in the last year have	<ul style="list-style-type: none"> ▪ Parent support: composite measure of six items ▪ Social cognitive processes

Reference	Violence Outcome Scales/Measures	Protective Factors Scale/Measures
Chui, W.H., Chan, H.C.O. (2012)	you hurt someone badly in a physical fight so that they had to be treated by a doctor or nurse?”, “in the past year have you started a fist fight or shoving match?”. Responses scored on a 5-point scale (0 times, 1 time, 2 times, 3-4 times, 5 or more times). Each severity category was scored from 0 to 4 and summed so that scores ranged from 0 to 24. Taken from 24-item questionnaire by Chapple et al. (2005) “Have you ever: slapped, shoved, or hit another student at school?, used force to get something you wanted from another person?, beaten up someone on purpose?” A single variable for violence was created by summing the total points for the 3 items.	Taken from 24-item questionnaire by Chapple et al. (2005).
Dornbusch, S.M., Erickson, K.G., Laird, J., Wong, C.A. (2001)	7 item scale: “During the past 12 months, how often did each of the following things happen? You pulled a knife or gun on someone? You shot or stabbed someone? How often did you get into a serious physical fight? How often did you use a weapon in a fight? How often did you hurt someone badly enough to need bandages or care from a doctor or nurse? How often did you use or threaten to use a weapon to get something from someone? How often did you take part in a fight where a group of your friends was against another group?” Frequency and intensity of violence scores were calculated using composite scores taken from the 8 items.	Questionnaire responses were scored on a 4 or 5-point Likert scale.
Griffin, K.W., Botvin, G.J., Scheier, L.M., Diaz, T., Miller, N.L. (2000)	5 item aggression scale: picking fights, hitting someone with the intention of hurting them, fighting if provoked, destroying others things, participating in group fights	
Nash, J.K., Mujanovic, E., Winfrey Jr, L.T. (2011)	6 items from the Youth Survey e.g. purposely damaged property, carried a hidden weapon, hit someone with the idea of hurting them	
Park, S., Morash, M., Stevens, T. (2010)	Frequency of assaults/attacks perpetrated in 2001 and 2002	
Pu, J., Chewning, B., St Clair, I.D., Kokotailo, P.K., Lacourt, J., Wilson, D. (2013)	Three items measured on a 3-point scale: In the past 3 months: Did you tell someone you were going to beat them up? Were you in a physical fight? Were you in a physical fight in which you were badly hurt? Responses measured on a 3-point scale (0: never, 1: not in the past 3 months,	

Reference	Violence Outcome Scales/Measures	Protective Factors Scale/Measures
Resnick, M.D., Ireland, M., Borowsky, I. (2004)	<p data-bbox="336 174 416 203">2: Yes)</p> <p data-bbox="336 327 767 689">Violence involvement scale, 8 items. “In the past 12 months how often did you: use or think to use a weapon to get something from someone; take part in a group fight; pull a knife/gun on someone; shoot/ stab someone; get into a serious physical fight; use a weapon in a fight; get into a fight where you had to be treated by a doctor or nurse; hurt someone badly enough to need bandages or care from a doctor or nurse.”</p>	

Appendix 4 SAVRY items

Historical items

- H1. History of violence
- H2. History of non-violent offending
- H3. Early initiation of violence
- H4. Past supervision/intervention failures
- H5. History of self-harm or suicide attempts
- H6. Exposure to violence in the home
- H7. Childhood history of maltreatment
- H8. Parental/caregiver criminality
- H9. Early caregiver disruption
- H10. Poor school achievement

Social/contextual items

- SC11. Peer delinquency
- SC12. Peer rejection
- SC13. Stress and poor coping
- SC14. Poor parental management
- SC15. Lack of personal/Social support
- SC16. Community disorganization

Individual items

- I17. Negative attitudes
- I18. Risk taking/impulsivity
- I19. Substance use difficulties
- I20. Anger management problems
- I21. Low empathy/remorse
- I22. Attention deficit/hyperactivity difficulties
- I23. Poor compliance
- I24. Low interest/Commitment to school or work

Protective items

- P1. Prosocial involvement
- P2. Strong social support
- P3. Strong attachments and bonds
- P4. Positive attitude towards intervention and authority
- P5. Strong commitment to school or work
- P6. Resilient personality

Appendix 5 Ethical Approval Letter



11/12/17

Dear Dr McLeod,

MVLS College Ethics Committee

Project Title: An Investigation of the Inter-rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) in Scotland

Project No: 200170042

The College Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study. It is happy therefore to approve the project, subject to the following conditions:

- Project end date: End July 2018
- The data should be held securely for a period of ten years after the completion of the research project, or for longer if specified by the research funder or sponsor, in accordance with the University's Code of Good Practice in Research: (http://www.gla.ac.uk/media/media_227599_en.pdf)
- The research should be carried out only on the sites, and/or with the groups defined in the application.
- Any proposed changes in the protocol should be submitted for reassessment, except when it is necessary to change the protocol to eliminate hazard to the subjects or where the change involves only the administrative aspects of the project. The Ethics Committee should be informed of any such changes.
- You should submit a short end of study report to the Ethics Committee within 3 months of completion.

Yours sincerely

A handwritten signature in black ink, appearing to be 'JD' with a flourish underneath.

Jesse Dawson
MD, BSc (Hons), FRCP, FESO
Professor of Stroke Medicine
NRS Stroke Research Champion / Clinical Lead for Scottish Stroke Research Network
Chair MVLS Research Ethics Committee

Institute of Cardiovascular and Medical Sciences
College of Medical, Veterinary & Life Sciences
Room M0.05
Office Block
Queen Elizabeth University Hospital
Glasgow
G51 4TF
Tel – 0141 451 5868
jesse.dawson@glasgow.ac.uk



Administrator: Mrs Elaine O'Neill
Telephone Number: 0141 232 1815
E-Mail: elaine.o'neill2@ggc.scot.nhs.uk
Website: www.nhsggc.org.uk/r&d

R&D Management Office
West Glasgow ACH
Dalnair Street
Glasgow G3 8SW

27 February 2018

Mrs Sarah E Selby
Mental Health Services
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow G12 0XH

NHS GG&C Board Approval

Dear Mrs S Selby,

Study Title: An investigation of the Inter-rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY)
Principal Investigator: Mrs Sarah E Selby
GG&C HB site: Forensic CAMHS – West Glasgow ACH
Sponsor: NHS Greater Glasgow and Clyde
R&D reference: GN18MH019
REC reference: n/a
Protocol no: V2, 24/01/18

I am pleased to confirm that Greater Glasgow & Clyde Health Board is now able to grant **Approval** for the above study.

Conditions of Approval

1. **For Clinical Trials** as defined by the Medicines for Human Use Clinical Trial Regulations, 2004
 - a. During the life span of the study GGHB requires the following information relating to this site
 - i. Notification of any potential serious breaches.
 - ii. Notification of any regulatory inspections.

It is your responsibility to ensure that all staff involved in the study at this site have the appropriate GCP training according to the GGHB GCP policy (www.nhsggc.org.uk/content/default.asp?page=s1411), evidence of such training to be filed in the site file.



An Investigation of the Inter-Rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) in Scotland

Vignette - Abbie

Reason for Referral

Abbie is a 16-year-old girl who lives at home with her father and step-mother. Abbie was recently charged, along with several other young people, with assault to severe injury following an alleged attack on two 15-year-old males. She has been bailed with a condition to engage with intensive social work supports.

Presentation at Interview

Abbie was articulate and polite at interview. She expressed remorse for her behaviour and was visibly upset about the physical injuries sustained by the victims. She engaged well in the interview and answered all the questions asked. She preferred to focus on the positive aspects of her life currently and plans for the future. She was more reluctant to speak about her childhood as she viewed it as “in the past”.

Family Life

Abbie was born and raised in an affluent suburb of East Renfrewshire. She is the only child from her parents’ relationship. Her parents, Robert and Elaine, separated when she was four years old. Abbie remained with her mother within the family home, and spent alternate weekends with her father. Relations between her parents have been described as tense and Abbie was exposed to arguments between them both during the time they lived together and after they separated. Elaine died of cancer when Abbie was 13 years old and she returned to live with her father and his new partner. Robert has a senior job in a bank which can result in him being away for long hours. Abbie has a lot of time to herself which she spends at home or out in the community with friends.

Abbie has a supportive extended family, in particular her maternal grandparents who also played an active role in her upbringing as a young child. Abbie spoke fondly of her grandparents. They have always believed in her and she was upset that she felt she had let them down. She worried that the stress of her legal situation might impact on their health. Since the assault, Abbie’s father has reduced his working hours so that he is more available to support Abbie.

Developmental History

Abbie was born in hospital 3 weeks early at 37 weeks. Concern was raised about possible developmental dysplasia of the hip at her neonatal assessment but upon further investigation no issues were identified and she has not experienced any difficulties as she has grown up. Abbie met all other developmental milestones.

Robert described having a good relationship with Abbie but there were some slight concerns about his emotional availability for her. Robert felt that Abbie was of an age where she should be more independent and self-reliant.

School History and Adjustment

Abbie's teachers describe her as very bright and academically capable. She was reported to have generally coped well with her transition to high school despite this occurring at a challenging time in her home life. There have been no concerns about her attainment until the last few months. She has continued to attend school but has started to fall behind in her performance in classes due to not completing some homework assignments. Two months ago, Abbie was involved in a "scuffle" with another pupil who made offensive comments about her family situation and not having a mother.

Peer Relationships

Abbie is reported by her teachers to be generally well-liked by peers. She has a group of friends from school, most of whom she also plays hockey with. She has spent less time with them in recent months, since sustaining an injury which prevented her from playing. It is reported that she began spending time with a group from school who are known truants. They are reported to be involved in fighting and other anti-social and nuisance behaviour. Abbie reported that, since her injury, she felt left out of her usual group of friends and did not want to spend time with them as much.

Substance Use and Lifestyle

Abbie admitted first drinking alcohol with friends when she was 13 years old. She reported that she has been drunk on a few occasions. She admitted that her alcohol use had increased over the past few months since socialising with a new group of friends. She reported that she smoked cannabis on one occasion a couple of months ago but did not enjoy the way it made her feel so has never used it again. She denied any other substance use.

Violent and Anti-Social Behaviour

Abbie reported that she intended to plead guilty to the offence. She seemed distressed at points when discussing it. Police reports indicated that Abbie was involved in the assault with five other young people. They had been drinking together in the park one evening and were walking back to one of their houses. Abbie denied that she was intoxicated and said they had just had "a few cans". They saw two boys walking towards them, one of whom was a boy from the year below at school who Abbie reported had recently moved to the school and had been trying to pick fights with older pupils, including some of Abbie's friends. Abbie reported that two of her friends started shouting and swearing at the two boys who retaliated by throwing stones at them. Abbie then said everything happened very quickly and "suddenly" everyone was fighting. She described being "caught up in the moment" and "acting without thinking". One of her friends shouted at Abbie to grab a metal pole that was lying nearby and hit the boy with it and Abbie did so. Abbie was uncertain how many times she hit him but said it was "a few". The boy sustained serious injuries to his chest and legs resulting in him being hospitalised for several weeks.

Abbie's family and teachers have reported that the assault was "out of character" and expressed shock and disbelief that she had been involved.

Mental Health

Abbie has no previous involvement with mental health services. She had been referred to the school counsellor following her mother's death but then decided not to go. Since the assault, she has described struggling more with her mood and feeling low. She reported that she had been having "bad dreams" about the assault and has been unable to walk past the location where it occurred. She denied any suicidal ideation or self-harm. In general, Abbie described herself as someone who focuses on what is happening in her life currently and thinking about the future rather than thinking about the past. She is also

aware that she has a tendency to avoid talking about her feelings and attributes this to her experiences growing up in a family where she was taught to just “get on with things”. Before the assault, her father and step-mother described her as a generally happy girl who copes well with life. They commented that she had been slightly more irritable and withdrawn in the weeks leading up to the assault however they attributed this to “typical teenage mood swings”.

Current Support/Intervention Plan

Abbie has attended all her appointments with her social worker and is reported to be engaging well.

Interests

Abbie enjoys playing computer games and is interested in computer programming. She has recently returned to playing hockey and is keen to re-engage with her old friends.

Goals and Future Plans

Abbie is keen to engage in further education and possibly go to university. She discussed how she has enjoyed learning about computer programming at school and is considering pursuing this for a career. Her father is helping her to look for suitable courses to apply for and she has recently joined an extra-curricular programming club at her school. Abbie also spoke about wanting to travel, possibly after she has been to university.

PARTICIPANT INFORMATION SHEET - EXPERT RATER

An Investigation of the Inter-Rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) in Scotland

You are being invited to take part in a research study investigating the inter-rater reliability (IRR) of the Structured Assessment of Violence Risk in Youth (SAVRY). Before you decide if you want to participate it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. You can contact us if there is anything that is not clear or if you would like more information before you decide about your involvement. Our contact details are included at the end of this information sheet.

1. What is the purpose of the study?

The SAVRY is a tool used internationally to assess the risk of violence in young people aged between 13 and 18 years old. It is used in NHS and youth justice settings in Scotland to make judgements about risk and need, and to inform treatment planning and risk management decisions.

This study aims to determine the level of IRR of the SAVRY in Scotland. IRR is the extent to which two or more raters get the same result when using the same tool. Although several studies have assessed IRR, no studies have included qualified professionals or been conducted in Scotland. This study will attempt to address these gaps. It will explore the extent to which raters agree with each other, and with expert raters. It will also explore the effect of rater characteristics on SAVRY ratings and IRR. Rater characteristics include how many years of experience and work setting professional variables.

2. Why have I been invited?

As a qualified psychologist, psychiatrist or social worker with several years' experience of completing adolescent risk assessments you have the relevant expertise to fulfil the role as an expert rater in this study.

3. Do I have to take part?

No, it is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason.

4. What will happen to me if I take part?

If you agree to take part, you will be sent 4 vignettes to rate using the SAVRY. You will also provide your feedback on the validity of the vignettes by completing a brief questionnaire. It should take around thirty minutes to rate each case, plus the time to complete the questionnaire. Thus, the total time for your participation time is 2-3 hours.

5. What are the possible disadvantages and risks of taking part?

Taking part in this study requires a commitment of time and effort. We greatly appreciate this commitment and understand that taking part in this study will be in addition to your workload. It is possible that case examples may contain details that you find disturbing. However, due to your job you are likely to be familiar with such material.

6. What are the possible benefits of taking part?

By taking part you will be contributing to an improved understanding of the use of the SAVRY.

7. Will my taking part in this study be kept confidential?

All the responses you give will be kept strictly confidential. You will be identified by a randomly assigned unique ID number. All information you provide will be stored in a locked filing cabinet within NHS premises or the Institute of Health and Wellbeing, Gartnavel Royal Hospital. An anonymised electronic copy of the study data will be stored on a password protected NHS computer or University network hard-drives. After a period of ten years, the data will be destroyed.

8. What will happen to the results of the research study?

Once the study is complete, we will produce a report that describes our findings. No personally identifiable information will be used in any report or publication. You may request to see a copy of the final report.

9. Who is organising and funding the research?

This study is being funded by the University of Glasgow and NHS Education Scotland as part of the University of Glasgow Doctorate of Clinical Psychology training programme.

10. Who has reviewed the study?

This study has been given ethical approval by the MVLS (Medical, Veterinary & Life Sciences) Ethics Committee for Glasgow University. It has also been approved by the University of Glasgow Doctorate of Clinical Psychology, Major Research Project submissions process.

11. Contact for Further Information

If you have any questions or concerns about any aspect of this study, please contact Sarah Selby. Alternatively, you can contact other members of the research team, Dr Jennifer McDonald or Dr Hamish McLeod. If you would like to speak to an independent person about this project then please contact Prof Tom McMillan (Tel no. 0141 211 3920). If you still have concerns and wish to complain formally, you can do this through the NHS Complaints Procedure.

Research Team:

Principal Investigator:

Sarah Selby
Trainee Clinical Psychologist
Tel: 0141 211 0607
Email: s.selby.1@research.gla.ac.uk
Email: Jennifer.McDonald@ggc.scot.nhs.uk

Project Supervised by:

Dr Jennifer McDonald
Consultant Clinical Psychologist
Tel: 0141 201 0808

Dr Hamish McLeod
Programme Director and
Honorary Consultant Clinical Psychologist
Tel: 0141 211 3922
Email: Hamish.Mcleod@glasgow.ac.uk

THANK YOU FOR TAKING THE TIME TO READ THIS INFORMATION SHEET



University
of Glasgow



PARTICIPANT CONSENT FORM – EXPERT RATER

An investigation of the Inter-rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY)

Please read each statement carefully and write your initials in the box if you agree with it.

1. I confirm that I have read and understood the Participant Information Sheet dated xx/xx/xxxx (version number X) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving reason.
3. I understand that the information I give will be confidential and that personally identifiable information will not be included in any report or publication of this study.
4. I agree to any comments I include in the study questionnaire being included in a written report or publication anonymously.
5. I understand that my information may be looked at by representatives of the study Sponsor, NHS GG&C, for audit purposes.
6. I agree to take part in the above study.

Name of Participant:

Date:

Signature:

Name of Principal Investigator:

Date:

Signature:

When completed: one copy to be retained by the participant and one copy to be returned to research team with study materials.

Version 1
20th February 2018

An investigation of the Inter-rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) in Scotland

Expert Panel Review

Case Being Reviewed _____

I would like to take this opportunity to thank you for participating in the expert review process for this study. The participants in this study will be required to rate six fictitious case vignettes using the SAVRY. We have developed case vignettes that vary in the severity of violence risk posed and introduced this expert review process as a way of maximising the validity of these cases. These case vignettes were developed to represent young people where it was felt that an assessment of violence risk may be warranted.

Your expert judgement will allow us to ensure the authenticity of this case and will be used within our data analysis.

A. Please state your rating for each of the following risk factors:

Historical items	Risk Rating
1. <i>History of violence</i>	<input type="checkbox"/> <i>LOW</i> <input type="checkbox"/> <i>MODERATE</i> <input type="checkbox"/> <i>HIGH</i>
2. <i>History of non-violent offending</i>	<input type="checkbox"/> <i>LOW</i> <input type="checkbox"/> <i>MODERATE</i> <input type="checkbox"/> <i>HIGH</i>
3. <i>Early initiation of violence</i>	<input type="checkbox"/> <i>LOW</i> <input type="checkbox"/> <i>MODERATE</i>

	<input type="checkbox"/> HIGH
4. Past supervision/intervention failures	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
5. History of self-harm or suicide attempts	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
6. Exposure to violence in the home	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
7. Childhood history of maltreatment	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
8. Parental/caregiver criminality	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
9. Early caregiver disruption	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
10. Poor school achievement	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH

Social/contextual items	Risk Rating
11. Peer delinquency	<input type="checkbox"/> LOW

	<input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
12. Peer rejection	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
13. Stress and poor coping	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
14. Poor parental management	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
15. Lack of personal/Social support	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
16. Community disorganization	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH

Individual items	Risk Rating
17. Negative attitudes	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
18. Risk taking/impulsivity	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH

19. Substance use difficulties	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
20. Anger management problems	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
21. Low empathy/remorse	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
22. Attention deficit/hyperactivity difficulties	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
23. Poor compliance	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH
24. Low interest/Commitment to school or work	<input type="checkbox"/> LOW <input type="checkbox"/> MODERATE <input type="checkbox"/> HIGH

Protective items	Rating
P1. Prosocial involvement	<input type="checkbox"/> PRESENT <input type="checkbox"/> ABSENT
P2. Strong social support	<input type="checkbox"/> PRESENT <input type="checkbox"/> ABSENT
P3. Strong attachments and bonds	<input type="checkbox"/> PRESENT

	<input type="checkbox"/> ABSENT
P4. Positive attitude towards intervention and authority	<input type="checkbox"/> PRESENT <input type="checkbox"/> ABSENT
P5. Strong commitment to school or work	<input type="checkbox"/> PRESENT <input type="checkbox"/> ABSENT
P6. Resilient personality	<input type="checkbox"/> PRESENT <input type="checkbox"/> ABSENT

B. In your own opinion, please state whether you think this case presents with an overall violence risk rating of Low, Moderate or High?

LOW

MODERATE

HIGH

C. Does this case vignette ‘feel’ authentic?

YES

NO

If you responded ‘No’, please suggest what information should be included, altered or removed to improve the authenticity of the case?

D. Are there any items where you feel there is insufficient information provided to rate the item? If so, please provide the name of the item and suggest what information should be included.

E. Are there any suggestions or improvements you would make for this case? If so, please state what information should be included, altered or removed to improve the case vignette.

Thank you for your participation.

PARTICIPANT INFORMATION SHEET

An Investigation of the Inter-Rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) in Scotland

You are being invited to take part in a research study investigating the inter-rater reliability (IRR) of the Structured Assessment of Violence Risk in Youth (SAVRY). Before you decide if you want to participate it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. You can contact us if there is anything that is not clear or if you would like more information before you decide about your involvement. Our contact details are included at the end of this information sheet.

1. What is the purpose of the study?

The SAVRY is a tool used internationally to assess the risk of violence in young people aged between 13 and 18 years old. It is used in NHS and youth justice settings in Scotland to make judgements about risk and need, and to inform treatment planning and risk management decisions.

This study aims to determine the level of IRR of the SAVRY in Scotland. IRR is the extent to which two or more raters get the same result when using the same tool. Although several studies have assessed IRR, no studies have included qualified professionals or been conducted in Scotland. This study will attempt to address these gaps. It will explore the extent to which raters agree with each other, and with expert raters. It will also explore the effect of rater characteristics on SAVRY ratings and IRR. Rater characteristics include how many years of experience and work setting professional variables.

The study is being conducted as in part fulfilment of the research component of the qualification for the Doctorate in Clinical Psychology.

2. Why have I been invited?

As a qualified professional working with children and young people you are likely to have experience of assessing and managing violence risk in young people. We are intending to compare the results of at least 22 clinicians as part of this study.

3. Do I have to take part?

No, it is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason.

4. What will happen to me if I take part?

If you agree to take part, you will be invited to attend a SAVRY training workshop which will be held on *[insert date]* at *[insert location]*. The training is provided free of charge. In the morning, you will receive training on using the SAVRY for risk assessment and management. In the afternoon, you will be given six fictitious case studies to rate using the SAVRY. Rating the case studies is part of the training workshop, but if you consent to participate in the study your responses will be used as data for the inter-rater reliability study. You will also complete a short questionnaire gathering professional information. It should take around thirty minutes to rate each case. Thus, in total, it should take approximately three hours for you to complete this study.

Refreshments, including lunch, will be provided.

5. What are the possible disadvantages and risks of taking part?

Taking part in this study requires a commitment of time and effort. We greatly appreciate this commitment and understand that taking part in this study will be in addition to your workload. It is possible that case examples may contain details that you find disturbing. However, as you are employed in an NHS or youth justice setting you are likely to be familiar with such material.

Unfortunately, it will not be possible to provide travel expenses if you attend.

6. What are the possible benefits of taking part?

By taking part you will develop your experience and skills in using the SAVRY. In addition, as this study is seeking to address gaps in current research you will be contributing to an improved understanding of the use of the SAVRY.

7. Will my taking part in this study be kept confidential?

All information collected about you and the responses you give will be kept strictly confidential. You will be identified by a randomly assigned unique ID number. All information provided will be stored anonymously in a locked filing cabinet in NHS premises at Forensic CAMHS in the West Glasgow Ambulatory Care Hospital. An anonymised electronic copy of the study data will be stored on a password protected NHS computer or University network hard-drives. After a period of ten years, the data will be destroyed. Comments you include in the study questionnaire may be included anonymously in a written report or publication. Your information may be looked at by representatives of the study Sponsor, NHS GG&C, for audit purposes.

8. What will happen to the results of the research study?

Once the study is complete, we will produce a report that describes our findings. This report will be included in the thesis submitted in part fulfilment of the Doctorate in Clinical Psychology. Direct quotes may be included anonymously in the report or subsequent publication, however no personally identifiable information will be used. You can request a summary of the results from which you

will be able to identify your own results by searching for your unique number. You may request to see a copy of the final report.

9. Who is organising and funding the research?

This study is being funded by the University of Glasgow and NHS Education Scotland as part of the University of Glasgow Doctorate of Clinical Psychology training programme.

10. Who has reviewed the study?

This study has been given ethical approval by the MVLS (Medical, Veterinary & Life Sciences) Ethics Committee for Glasgow University. It has also been approved by the University of Glasgow Doctorate of Clinical Psychology, Major Research Project submissions process.

11. Contact for Further Information

If you have any questions or concerns about any aspect of this study, please contact Sarah Selby. Alternatively, you can contact other members of the research team, Dr Jennifer McDonald or Dr Hamish McLeod. If you would like to speak to an independent person about this project then please contact Prof Tom McMillan (Tel no. 0141 211 3920). If you still have concerns and wish to complain formally, you can do this through the NHS Complaints Procedure.

Research Team:

Principal Investigator:

Sarah Selby
Trainee Clinical Psychologist
Tel: 0141 211 0607
Email: s.selby.1@research.gla.ac.uk
Jennifer.McDonald@ggc.scot.nhs.uk

Project Supervised by:

Dr Jennifer McDonald
Consultant Clinical Psychologist
Tel: 0141 201 0808
Email:

Dr Hamish McLeod
Programme Director and
Honorary Consultant Clinical Psychologist
Tel: 0141 211 3922
Email: Hamish.Mcleod@glasgow.ac.uk

THANK YOU FOR TAKING THE TIME TO READ THIS INFORMATION SHEET



University
of Glasgow



PARTICIPANT CONSENT FORM

An investigation of the Inter-rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY)

Please read each statement carefully and write your initials in the box if you agree with it.

1. I confirm that I have read and understood the Participant Information Sheet dated 23/01/2018 (version number 2) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving reason.
3. I understand that the information I give will be confidential and that personally identifiable information will not be included in any report or publication of this study.
4. I agree to any comments I include in the study questionnaire being included in a written report or publication anonymously.
5. I understand that my information may be looked at by representatives of the study Sponsor, NHS GG&C, for audit purposes.
6. I agree to take part in the above study.

Name of Participant:

Date:

Signature:

Name of Principal Investigator:

Date:

Signature:

When completed: one copy to be retained by the participant and one copy to be returned to research team with study materials.

Version 2
11th January 2018



An investigation of the Inter-rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) in Scotland

Participant Information Questionnaire

Participant ID Number: _____

Section A: Background Information

1. What is your gender?

Male Female

2. Which professional group do you belong to?

Social Work Clinical Psychology Forensic Psychology
Psychiatry Nursing Other *please*
state _____

3. How many years have you been qualified in your profession?

4. i) What setting(s) do you currently work in? (*please tick all applicable*)

Area/community team Residential Secure Inpatient

Other , please state _____

ii) How many years/months in this/these setting(s)?

5. i) What setting(s) have you previously worked in, if different from current
post? (*please tick all applicable*)

Area/community team Residential Secure Inpatient

Other , please state _____

ii) How many years/months in this/these setting(s) in total?

6. i) Do you have experience working in youth justice settings?

Yes No

ii) If applicable, how many years of experience do you have you working in youth justice settings?

	Are you familiar with this measure?	Have you attended formal training in this measure?	If yes, please estimate the number of days of formal training	Have you used this measure in your clinical practice?	If you have used this measure, please estimate how many times you have administered it.	How useful was this measure? 1 Not at all useful 2 3 4 5 Very useful
ASSET	Yes/No	Yes/No		Yes/No		
Early Assessment Risk List for Boys Version 2 (EARL-20B)	Yes/No	Yes/No		Yes/No		
Youth Level of Service Case Management Inventory (YLS-CMI)	Yes/No	Yes/No		Yes/No		
Structured Assessment of Violence Risk in Youth (SAVRY)	Yes/No	Yes/No		Yes/No		
Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR)	Yes/No	Yes/No		Yes/No		
Juvenile Sex Offender Protocol-II (JSOAP-II)	Yes/No	Yes/No		Yes/No		

Appendix 14 Calculating Percentage Agreement

The modal score for a SAVRY item was defined as expected as the most common score across all participants for a single vignette (that is up to 6 participants contributed to each modal score) with two equally popular scores being averaged such that 5 incidents of score 1, 5 incidents of score 2 and 2 incidents of score 3 would give a modal score of 1.5.

The mean score for a SAVRY item was defined as the mean score across all participants for a single vignette (that is up to 6 participants contributed to each mean score), rounded to the nearest whole number.

The method for calculating percentage agreement was consistently applied for evaluating agreement with the scores of expert raters and the mean and modal score of participants.

The percentage agreement score is defined as the proportion of participant scores for each SAVRY item that exactly match the expert rater score or the mean or modal scores of all participants. The SAVRY item scores included in any given percentage agreement were determined according to the results required, for example “all SAVRY item scores across all participants for a given vignette” (Figure 7).

Percentage agreement was calculated using the equation specified by Araujo and Born (1985). A worked example of the percentage agreement with the mean from Figure 1 for SAVRY item H1 follows. There were 6 vignettes rated by 6 participants and 6 vignettes rated by 5 participants giving 132 potential individual SAVRY item scores for H1. 2 participants failed to record a score on against one SAVRY item for one vignette. 85 SAVRY item scores matched the rounded mean for the respective vignette. 45 SAVRY item scores did not match the rounded mean for the respective vignette. The percentage agreement (rounded to the nearest one percent) is then $85 / (85 + 45 + 2) \times 100 = 64$.

SAVRY items where the participant failed to record score have been treated as a disagreement in the calculation of percentage agreement but have been excluded from the calculation of mean and modal scores.

A limitation of the modal score is that no participant can agree with a mode that is not a whole number. Both the modal and mean methods have a limitation in their potential sensitivity to additional or excluded data, for example a mean score of 2.51 (round to 3) could be reduced to 2.49 (rounded to 2) if an individual score of 3 were excluded from the analysis. A modal score could change from or to a split result (eg 1.5) with the addition or removal of a single rating. Thus care was exercised in assessing the impact of included or excluding data from the analysis.

Reference

Araujo, J., & Born, D. G. (1985). Calculating percentage agreement correctly but writing its formula incorrectly. *The Behavior analyst/MABA*, 8(2), 207-208.

Appendix 15 Rounding to the Nearest Integer (Method and Impact)

SAVRY risk factors and SRR are rated on a 3-point scale (Low, Moderate, High) and are not normally distributed. When calculating the modal score, for some items there were equal numbers of two different ratings therefore resulting in a modal score that was a fraction e.g. 2.5. Rounding to the nearest integer to achieve a whole number skews the percentage agreement since half of the participants would be categorised as not agreeing with the mode which is not representative of the actual ratings achieved in the study. Similarly, when calculating the mean, rounding scores to the nearest integer can provide a misleading percentage of how many people agree with the mean. For example, H2 has a mean score of 1.52 (see Table 4) and this would be rounded to the nearest integer of 2 to generate the percentage agreement statistic. However the modal score for H2 is 1. Therefore there will be a mismatch between the mean and modal percentage agreement which means they should not be directly compared. The median and interquartile ranges (IQR) are also reported in Table 4 to provide a more accurate summary of the data. The median and IQR are better able to describe data that is not normally distributed as the median is less sensitive to the influence of small changes in the data. For example, a true mean of 2.45 for a sample of 10 would be rounded to 2, and a small change such as a participant score changing from 2 to 3 would move the true mean to 2.55 and the rounded value to 3, dramatically altering the percentage agreement.

Table 4 Mean, mode and median scores for SAVRY individual and protective factors

	Mean	Mode	Median	IQR
H1	2.07	2	2.00	2.00 – 2.25
H2	1.52	1	1.00	1.00 – 2.00
H3	1.35	1	1.00	1.00 – 2.00
H4	1.32	1	1.00	1.00 – 1.25
H5	1.63	1	1.00	1.00 – 2.00
H6	1.90	1	2.00	1.00 – 3.00
H7	1.94	1	2.00	1.00 – 3.00
H8	1.65	1	1.00	1.00 – 2.00
H9	1.95	2	2.00	1.00 – 3.00

	Mean	Mode	Median	IQR
H10	1.87	1	2.00	1.00 – 2.75
SC11	1.82	1	2.00	1.00 – 3.00
SC12	1.92	1	2.00	1.00 – 3.00
SC13	2.40	2	2.00	2.00 – 3.00
SC14	2.13	3	2.00	1.00 – 3.00
SC15	1.89	1	2.00	1.00 – 3.00
SC16	1.92	1	1.00	1.00 – 3.00
I17	1.94	1	2.00	1.00 – 3.00
I18	1.82	2	2.00	1.00 – 2.00
I19	2.00	1*	2.00	1.00 – 3.00
I20	1.82	2	2.00	1.00 – 2.00
I21	1.79	1	1.00	1.00 – 3.00
I22	1.23	1	1.00	1.00 – 1.00
I23	1.66	1	1.00	1.00 – 3.00
I24	1.70	1	1.00	1.00 – 3.00
P1	0.44	0	0.00	0.00 – 1.00
P2	0.49	0	0.00	0.00 – 1.00
P3	0.65	1	1.00	0.00 – 1.00
P4	0.60	1	1.00	0.00 – 1.00
P5	0.63	1	1.00	0.00 – 1.00
P6	0.54	1	1.00	0.00 – 1.00
SRR	1.90	1	2.00	1.00 – 3.00

**Multiple mode exists. Smallest value is shown.*

Appendix 16 Results of Percentage Agreement with Expert

	<i>% Agreement with Expert</i>
SRR	48%
H1	61%
H2	78%
H3	88%
H4	55%
H5	92%
H6	71%
H7	57%
H8	79%
H9	59%
H10	73%
SC11	86%
SC12	64%
SC13	54%
SC14	55%
SC15	67%
SC16	92%
I17	77%
I18	41%
I19	64%
I20	55%
I21	57%
I22	79%
I23	74%
I24	45%
P1	71%
P2	83%
P3	91%
P4	87%
P5	66%
P6	59%

Appendix 17 Comparison of male and female vignettes

Table 5 Mean, mode, median ratings for individual items by gender

	Female				Male			
	Mean	Mode	Median	IQR	Mean	Mode	Median	IQR
H1	2.11	2	2.00	2.00 – 3.00	2.03	2	2.00	2.00 – 2.00
H2	1.47	1	1.00	1.00 – 2.00	1.56	1	1.00	1.00 – 2.00
H3	1.36	1	1.00	1.00 – 2.00	1.34	1	1.00	1.00 – 2.00
H4	1.35	1	1.00	1.00 – 2.00	1.28	1	1.00	1.00 – 1.00
H5	1.62	1	1.00	1.00 – 2.00	1.65	1	1.00	1.00 – 2.00
H6	1.88	1	2.00	1.00 – 3.00	1.92	1	2.00	1.00 – 3.00
H7	1.91	1	2.00	1.00 – 3.00	1.97	1	2.00	1.00 – 3.00
H8	1.63	1	1.00	1.00 – 2.00	1.68	1	1.00	1.00 – 2.00
H9	1.86	2	2.00	1.00 – 2.00	2.05	2	2.00	1.00 – 3.00
H10	1.79	1	2.00	1.00 – 2.00	1.95	2	2.00	1.00 -3.00
SC11	1.86	1	2.00	1.00 – 3.00	1.77	1	1.00	1.00 – 3.00
SC12	1.88	1	2.00	1.00 – 3.00	1.97	1	2.00	1.00 – 3.00
SC13	2.37	3	2.00	2.00 – 3.00	2.44	2	2.00	2.00 – 3.00
SC14	2.14	3	2.00	1.00 – 3.00	2.12	3	2.00	1.00 – 3.00
SC15	1.91	1*	2.00	1.00 – 3.00	1.86	1	2.00	1.00 – 3.00
SC16	1.88	1	1.00	1.00 – 3.00	1.95	1	2.00	1.00 – 3.00
I17	1.98	1	2.00	1.00 – 3.00	1.89	1	2.00	1.00 – 3.00
I18	1.80	1*	2.00	1.00 – 2.00	1.83	2	2.00	1.00 – 2.00
I19	2.00	1*	2.00	1.00 – 3.00	2.00	1*	2.00	1.00 – 3.00
I20	1.77	2	2.00	1.00 – 2.00	1.88	2	2.00	1.00 – 2.00
I21	1.86	1	2.00	1.00 – 3.00	1.72	1	1.00	1.00 – 3.00
I22	1.24	1	1.00	1.00 – 1.25	1.21	1	1.00	1.00 – 1.00
I23	1.68	1	1.00	1.00 – 3.00	1.64	1	1.00	1.00 – 3.00
I24	1.64	1	1.00	1.00 – 2.00	1.77	1	1.00	1.00 – 3.00
P1	0.48	0	0.00	0.00 – 1.00	0.39	0	0.00	0.00 – 1.00
P2	0.48	0	0.00	0.00 – 1.00	0.50	0*	0.50	0.00 – 1.00
P3	0.66	1	1.00	0.00 – 1.00	0.63	1	1.00	0.00 – 1.00
P4	0.58	1	1.00	0.00 – 1.00	0.62	1	1.00	0.00 – 1.00
P5	0.64	1	1.00	0.00 – 1.00	0.62	1	1.00	0.00 – 1.00
P6	0.62	1	1.00	0.00 – 1.00	0.46	0	0.00	0.00 – 1.00
SRR	1.89	1	2.00	1.00 – 3.00	1.91	2	2.00	1.00 – 3.00

*Multiple modes exist, the smallest value is shown

Table 6 Percentage Agreement (mean, mode, expert) for male and female vignettes

	Female Vignettes			Male Vignettes		
	Mean	Mode	Expert	Mean	Mode	Expert
SRR	77%	77%	47%	68%	52%	50%
H1	65%	65%	59%	64%	58%	64%
H2	91%	91%	79%	77%	61%	77%
H3	89%	89%	89%	86%	86%	86%
H4	70%	74%	55%	73%	73%	55%
H5	92%	92%	92%	91%	91%	91%
H6	77%	70%	68%	74%	68%	74%
H7	52%	55%	55%	67%	59%	59%
H8	85%	85%	79%	77%	79%	79%
H9	65%	65%	59%	53%	61%	59%
H10	79%	79%	79%	62%	59%	68%
SC11	89%	89%	89%	82%	82%	82%
SC12	79%	79%	65%	77%	77%	64%
SC13	56%	52%	52%	71%	71%	56%
SC14	70%	70%	48%	73%	65%	62%
SC15	70%	71%	71%	68%	68%	64%
SC16	92%	92%	92%	91%	91%	91%
I17	77%	77%	74%	80%	80%	80%
I18	55%	59%	42%	50%	61%	39%
I19	76%	76%	59%	80%	80%	68%
I20	71%	64%	61%	50%	61%	50%
I21	86%	86%	62%	91%	91%	52%
I22	82%	74%	82%	76%	71%	76%
I23	82%	83%	76%	86%	79%	73%
I24	68%	71%	45%	68%	77%	44%
P1	85%	47%	70%	82%	39%	73%
P2	80%	47%	80%	86%	50%	86%
P3	89%	65%	89%	92%	62%	92%
P4	91%	58%	88%	95%	62%	86%
P5	86%	62%	67%	86%	62%	65%
P6	82%	17%	52%	73%	18%	67%



University
of Glasgow

DOCTORATE IN CLINICAL PSYCHOLOGY

SUBMISSION FRONT PAGE

Name: Sarah Selby

Matriculation Number: 2230365

Name of Assessment: MRP Proposal

Title of Project: An investigation of the Inter-rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY)

Academic Supervisor: Dr Hamish McLeod

Field Supervisor: Dr Jennifer McDonald

Submission Date to Supervisor: 30.01.17

Version Number: 1

Word Count: 3243

DOCTORATE IN CLINICAL PSYCHOLOGY

SUBMISSION COVER PAGE

Matriculation Number: 2230365s

Name of Assessment: MRP Proposal

Title of Project: An investigation of the Inter-rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY)

Date of Submission: 30.01.17

Version Number: 1

Word Count, including reference list (excluding appendices): 3243
(maximum word count is 3,000)

Abstract

The Structured Assessment of Violence Risk in Youth (SAVRY) is a risk assessment tool used internationally to assess the risk of violence in young people aged between 13 and 18 years old. It is based on the Structured Professional Judgement (SPJ) approach and involves the detailed evaluation of a set of empirically-based risk factors for violent behaviour to inform risk formulation, scenario planning, and risk management. This study intends to extend existing research into the inter-rater reliability (IRR) of the SAVRY by examining the impact of rater and case study characteristics on the reliability of risk ratings. This will include the impact of gender bias on decision making and the impact of professional background and level of experience of raters. In this study, raters will be clinicians recruited from health and social work services in Scotland. This contrasts with previous IRR studies which have tended to use academic staff. Experts in adolescent violent risk assessment will also rate the case studies to provide 'gold standard' ratings against which to assess clinician rater performance.

Introduction

Youth Violence

Violence can be defined as “the intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community, that either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment or deprivation” (Krug, Dahlberg, Mercy, Zwi, Lozano, 2002). This definition is necessarily broad as it seeks to convey the potential complexity of violent behaviour. Youth violence also tends to be described in broad terms, for example “harmful behaviours that can start early and continue into young adulthood. The young person can be a victim, an offender, or a witness to the violence” (Centers for Disease Control and Prevention, 2016). Violence is recognised as a significant problem across the world with the World Health Organisation (WHO) describing it as a “global public health problem”. Homicide is the fourth leading cause of death amongst young people aged 10-29 years old with an estimated 200 000 homicides occurring worldwide each year (WHO, 2016). For every death, many more young people suffer non-fatal injuries due to violence which can lead to lifelong physical, psychological and social consequences.

Violence Risk Assessment

Given the scale of the problem, the ultimate goal is to stop youth violence before it starts. However, for those young people already engaging in violent behaviour understanding why they are at risk is crucial to implement effective interventions to minimise the risk of future harm for themselves and others. The process of assessing risk has evolved from unstructured clinical judgements focusing on risk prediction to the Structured Professional Judgement (SPJ) approach which draws on the strengths of actuarial and clinical judgements and uses empirically supported risk factors to assess the behaviour in question. The SPJ approach is more concerned with developing an understanding of the presence and relevance of risk factors for the individual to support risk management, rather than making direct assertions about the predictability of a behaviour. This approach to risk assessment is strongly supported within the extant empirical and professional literature (Sutherland et al., 2012).

Structured Assessment of Violence Risk in Youth

The Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2006) is a risk assessment tool used internationally to assess violence risk in young people aged between 13 and 18 years old. Based on the SPJ approach, it evaluates a set of empirically-based risk factors for violent behaviour to inform risk formulation, scenario planning, and risk management.

The SAVRY defines violence as “an act of physical battery sufficiently severe to cause injury that would require medical attention, a threat with a weapon in hand, or any act of forcible sexual risk”. Risk arises from the dynamic, reciprocal interaction between factors that increase and decrease the likelihood of offending in the developing young person over time. There are 24 risk factors in three domains (Historical, Social/Contextual, and Individual/Clinical Factors). The SAVRY uses a descriptive rating code (Low, Moderate, High) rather than numerical scores. Six items are also included as Protective Factors, these are rated as Present or Absent. An overall judgment of the risk and protective factors is made in a Summary Risk Rating (SRR), again using the descriptive code.

The authors are clear that the SAVRY should not be used to quantify risk, but rather to structure risk assessment and formulation based on the risk and protective factors. Furthermore, they assert that how an item is coded is less critical than the assessment of how a particular factor is associated with violence.

The SAVRY can be used to assess risk for males and females. The SAVRY authors report that many risk and protective factors operate similarly for both genders, although sensitivity and rates of exposure for each may differ (Borum, Lodewijks, Bartel, Forth, 2010).

Reliability and Validity of the SAVRY

Research into the psychometric properties of the SAVRY has identified high levels of predictive validity. For example, Gammelgård et al. (2015) showed it to be a good predictor of violent and general offending over a four-year period. However, it has been argued that traditional concepts of predictive reliability and validity may not be the most useful method of assessing the utility of SPJ tools (Hart, Michie, & Cooke, 2007; Sutherland et al., 2012). Hart et al. (2007) highlight that the error margins are unacceptably large and therefore we cannot predict the future with any certainty. Crucially, the SPJ approach is about the prevention and management of risk rather than its prediction.

Focusing on predictive validity could be argued to be of less relevance than better understanding how the SAVRY can inform risk management decisions.

Given the central role of professional judgement in the SPJ approach it is argued that the evaluation of inter-rater reliability (IRR) should be a particular focus of research (Sutherland et al., 2012). Six studies have examined the SAVRY's IRR. Results show good to excellent agreement between raters with ICCs ranging from .81 to .97 for the SAVRY Risk Total, and .72 and .95 for the SAVRY Summary Risk Rating (Catchpole & Gretton, 2003; Dolan & Rennie, 2008; Lodewijks et al., 2008; McEachran, 2001; Meyers & Schmidt, 2008; and Viljoen et al., 2008). Nonetheless, these studies are argued to be limited in scope. Firstly, comparisons were made between two or three raters who are academics rather than clinicians. Research suggests that the professional background of raters may be an important moderator of validity and reliability (Sutherland et al., 2012). Secondly, they assessed total and domain scores but not individual item scores (Borum et al., 2010). Given the key purpose of the tool is risk management then an understanding of specific risk factors is essential to developing an idiosyncratic risk management plan. Thirdly, these studies did not address the possible impact of gender bias in violence risk assessment. Evidence suggests that mental health professionals underestimate future violence by females, and this is not due to gender-related differences in violence (Skeem et al., 2005). Fourthly, these studies did not examine the impact of case specific factors on IRR. The level of complexity and risk can affect the degree of IRR with cases at the high or low extremes achieving higher rates of IRR than those in between (Sutherland et al., 2012). Finally, whilst one study has been conducted in the UK, no research has been undertaken in Scotland. Further reliability studies are therefore warranted to build on the existing empirical literature, and assess how this tool is used by health and social care professionals within Scotland.

Aims/Hypotheses

This research aims to assess:

- The level of IRR achieved by Mental Health professionals using the SAVRY to assess violence risk.
- The level of agreement between ratings made by Mental Health professionals and experts (professionals with expertise in the use of the SAVRY).
- The association between the IRR and case characteristics (i.e. gender, severity of violence risk).
- The association between the IRR and rater characteristics (i.e. professional background, years of experience).

Hypotheses:

1. Raters with less experience of risk assessment will result in lower IRR scores compared to the published literature that has relied on more experienced raters.
2. IRR will be higher for cases with low or high levels of violence risk compared to cases with a moderate level of violence risk (risk level defined by the SRR).
3. Based on the observation that female violence risk is typically underestimated, the level of IRR will be lower for female than male case studies.
4. Raters with more SAVRY experience/training (measured using a self-report questionnaire) will demonstrate greater concordance with 'gold standard' ratings and have higher rates of IRR.

Plan of Investigation

Participants. NHS staff will be recruited from Tier 3 Child and Adolescent Mental Health Services across Greater Glasgow and Clyde. This group is likely to include psychology, psychiatry and nursing staff. Social workers will be recruited from youth justice services across Glasgow.

Inclusion and Exclusion Criteria. The SAVRY manual states that professionals who use the SAVRY do not require formal training in the tool but should meet the minimum criteria of having expertise (i.e. knowledge, training and experience) in child/adolescent development, youth violence and delinquency, and conducting individual assessments.

“...generally advised that psychologists, psychiatrists, trained youth probation officers, and social workers would have the requisite expertise to use the SAVRY” (Borum et al., 2006). Individuals who have completed professional training in these areas will therefore be included, trainees in these fields will be excluded.

Recruitment Procedures. An email invitation to the SAVRY training workshop will be sent to relevant NHS and social work employees. The invitation will specify the optional participation in the IRR study. Further invitation to participate in the study will be offered at the training event(s).

Measures.

- SAVRY
- Staff information questionnaire based on previous research (Sutherland et al., 2012; Dickson, 2014). This will request information on professional characteristics such as number of years qualified, participation in previous training on the SAVRY and/or other SPJ tools.

Design.

3x2 Factorial Design. Independent variables are:

- Level of violence risk (low, moderate, high)
- Gender of case vignette (male, female)

Research Procedures.

- Six fictitious case vignettes (2 x low, moderate, high violence risk) will be developed by the researcher in collaboration with the Dr McDonald and other clinicians in the Forensic Child and Adolescent Mental Health Service.
- Case vignettes will be replicated and gender changed to create a total of 12 vignettes.
- Clinicians experienced in adolescent risk assessment will be contacted to request their participation as ‘expert’ raters for the ‘gold standard’ judgments. Likely to include the authors of the SAVRY and clinicians from NHS Scotland with several years’ experience of completing adolescent risk assessments. Six expert raters are required.

- Vignettes will be given to two expert raters for their feedback. Following any amendments made on the basis of the feedback, two vignettes will then be randomly allocated to each expert rater for them to complete their SAVRY ratings.
- Email invitation will be sent to NHS and social work staff detailing the training and information about the IRR study. Staff will be clearly informed that participation in the study is voluntary.
- Participants will attend a two-hour accredited SAVRY training workshop delivered by Dr McDonald or another appropriately qualified clinician. Information about the study will be provided again as well as the opportunity to ask questions. Willing participants will be provided with the consent form to sign.
- Participants will be randomly assigned one of two sets of case vignettes. Ordering of the vignettes will be counterbalanced to limit possible scoring bias in the order in which vignettes are presented.
- Participants will complete the staff information questionnaire and SAVRY ratings on their set of case vignettes.
- To enhance the clinical utility of the study, following the training, participants will be provided with feedback on the reliability of their risk assessments in comparison to the ‘gold standard’ ratings. If any participant demonstrates poor reliability, recommendations will be made on how to develop their assessment skills, for example seeking supervision.

Data Analysis.

- Hypotheses 1 and 2: Case 2 Intraclass Correlation Coefficients (ICC) and percentage agreement statistics (with the mean, mode and expert ratings) (Uebersax, 2015).
- Hypothesis 3: ICC and percentage agreement statistics to make comparisons across cases in terms of violence risk and gender. Gender and violence level will be analysed using ANOVA.
- Hypothesis 4: correlations between continuous professional characteristics and participant agreement on SAVRY ratings.

Justification of Sample Size.

Previous IRR studies of SPJ risk assessment tools using similar methodologies (Sutherland et al., 2012; Dickson, 2014) achieved participant numbers of 28 and 19 respectively. To

determine the sample size in this study the formula outlined in Walter, Eliasziw, and Donner (1998) was used. Based on power being set at .8, a null hypothesis of ICC.3, an alternative hypothesis of ICC .7, and a significance level of .05 a minimum of six vignettes and 22 raters are required.

Settings and Equipment.

The training workshop(s) will take place on NHS Greater Glasgow & Clyde (NHS GG&C) premises. Audiovisual equipment will be required for delivery of the training.

Health and Safety Issues

No specific health and safety issues are anticipated specific to this study.

Ethical Issues

Case vignettes will be based upon the clinical and theoretical experience of the research team. Where details are taken from real cases, information will be significantly altered to ensure anonymity. Participants will receive an information sheet and given the opportunity to discuss any questions. Written informed consent will be obtained at the training workshop.

Information in the vignettes could be potentially distressing. As participants are NHS clinicians working in mental health services and social workers they are likely to encounter similar material through their work therefore this is not anticipated to be a significant issue. The research team will be available at the end of training should participants have any concerns or questions they wish to discuss. Participant information will be anonymized and handled in line with the Data Protection Act (1998) and NHS Scotland procedures.

Financial Issues

The project requires funding for photocopying. Audiovisual equipment for delivery of the training can be sourced from CAMHS.

Limitations

It is acknowledged that the use of case vignettes is a potential limitation to the generalisability of the findings from this study. To ensure the authenticity of cases and minimise the extent to which risk judgements appear too obvious for each risk level, vignettes will be developed in collaboration with expert risk assessors as outlined above.

Research on the use of standardised vignettes will also be consulted, for example Gonsalvez et al. (2013).

Provisional Timetable

30 th Jan 2017	MRP Proposal Submission
15 th May 2017	Final approved MRP Proposal & Paperwork
June 2017	Ethics application to university ethics committee.
At some point from Jan – May 2018	Training workshop
May 2018 onwards	Data analysis
June 2018	‘Gold standard’ feedback to participants
July 2018	Thesis submission

Practical Applications

This research will add to the empirical literature on the psychometric properties of risk assessment tools using the SPJ approach by identifying the influences of rater and case characteristics on the IRR of the SAVRY. This will include exploration of gender bias in rater assessments of violence risk in females. Furthermore, the results may be used to inform staff training and how the SAVRY is used for risk assessment purposes within Scotland.

Providing SAVRY training to Tier 3 CAMHS clinicians is particularly timely as, following a recent redesign of Tier 4 CAMHS in NHS GG&C, there is an expectation of greater involvement of Tier 3 CAMHS with young people with forensic risk issues. Specifically, all Tier 4 cases will now be case managed at Tier 3. Thus, it is important for Tier 3 clinicians to be aware of the SAVRY and equipped to use and interpret its findings. The recent integration of health and social care also makes the inclusion of social work staff more pertinent as professionals will need to work more closely together to meet the need of this vulnerable client group.

References

- Borum, R., Bartel, P., & Forth, A. (2006). *Structured Assessment of Violence Risk in Youth (SAVRY)*. Odessa, FL: Psychological Assessment Resources.
- Borum, Lodewijks, Bartel, Forth (2010) *Structured Assessment of Violence Risk in Youth (SAVRY)*. In R.K. Otto & K.S. Douglas (Eds) *Handbook of Violence Risk Assessment* (pp63-79). New York: NY, US: Routledge/Taylor & Francis Group.
- Catchpole, R.E.H. & Gretton, H.M. (2003). The Predictive Validity of Risk Assessment with Violent Young Offenders: A 1 –year examination of criminal outcome. *Criminal Justice and Behaviour*, 30(6), pp 688-708.
- Centers for Disease Control and Prevention (2016) <https://www.cdc.gov/violenceprevention/youthviolence/> Date accessed 18.12.16.
- Childs, K.K., Frick, P.J., Katie Gottlieb, K. (2016). Sex Differences in the Measurement Invariance and Factors That Influence Structured Judgments of Risk Using the Structured Assessment of Violence Risk in Youth (SAVRY). *Youth Violence and Juvenile Justice*, 14(1), pp 76-92.
- Dickson, S. (2014). *The Psychopathy Checklist Youth Version (PCL:YV): an investigation into its inter-rater reliability*. Unpublished thesis.
- Dolan, M.C., & Rennie, C.E. (2008). The Structured Assessment of Violence Risk in Youth as a Predictor of Recidivism in a United Kingdom Cohort of Adolescent Offenders With Conduct Disorder. *Psychological Assessment*, 20(1), pp35-46.
- Gammelgård, M., Koivisto, A-M., Eronen, M., Kaltiala-Heino, R. (2015). Predictive validity of the structured assessment of violence risk in youth: A 4-year follow-up. *Criminal Behaviour and Mental Health*, 25, pp192-206.

- Gonsalvez, C. J., Bushnell, J., Blackman, R., Deane, F., Bliokas, V., Nicolson-Perry, K., Shires, A., Nasstasia, Y., Allan, C. (2013). Assessment of psychology competencies in field placements: standardized vignettes reduce rater bias. *Training and Education in Professional Psychology*, 7(2), 99-111.
- Hart, S. D., Michie, C., and Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the “margins of error” of group v. individual predictions of violence. *British Journal of Psychiatry*, 190(49), 60-65.
- Krug, E.G., Dahlberg, L.L., Mercy, J.A., Zwi, A.B., Lozano, R. (2002). *World report on violence and health*. Geneva, World Health Organization.
- Lodewijks, H.P.B., Doreleijers, T.A.H., de Ruiter, C., Borum, R. (2008). Predictive Validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during Residential Treatment. *International Journal of Law and Psychiatry*, 31, pp263-271.
- McEachran, A. (2001). The Predictive Validity of the PCL:YV and the SAVRY in a Population of Adolescent Offenders. Unpublished thesis.
- Meyers, J.R. & Schmidt, F. (2008). Predictive Validity of the Structured Assessment for Violence Risk in Youth (SAVRY) with Juvenile Offenders. *Criminal Justice and Behaviour*, 35(3), pp 344-355.
- Risk Management Authority (2016) Risk Assessment Tools Evaluation Directory (RATED) <http://rated.rmascotland.gov.uk> Date accessed 05.12.16.
- Skeem, J., Schubert, C., Stowman, S., Beeson, S., Mulvey, E., Gardner, W., Lidz, C. (2005). Gender and Risk Assessment Accuracy: Underestimating Women’s Violence Potential. *Law and Human Behavior*, 29(2), pp173-186.
- Sutherland, A.A., Johnstone, L., Davidson, K.M., Hart, S.D., Cooke, D.J., Kropp, P.R., Logan, C., Michie, C., Stocks, R. (2012). Sexual Violence Risk Assessment: An

Investigation of the Interrater Reliability of Professional Judgments Made Using the Risk for Sexual Violence Protocol. *International Journal of Forensic Mental Health*, 11, pp 119-133.

Uebersax, J. (2015). Statistical Methods for Rater and Diagnostic Agreement. Accessed January, 2017, from www.john-uebersax.com/stat/agree.

Viljoen, J.L., Scalora, M., Cuadra, L., Bader, S., Chávez, V., Ullman, D., Lawrence, L. (2008). Assessing Risk for Violence in Adolescents who have Sexually Offended. A Comparison of the J-SOAP-II, J-SORRAT-II, and SAVRY. *Criminal Justice and Behaviour*, 35(1), pp 5-23.

Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in medicine*, 17(1), 101-110.

WHO (2016). World Health Organisation Youth Violence Factsheet. <http://www.who.int/mediacentre/factsheets/fs356/en/> Date accessed 23.01.17.

Plain English Summary

An investigation of the Inter-rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY)

Violence is a significant problem worldwide due to the consequences caused by being a victim, offender, or witness to violence. The World Health Organisation describes it as a “global health problem” (WHO). Homicide is the fourth leading cause of death amongst young people aged 10-29 years old with an estimated 200 000 homicides occurring globally each year (WHO, 2016). For every death, many more young people suffer non-fatal injuries due to violence which can lead to lifelong physical, psychological and social consequences.

To limit these consequences, clinicians need to understand why young people are violent so that they can provide effective interventions. The Structured Assessment of Violence Risk in Youth (SAVRY; Borum et al., 2010) is a clinician guide for assessing violence risk in 13-18 year olds. It contains 24 factors that increase violence risk, as well as 6 protective factors that may reduce violence risk. Clinicians rate the presence of each factor for the young person they are assessing. To ensure the SAVRY is useful in assessing violence it is important to know how much clinicians agree in their risk ratings when rating the same person. This is known as inter-rater reliability (IRR).

Previous IRR studies of the SAVRY have used just two or three raters. In clinical practice, many clinicians with varying levels of experience complete risk assessments. Experience level may impact on the reliability of their ratings. Research has also shown that clinicians may have a gender bias and rate girls as being less violent than boys (Skeem et al., 2005).

This study will assess the IRR of the SAVRY by asking a group of clinicians to rate several case studies. Participants will be NHS clinicians and social workers who work with adolescents. It will use a larger group of raters than previous research and examine possible sources of bias that may impact on risk assessment decision-making, such as rater experience and gender bias.

Participants will be contacted via email with an invitation to attend SAVRY training

and information about the IRR study. The study will take place directly after the training but will be entirely voluntary. At the training, participants will again be provided with information about the study and asked if they wish to participate. Signed consent will be taken.

The study will develop understanding of factors affecting how violence risk is assessed using the SAVRY. This could inform future staff training. Results will be shared with relevant health and social care services.

References

Borum, Lodewijks, Bartel, Forth (2010) Structured Assessment of Violence Risk in Youth (SAVRY). In R.K. Otto & K.S. Douglas (Eds) Handbook of Violence Risk Assessment (pp63-79). New York: NY, US: Routledge/Taylor & Francis Group.

Skeem, J., Schubert, C., Stowman, S., Beeson, S., Mulvey, E., Gardner, W., Lidz, C. (2005). Gender and Risk Assessment Accuracy: Underestimating Women's Violence Potential. *Law and Human Behavior*, 29(2), pp173-186.

WHO (2016). World Health Organisation Youth Violence Factsheet. <http://www.who.int/mediacentre/factsheets/fs356/en/> Date accessed 23.01.17.

Word Count: 500

WEST OF SCOTLAND/ UNIVERSITY OF GLASGOW

DOCTORATE IN CLINICAL PSYCHOLOGY

HEALTH AND SAFETY FOR RESEARCHERS

1. Title of Project	An investigation of the Inter-rater Reliability of the Structured Assessment of Violence Risk in Youth (SAVRY)
2. Trainee	X
3. University Supervisor	X
4. Other Supervisor(s)	X
5. Local Lead Clinician	X
6. Participants: (age, group or sub-group, pre- or post-treatment, etc)	NHS clinicians, social workers
7. Procedures to be applied (e.g., questionnaire, interview, etc)	Questionnaire, rating case studies
8. Setting (where will procedures be carried out?) i) Details of all settings	NHS department or similar
ii) Are home visits involved	Y/N

WEST OF SCOTLAND/ UNIVERSITY OF GLASGOW

DOCTORATE IN CLINICAL PSYCHOLOGY

HEALTH AND SAFETY FOR RESEARCHERS

<p>9. Potential Risk Factors Considered (for researcher and participant safety):</p> <ul style="list-style-type: none">i) Participantsii) Proceduresiii) Settings	<ul style="list-style-type: none">i) Possible impact on emotional well-being due to possible distressing content in case vignettes, however this is deemed to be a minimal risk due to clinicians as they are likely to encounter similar material through their work.ii) None identifiediii) No additional factors in addition to standard health and safety issues in the workplace
<p>10. 10. Actions to minimise risk (refer to 9)</p> <ul style="list-style-type: none">i) Participantsii) Proceduresiii) Settings	<ul style="list-style-type: none">i) Advising participants that participation in the study is voluntary and they can withdraw at any time. Trainers will be available at the end of the study of participants wish to discuss anything with them.ii) N/Aiii) Informing participants of standard health and safety issues at the start of the training session e.g. location of fire exit

Trainee signature: Date:

University supervisor signature: Date:



RESEARCH EQUIPMENT, CONSUMABLES AND EXPENSES

TraineeX.....

Year of Course2ND Year..... Intake Year.....2015.....

Please refer to latest stationary costs list (available from student support team)

Item	Details and Amount Required	Cost or Specify if to Request to Borrow from Department
Stationery	Paper	Subtotal: £5
Postage	n/a	Subtotal:
Photocopying and Laser Printing	Training handouts, vignettes, SAVRY rating sheets	Subtotal: £50
Equipment and Software	Audiovisual equipment for training	To be borrowed from CAMHS

		Subtotal:
Measures	SAVRY Manuals	To be borrowed from CAMHS Subtotal:
Miscellaneous		Subtotal:
Total		£55

For any request over £200 please provide further justification for all items that contribute to a high total cost estimate. Please also provide justification if costing for an honorarium:

Trainee Signature..... Date.....

Supervisor's Signature Date