

Assessing Violence Risk: A Framework for Practice

Final Report

David P. Farrington
Darrick Jolliffe
Lorraine Johnstone

May 2008

Main Contact: Professor David P. Farrington
Institute of Criminology, Cambridge University
Sidgwick Avenue, Cambridge, CB3 9DT
Tel: 01223 335 360; Fax: 01223 335 356
email: dpf1@cam.ac.uk

Executive Summary

The main aim of this report is to assess the usefulness of risk assessment instruments for violence. We assess this in two ways, first by reviewing empirical evaluations of predictive accuracy and, second by surveying practitioner opinions about usefulness. The major instruments that we can compare are the OGRS and GSIR (static, statistical predictors for offender samples), the VRAG and PCL-R/PCL:SV (mainly static predictors focusing on personality disordered offenders), the HCR-20 and RAG-F (containing both static and dynamic items) and the LSI-R/LS/CMI (a mainly dynamic predictor focusing on criminogenic needs that have implications for treatment).

In this report, we try to bridge the gap between the evaluation standards of the scientific research community and those of practitioners and users of risk assessment instruments. This report is in two parts. The first presents a systematic review of research on the accuracy of risk assessment instruments in predicting violence. We conclude that a simple static instrument such as the OGRS and an instrument that incorporates both static and dynamic risk factors such as the HCR-20 seem most useful on the basis of currently available evidence. We recommend more analyses of the predictive validity of individual items, more efforts to develop a dynamic risk assessment instrument, and more evaluation research in Scotland.

The second part presents the results of a survey of practitioners to ascertain their opinions about different violence risk assessment instruments. They were asked to evaluate the instruments on standards such as the existence of a user manual, the need for user qualifications and competencies, accessible training, ease of use, administration time, ease of scoring, concordance of results with other information, and confidence in risk assessment. Most instruments met most standards, but the least satisfactory instruments were the RAG-F

and PCL-R. No instruments were viewed as providing satisfactory coverage of protective factors. Based on quantitative and qualitative information from practitioners, the most satisfactory instrument seems to be the HCR-20.

We conclude by making recommendations about a framework for assessing the risk of violence and what the RMA might do next to develop a screening instrument. Our main recommendations are as follows:

- The purpose and function of the instrument should be clearly stated.
- The instrument should have a user manual and a training programme.
- The instrument should include static and dynamic, risk and protective factors.
- The instrument should be used in an individualised way as part of a structured clinical judgement.

The instrument should be designed to identify those who do or do not require a more intensive assessment.

Finally, users should always bear in mind the difficulties involved in moving from predictions about individuals, and should be extremely cautious in drawing any conclusions about a person's risk of future violence.

PART 1. PREDICTIVE ACCURACY

Risk Assessment Instruments for Violence

The prediction of the future risk of violence is an essential task within numerous forensic settings (Hanson, 2005). In the past these predictions were largely based on clinical opinions, but research on the accuracy of these predictions (often referred to as first-generation risk predictions) demonstrated that these opinions were not much better than chance in predicting future violence (Quinsey et al., 1998). In direct reaction to the poor predictive validity of clinical opinions, researchers set about developing devices that would improve the prediction of future violence (e.g. Andrews & Bonta, 1995; Quinsey et al., 1998). While the overall validity of these empirically driven risk assessment devices is often disputed (e.g. Hart, Mitchie & Cooke, 2007), the tone of the research literature has changed from questioning whether future violence can be predicted to debating the best methods of developing risk assessment instruments (Hanson, 2005).

Table 1 shows a number of the instruments designed by various methods to predict the future risk of violence. The actuarial risk assessment instruments are statistically derived measures that are based on determining the presence or absence of static characteristics that predict recidivism (e.g. age at first offence, gender). For example, the OGRS (Offender Group Reconviction Scale; Copas & Marshall, 1998) is a criminogenic risk assessment instrument designed in England that is based completely on unchangeable demographic features and previous offending history. The OGRS estimates the probability that an offender will be reconvicted within two years after release on the basis of six variables (e.g. age at first offence, gender and type of offence). The OGRS score cannot be calculated for people who do not have a previous conviction, and it does not include any weighting of mental health variables.

Table 1. Key Features of the Main Risk Assessment Instruments

Instrument, Country	Citation	No. Items	Scoring	Dynamic/Static	Interview with Individual	Validation Sample
Violence Risk Appraisal Guide (VRAG), Canada.	Harris, Rice and Quinsey, 1993	12	Scores range from -26 to +38 with higher scores reflecting greater likelihood of violence	Static	Not necessary	800 Violent men, approximately half from a secure psychiatric facility and half from federal prisons.
Level of Service Inventory - Revised (LSI-R), Canada	Andrews & Bonta, 1995	54	Items scored on a binary scale. Scores range from 0–54	Dynamic	Desirable, but not necessary	Originally designed to help probation officers in planning their supervision of probationers and parolees.
Psychopathy Checklist – Revised (PCL-R), Canada	Hare, 1991	20	Items scored on a three-point scale (0 – absent, 1 – possible present, 2 present) Scores range from 0 – 40	Static	Desirable, but not necessary	Designed using a correctional sample to assess the personality disorder of psychopathy.
Psychopathy Checklist Screening Version (PCL: SV), Canada	Hart, Cox & Hare, 1995	12	Items scored on a three-point scale (0 – absent, 1 – possible present, 2 present) Scores range from 0 – 24	Static	Desirable, but not necessary	Designed using a correctional sample to assess the personality disorder of psychopathy.
Historical, Clinical and Risk Management Scales (HCR-20), Canada	Webster, Douglas, Eaves & Hart, 1997	20	10 Historical items, 5 Clinical and 5 Risk Management. Items scored on a three-point scale (0 – absent, 1 – possible present, 2 present) Scores range from 0 to 40.	Dynamic	Desirable, but not necessary	Developed from a thorough consideration of the empirical literature concerning factors that relate to violence.
Offender Group Reconviction Scale (OGRS), England	Copas & Marshall, 1998	6	Scoring based on the presence or absence of six items. Scores range from 0 – 100.	Static	Not necessary	Developed using four samples of 3000 – 4000 of those given probation or community service orders.
Violence Risk Scale (VRS), Canada	Wong & Gordon, 2006	26	Items are scored on a four-point scale (0, 1, 2, 3) Scores range from 0 – 78	Dynamic	Necessary	Developed on the theoretical basis of the psychology of criminal conduct and the risk, need responsivity principles using a high-risk correctional sample.

Table 1 (continued)

Instrument, Country	Citation	No. Items	Scoring	Dynamic/ Person Static	Interview with Individual	Validation Sample
Self-Assessment Questionnaire (SAQ), Canada	Loza & Loza-Fanous, 2000	72	A self-reported True/False Questionnaire. Scores range from 0 – 36	Dynamic	Not necessary	Theoretically and empirically based on the recidivism literature. Based on correctional sample.
General Statistical Information on Recidivism (GSIR), Canada	Nuffield, 1982	15	Scoring based on the presence or absence of 15 items	Static	Not necessary	Developed using 2500 federally sentenced offenders.
Risk Matrix 2000 Violence (RM2000V), England	Thornton, 2007	3	Scoring based on age (in four categories), number of sentences for non-sexual violence and conviction for burglary	Static	Not necessary	Developed to assess the non-sexual violence of sexual offenders. Based on structured anchored clinical judgment
OASys Violence Predictor (OVP), England	Howard, 2007	13	Scoring is based on five static (mostly offending) factors and eight dynamic factors. Scores range from 0 – 100.	Dynamic	Necessary	Developed using 26619 offenders for whom the OASys structural clinical risk/needs assessment and management tool had been administered

Note: The coding of static or dynamic refers to the most important feature of an instrument. For example, the HCR-20 has 10 static items and 10 dynamic items, but its dynamic nature is its most noteworthy feature.

Similarly, the General Statistical Information on Recidivism (GSIR; Nuffield, 1982) is an actuarial instrument developed on 2500 Canadian offenders based on 15 items including current and past offending and age at first offence. The Risk Matrix 2000V is another example of an actuarial device designed to assess the risk of future violence of sex offenders. The RM2000V includes only three items: age when next at risk (i.e. age on release), number of previous violent convictions and any previous burglaries (Thornton, 2007).

The Violence Risk Appraisal Guide or VRAG (Harris, Rice & Quinsey, 1993) is also a purely actuarial device. The VRAG is a twelve-item scale that was designed to predict violent recidivism among mentally disordered offenders. Prediction results with the VRAG have shown that the instrument predicts violence at follow-up periods of 3.5, 6 and 10 years regardless of the base rate for failure. Furthermore, the reliability and the validity of the VRAG has been extended to non-psychiatric incarcerated males (Loza & Dhaliwal, 1997).

Embedded within the VRAG is another device that, while not specifically designed to predict the risk of future violence, has shown significant predictive validity (Gendreau, Goggin & Smith, 2002). The Psychopathy Checklist – Revised (PCL-R) is a twenty-item scale that was designed to assess the construct of psychopathy. The prototypical psychopath possesses significant interpersonal, affective and behavioural abnormalities that make functioning in a prosocial manner difficult. The PCL-R has been shown to have utility in predicting future criminal behaviour (Hemphill, Hare & Wong, 1998), and it also predicts violent behaviour (Serin, 1996). The Psychopathy Checklist – Screening Version or PCL-SV (Hart, Cox & Hare, 1995) is shorter than the PCL-R (12 items as opposed to 20), but it has shown similar levels of predictive validity for general and violent reoffending (Edens et al., 2006; Nicholls et al., 2004).

The Historical Clinical Risk Management – 20 instrument (Webster et al., 1997) is a 20 – item broad based violence risk assessment device that is applicable to a number of settings. The HCR-20 is unique amongst risk assessment devices in that it includes an assessment of both current functioning (with a 5-item clinical subscale) and future functioning (with a 5-item risk management scale). This is in addition to the 10 historical (static) features that are assessed in most other risk assessment devices. There is some evidence for the reliability and validity of the HCR-20. For example, over a follow-up period of two years, the historical and risk management items predict violent crime (Douglas, 1996).

The Level of Service Inventory – Revised (LSI-R; Andrews & Bonta, 1995) is a criminological risk and need for treatment measure that was developed within a social learning perspective. The focus of this instrument is on personal history and interactions with others in a social context. It is suggested that the four areas of antisocial cognitions, antisocial associates, history of antisocial behaviour and antisocial personality influence criminal behaviour (Andrews & Bonta, 2006). This instrument was developed using probationers and offenders with sentences of less than two years to aid in determining supervision requirements. There is research to suggest that this device reliably predicts future violent and nonviolent convictions (Kroner & Mills, 2001). The LS/CMI is a further development of this instrument (Andrews, Bonta & Wormith, 2004).

The self-assessment questionnaire (SAQ) is a self-report questionnaire that consists of 72 true or false items (Loza & Loza-Fanous, 2000). This instrument consists of eight subscales, six of which (criminal tendencies, antisocial personality problems, conduct problems, criminal history, alcohol/drug abuse, antisocial associates) are used to predict both violent and non-violent recidivism. There is evidence to support the concurrent and construct validity of the SAQ (Loza & Loza-Fanous, 2000). There is also evidence to suggest that

scores on the SAQ significantly predict future violence after release (Kroner & Loza, 2001).

Another more recently developed risk assessment device is the Violence Risk Scale (VRS - Wong & Gordon, 2001). Unlike most other risk assessment devices this 26 – item scale gives significantly more weight to dynamic (changeable) factors (20 items) than to the static (historical) factors (6 items) that comprise most of the previously mentioned risk assessment instruments. The test designers suggest that the benefit of including dynamic items is that the VRS identifies targets for treatment and that successfully addressing these factors will reduce an individual's VRS score and their risk of violence. There is evidence to suggest that scores on the VRS correlate with violence two years later ($r = .46$; Wong & Gordon, 2001), but thus far less research has been completed on this measure compared to the other instruments listed above.

The Offender Assessment System (OASys) is the principal risk assessment and management tool used by the National Offender Management Service in England and Wales. Based on logistic regression and the information available in OASys, the OASys violence predictor (OVP) was created (Howard, 2007). This risk assessment device combines five static measures (e.g. age at first conviction, gender) with eight dynamic measures (e.g. recognizes impact of offending, alcohol misuse) to produce a score out of 100 that significantly predicts the likelihood of future violence two years after release (Howard, 2007).

Comparing Risk Assessment Instruments

All of the above instruments have advantages and limitations. For example, the VRAG and PCL-R can both be completed with only documentary evidence and as such do not require access to, or an interview with, the individual (Gendreau, Goggin & Smith, 2002; Harris, Rice & Quinsey, 1993). This might be a benefit when an individual is reluctant to take part in the assessment process. However, a limitation of both of these devices is that they

are based solely on historical factors, and as such cannot measure individual change. Individuals will remain at a similar level of risk regardless of their engagement in or response to treatment.

A key issue is, in light of the benefits and limitations of these risk assessment instrument, which device is 'best' for predicting an individuals' future risk of violence? There are two methods that allow this question to be adequately addressed. The first is to administer a number of risk assessment instruments to the same population and then follow this group up after release to assess reoffending. This methodology was employed by Kroner and Mills (2001) who administered the VRAG, PCL-R, HCR-20 and LSI-R to a group of 97 federally sentenced males in a Canadian prison. The results provided some evidence that the LSI-R and VRAG were somewhat more accurate in predicting reconviction at two years after release, and that the LSI-R and the HCR-20 were somewhat more accurate in predicting violence. This methodology was also used by Cooke, Mitchie and Ryan (2001) in a study in which the predictive validity of the PCL-R, VRAG and HCR-20 was assessed using interviews and the files of 190 prisoners held in H.M. Prison, Barlinnie. The results suggested that all three devices had similar levels of predictive validity in predicting reconviction and violent reconviction two years after release.

The other method by which the predictive validity of the various measures can be compared is to carry out a systematic review and meta-analysis. Systematic reviews use rigorous methods for locating, appraising and synthesizing evidence from prior studies. They have explicit objectives, explicit criteria for including or excluding studies, and they are reported with the same level of detail that characterizes high quality reports of original research (e.g. Farrington & Petrosino, 2001; Farrington & Welsh, 2002). Meta-analysis is a form of analysis based on research reports rather than individuals. An effect size measure is

derived in each study and these effect sizes are summarized and related to methodological features of the studies (Lipsey & Wilson, 2001). The summary effect size addresses the question: How well does it work?

There have been a limited number of prior meta-analytic reviews of the effectiveness of various risk assessment devices for predicting future offending. Salekin, Rogers and Sewell (1996) undertook a meta-analysis of 18 studies examining the relationship between the Psychopathy Checklist (PCL) and Psychopathy Checklist Revised (PCL-R) and violent and general recidivism. They found that the PCL and the PCL-R appeared to be good predictors of violent behaviour, with an effect size of $d = .79$. However, interpretation of these results is limited by the fact that not all of the studies included were predictive (i.e. in some studies the PCL was scored with knowledge of later offending outcomes). Furthermore, the 'violent behaviour' examined in this study included a wide range of behaviour including institutional violence, community violence and aggression.

Gendreau, Goggin and Smith (2002) assessed the predictive validity of two risk assessment devices (PCL-R and LSI-R) using meta-analysis. They identified 57 prospective studies which examined the predictive validity of these devices, acknowledging that a number of these studies had been included in previous meta-analyses (Gendreau, Goggin & Smith, 1999; Hemphill, Hare & Wong, 1998; Salekin, Rogers & Sewell, 1996). The researchers used phi correlations to summarise the results across the different studies. Overall the results suggested that the LSI-R performed better than the PCL-R in predicting both general and violent recidivism, but that this difference was only significant for general (and not violent) recidivism.

There were a number of shortcomings of this meta-analysis which limit its usefulness for informing risk assessment. First, a detailed analysis of the mediators and moderators of

predictive accuracy were not reported. For example, the PCL-R might have performed better than the LSI-R in predicting future violence among violent offenders, but offenders were treated as a homogeneous group in all analyses. Second, only two risk assessment devices were assessed. As previously mentioned, there are currently many risk assessment devices, some of which might perform as well as or better than the PCL-R and LSI-R.

The third shortcoming involves the method by which the effect size was derived from each of the 57 studies; using the correlation between the risk prediction measure and recidivism. While this is not incorrect, research has clearly indicated that this is not the most sensitive measure to summarize the accuracy of risk assessment measures. Currently, researchers recommend the use of receiver operating characteristic (ROC) curves because these are less dependent on the base rate of the criterion variable in the sample (in this case general or violent recidivism) (Webster et al, 1997). Most recent studies of risk assessment devices report information about the ROC curve as an index of the accuracy of prediction.

The recently-published meta-analysis of risk assessment instruments by Campbell, French and Gendreau (2008) also uses the phi correlation but attempts to correct it for base rate problems. In predicting violent recidivism, they concluded that the best instrument was the VRAG (weighted mean $r = .32$), followed by the LSI-R ($r = .28$), the PCL-R ($r = .27$), the HCR-20 ($r = .22$) and the GSIR ($r = .22$). However, despite significant heterogeneity in effect size measures, they do not appear to have used random effects models in their meta-analysis.

It is now widely accepted (see e.g. Mossman, 1994) that the best measure of predictive efficiency in an $N \times 2$ table relating risk scores to an outcome such as violence is the Area under the ROC (Receiver Operating Characteristic) curve (AUC; see Swets, 1986). The ROC curve plots the probability of a “hit” or true positive (the fraction of convicted

persons identified at each cut-off point) against the probability of a false positive (e.g. the fraction of unconvicted persons identified at each cut-off point). AUC, like the odds ratio (but unlike phi), is a measure of predictive efficiency that is unaffected by changes in sample size or row and column totals (e.g. prevalence). It is equivalent to the proportion of correct predictions in a 2 x 2 table where chance expectation = .50 and perfect discrimination = 1.0 (Fergusson *et al.*, 1977).

If an AUC is greater than $.50 + 1.96 \times SE$, it is statistically significant at $p < .05$. The 95% confidence interval (CI) of the AUC is $\pm 1.96 \times SE$. If the lower confidence interval is less than .50, the AUC is not significant.

Method

This investigation involved a comprehensive systematic review and meta-analysis of empirical studies which evaluated the effectiveness of various risk assessment instruments for predicting future violence among adult males. Meta-analysis based on the AUC and its SE was used to quantify the results of the systematic review (see Kester & Buntinx, 2000). An effect size measure was derived in each study that was included in the systematic review and these effect sizes were summarized to provide an evaluation of which risk assessment instrument best predicted future violence.

The AUC and its standard error (SE) were estimated assuming underlying normal distributions and using maximum likelihood fitting techniques developed by Metz *et al.* (1998). The advantage of using this method is that it produces an AUC and SE from the N x 2 table presented in research reports. Other methods of calculating the AUC and its SE (e.g. using SPSS) require the raw data, which was not available.

Objectives of the Study

This systematic review had the following objectives:

1. To characterise (and as far as possible quantify) the evidence to date on the predictive validity of the most common risk assessment measures that are used to predict serious violent offending in Scotland, England and North America.
2. To characterise (and as far as possible quantify) the potential mediators and moderators of the relationships identified in 1 above. For example, the predictive validity might be influenced by the instrument employed (PCL-R, LSI – R, VRAG, VRS), geographical location (Scotland, England, North America), type of sample (correctional or psychiatric), or type of offender (violent, general).
3. In light of the findings of the review, to make recommendations about what future research is needed to advance knowledge about the predictive validity of risk assessment instruments for violence.

Inclusion Criteria

Below is a list of the criteria that were used for including a study in the systematic review and meta-analysis.

1. The study prospectively administered a risk assessment instrument to an offender who was subsequently released and therefore at risk of committing additional offences. Ideally, this would involve administration of the instrument before release, but studies which involved retrospective completion of the instrument based on case files were eligible for inclusion if the completion was blind to previously existing reconviction status.
2. The study reported information about the area under the curve (AUC) for the receiver operating characteristics (ROC) for the risk assessment device predicting violence. Where this was not reported the study is eligible for inclusion if this statistic could be calculated based on information available in the study (e.g. an N x 2 table) or if the information can be obtained from the author of the study.
3. The study administered the risk assessment instrument to at least 50 persons initially (before attrition). Smaller studies are likely to have low internal and external validity and insufficient statistical power.
4. The study was published between 1990 and the present day. We doubt that risk assessment studies published before 1990 used the AUC.
5. The study focused on a sample of males. If the sample included both males and females it was included if the results could be separated by gender, or if greater than half of the sample was male.
6. This review was limited to instruments designed for use with adult violent offenders, and administered to violent or general offenders. This is because there are a number of thorough systematic reviews and meta-analyses on the predictive ability of risk assessment devices when applied to sex offenders (e.g. Hanson & Morton-Bourgon, 2007) and young offenders (Edens, Campbell & Weir, 2007; Schwalbe, 2007).

Key Terms Searched

Below is a list of some of the key terms, searched in any part of a document, which formed the basis of our search for potentially relevant articles: Violen*, aggressiv*, serious* AND offend*, crim*, AND risk*, assessment, HCR-20, VRAG etc.

Search Strategy and Results of Searches

The search for relevant studies was based on a number of strategies. These included: (1) contact with leading researchers in the area, including all members of the Risk Group of the International Association of Forensic Mental Health Services; (2) electronic database searches (e.g. Criminal Justice Abstracts, Psychlit etc.); (3) focused internet searches; and (4) citation searches. This led to the identification of 127 potentially relevant studies of which 31 were judged to meet our inclusion criteria. Excluded studies published in 1990 or later are listed in the Technical Appendix, together with reasons for exclusion.

Included Studies

Fourteen of the 31 studies included in the review used multiple risk assessment instruments. This resulted in 70 independent evaluations of the relationship between a risk assessment instrument and later violence that could be measured by the AUC. Below are narrative descriptions of the 31 studies that were included in our review. Unfortunately, these descriptions are sometimes limited by the information available in the original reports.

Barnoski (2003) evaluated the predictive validity of the LSI-R using 22533 offenders released from prison in Washington State. The LSI-R was administered prospectively to all felony offenders by correctional staff as a method of assessing the risk of reoffending as part of the Offender Accountability Act (OAA) legislation enacted in 1999. Therefore, of the sample of offenders some would have been released from custody and others on a community sentence (e.g. probation). In all cases the LSI-R had been administered within 90

days of the date of placement in the community. The violent reconvictions of the 22533 offenders were searched 24 months after release, and the results suggested that the LSI-R significantly predicted future violence (AUC = .64, se = .01).

Bonta and Yessine (2005) reported on the accuracy of a number of risk assessment instruments for predicting violent and general reoffending among a high-risk sample of Canadian offenders on the National Flagging System. The sample comprised 256 male offenders (average age 35.2) from six of the eleven provinces and territories of Canada. Information to complete the risk assessment devices was collected from data available in the Offender Management System, and the instruments were completed retrospectively, but blind to the outcome. The risk assessment devices were the GSIR, the VRAG and the LSI-SV (similar to the LSI-R). The 202 men who were released were followed up for official convictions of violence at an average of 41 months after release. The results suggested that, based on the AUC all measures predicted violence at significantly better than chance levels. The AUC of the GSIR was .77 (se = .04), the AUC of the VRAG was .68 (se = .08) and the AUC of the LSI-SV was .66 (se = .04).

Coid et al. (2008) conducted a comprehensive evaluation of the efficacy of structured risk assessment devices for predicting general and violent reoffending in England and Wales. A total of 1396 offenders were administered five risk assessment devices. These were the PCL-R, the VRAG, the HCR-20, the RM2000V and the OGRS (Copas & Marshall, 1998). Prisoners were eligible for the study if they were serving two years or more in an English or Welsh prison (not serving life) for a violent or sexual offence, were at least aged 18 and had at least one year left of their sentence left to serve. Previous criminal history and OGRS were used to select a sample of high-risk offenders and a stratified sample of the remainder. An average of 2 years later the reconvictions of the original sample were searched using the

police national computer. Of the 1353 participants for which reconviction information was available, the results suggested that the OGRS was the most accurate in predicting future violence and general offending (AUC = .718 and .762 respectively). The VRAG was the next most accurate (AUC = .700 violence and .721 general), followed by the RM2000V (AUC = .685 violence and .711 general). The PCL-R was the least accurate (AUC = .641 violence and .649 general) followed by the HCR-20 (AUC = .667 violence and .663 general). All of these AUC values were significantly greater than the chance expectation of .50.

Cooke, Mitchie and Ryan (2001) undertook a study that examined the predictive validity of three risk assessment instruments (VRAG, PCL-R and HCR-20) for predicting future violence. The risk assessment measures were administered prospectively by trained psychologists using semi-structured interviews and comprehensive file reviews. The sample comprised 190 male inmates released from HMP Barlinnie. Twenty-four months after the offenders were released their violent convictions were collected by the research team. The results suggested that all three measures performed reasonably well (VRAG AUC = .67; PCL-R AUC = .65; HCR-20 AUC = .69) in predicting future violence, but that no measure outperformed any other.

Dahle (2006) used three risk assessment devices (LSI-R, HCR-20, and PCL-R) which were completed retrospectively by a criminologically trained researcher for a group of 397 adult male (mean age 29.8) offenders sentenced to prison between February and May 1976 in the former West Berlin. Of the original sample 309 males were available for follow-up ten years later. The results suggested that the HCR-20 had significantly greater predictive validity (AUC = .71, $se = .03$) than the LSI-R (AUC = .65, $se = .03$) for predicting future violence. Information about the predictive validity of the PCL-R using the AUC was not reported.

de Vogel et al. (2004) conducted a retrospective study that examined the predictive validity

of the HCR-20 and PCL-R coded on the basis of file information for 120 patients (89% men, mean age = 25.5) discharged from a Dutch forensic psychiatric hospital. The risk assessment instruments were completed by researchers using available file information. A small proportion of the sample was identified as having Axis I disorders (16%) but almost three-quarters were identified as having an Axis II disorder. Recidivism was assessed using official records when the patients had been released for an average of 72.5 months. The results suggested that both the HCR-20 (AUC=.82, se = .04) and the PCL-R (AUC = .75, se = .05) significantly predicted future violence.

Dolan and Khawaja (2004) used 19 of the 20 items of the HCR-20 to predict the likelihood of future convictions and self or collateral reports of violence among 70 violent male patients released from a medium security hospital in England. Hospital files were used to code data to score the HCR-20 and the Offenders Index and community case files were used to obtain data about offending. The subjects were hospitalized for an average of two years and were followed up into the community for an average of 59 months. The HCR-20 predicted the combined measure of official and self or collateral violence at a level significantly better than chance (AUC .76, se= .05).

Douglas et al. (1999) evaluated the predictive validity of the HCR-20 and the PCL: SV for predicting the violence of 193 civilly committed psychiatric patients. The information that was required to complete the risk assessment instruments was collected retrospectively using available file information, but the raters were blind to the outcomes. It was not possible to code all aspects of the R scale of the HCR-20 because this study was based on file data, so the scores on the HCR-20 were pro-rated. More than half of the sample was male (61%) and the average age of the sample was 39. The results suggested that both the HCR-20 and the PCL-SV significantly predicted future violence among this sample. For the HCR-20 the AUC

was .80 (se = .06) and for the PCL: SV the AUC was .79 (se = .05).

Douglas et al. (2003) reported on a study in which the HCR-20 was completed retrospectively (using file information) for a group of 100 forensic psychiatric patients who had been found guilty by reason of a mental disorder and subsequently released to the community in Canada. Over 90% of the sample was male and a large proportion (73%) had a primary diagnosis of schizophrenia. Approximately 43 months after release official records were searched for evidence of violence. Nurse and practitioner files were also used to collect data about self-reported and collateral evidence of violence. The HCR-20 predicted future violence among this sample (AUC = .67, se = .10), but not significantly.

Doyle and Dolan (2006) reported on the results of a prospective study undertaken in the UK that was modelled on the MacArthur Violence Risk Assessment Study. In this study 129 patients (67% male, mean age=40) were recruited from five psychiatric institutions in the North West of England (three forensic medium secure, two non-forensic), and semi-structured interviews, files and collateral information were used by forensic practitioners to complete the HCR-20, PCL: SV and the VRAG. Approximately 70% of the sample had a serious mental illness diagnosis of either schizophrenia-spectrum disorders or bipolar disorder. Of the 129 participants who were discharged, complete follow-up data (at 5.5 months) on violent recidivism was available for 112. In this study violent recidivism included a combination of self-report, collateral information and official violence. While all three instruments predicted violence at better than chance levels, the HCR-20 (AUC= .797, se=.05) had somewhat greater predictive ability than either the PCL: SV (AUC=.69, se= .06) or the VRAG (AUC=.66, se=.06).

Edens, Skeem and Douglas (2006) used data collected as part of the MacArthur Violence Risk Assessment Study to test the predictive validity of the PCL: SV and the VRAG. Over

1000 civil psychiatric patients, from three hospitals in the US, were interviewed by forensic practitioners, and this information was used by the researchers to complete the VRAG and the PCL: SV. The sample had an average age of 30, 59% were males and most (58%) were voluntary patients. About 70% of the sample was reported to have Axis I disorders. Five months after release patients, collateral informants and official records were interrogated for evidence of violence. The AUC of the VRAG was .73 (se = .02) and the AUC of the PCL: SV was slightly higher at .78 (se = .02).

Girard and Wormith (2004) tested the predictive validity of the LSI-Ontario Revision in Ontario (Canada), using a large sample of offenders aged about 32 which included male inmates sentenced to less than two years (n=454) and male probationers (n=176). The LSI-OR has since been adapted into the LS/CMI, but the General Risk/Needs section of this instrument is equivalent to the LSI-R (Girard & Wormith, 2004, pp. 153). For the purposes of comparability across studies only the results relating to the LSI-R are presented. The instrument was completed prospectively using semi-structured interviews and collateral file information by trained institutional classification officers or community probation officers in five separate locations. After an average of approximately 30 months information about violent reconvictions was obtained from the official records. The AUC for violent recidivism was .68 (se = .06), suggesting that this scale significantly predicted future violence.

Grann et al. (1999) used forensic records to rate a sample of Swedish offenders using the PCL-R. Specifically, the sample comprised all individuals subjected to a court-ordered forensic psychiatric evaluation and convicted for a violent crime in Sweden between 1988 and 1990 and clinically diagnosed with a personality disorder (with or without concomitant alcohol/drug abuse or dependence). Approximately 9% of the sample was female, and the offenders were predominantly from a working class background (85%). Three independent

raters, blind to the offending outcomes of the individuals, coded the PCL-R scores based on files and records. The inter-rater reliability for 38 randomly chosen cases was .78. Of the 352 individuals eligible for follow-up, 121 (34%) committed a least one new violent crime that led to a court conviction. The Area under the Curve (AUC) within two years of release was .72 (95% CI: .66 - .78). There was evidence that this AUC for violent reconviction did not vary substantially depending on the time of follow-up; from .67 (95% CI .58 - .76) at 6 months to .70 (95% CI .63 - .76) at 5 years.

Grann et al. (2000) explored the predictive validity of the VRAG and the historical part of the HCR-20. Only the results of the VRAG are included in our review. The VRAG was completed by the researchers retrospectively for two populations: one group of 358 violent offenders (mean age =32; 90% male) with personality disorders and one group of 202 violent male offenders (mean age =33) with schizophrenia. All were offenders who had been subjected to a court-ordered psychiatric evaluation between 1988 and 1993. Two years after release 293 personality disordered offenders and 111 offenders who had been diagnosed with schizophrenia were followed up for official convictions. The results suggested that the VRAG predicted future violence significantly better than chance for personality disordered offenders (AUC = .68; 95% CI .62 - .73), but not for schizophrenic offenders (AUC = .60; 95% CI .50 - .69).

Gray et al. (2004) retrospectively assessed the case files of 346 patients released from a medium-secure psychiatric facility in England in order to test the predictive validity of the PCL: SV, the HCR-20 and the OGRS. A large proportion of the sample was male (89%) and white (84%), with about half having a primary diagnosis of schizophrenia or another psychotic disorder. Of the original sample 315 were released and followed up 72 months after release. The results suggested that the OGRS (AUC = .71, se = .06) performed better in predicting

future violence compared to the HCR-20 (AUC = .56, se = .06) and the PCL-R (AUC=.58, se = .06) neither of which was significant.

Harris et al. (2002) undertook a retrospective study to test the predictive validity of the VRAG for predicting the likelihood of violence among 467 patients in forensic hospitals across Ontario, Canada. This sample was predominantly male, and a large proportion were diagnosed with a psychotic illness. Only 133 of the original sample of 467 forensic patients were 'at risk' for 5 years after release, and it was this group who had their violent reconvictions searched. The VRAG had significant predictive validity for predicting future violence (AUC = .80, se = .04).

Hildebrand et al. (2005) undertook a retrospective study of 156 violent forensic psychiatric patients in the Netherlands who were involuntarily admitted between September 1973 and July 1999 and released between January 1992 and December 2001. Available file information was coded by the researchers and used to complete the PCL-R and the HCR-20. The average length of detention at the hospital was 86 months and the average follow-up period after release was also 86 months. Using the AUC, the predictive validity of the PCL-R (AUC = .72, se = .05) and the HCR-20 (AUC=.67, se = .05) were significantly greater than chance.

Howard (2007) conducted a large scale study in England and Wales using the Offender Assessment System (OASys), which is the National Offender Management Service's principal risk assessment tool. This tool identifies static and dynamic risk factors and generates a summary score in order to assess the likelihood of reconviction and risk of serious harm and to prepare a Sentence Plan. Over 26000 offenders who were assessed at the start of their community supervision comprised the sample. Most of the sample was male (86%) and white (93%), and most had not been sentenced to custody (84%). The index

offence for 26% was violent (violence against the person, sex, robbery). The purpose of this study was to develop unique risk assessment measures based on the OASys system called the OASys General Re-offending predictor and the OASys Violence Predictor. The breadth and depth of information collected as part of OASys meant that a number of other risk assessment instruments could be completed on these offenders by proxy. These included the HCR-20 (18 out of 20 items scored), the PCL: SV (10 out of 12 items scored) and the VRAG (52 out of 64 points scored). Information to complete the Risk Matrix 2000- violence and the OGRS scale was also available¹. Police National Computer data was used to search for violent convictions at 24 months after the offenders had been released. The results suggested that all measures significantly predicted future violence, with only slight differences between them. The OASys Violence predictor had an AUC = .71, the VRAG and RM2000V had an AUC = .67, the HCR-20 and the OGRS had an AUC of .66, and the PCL: SV had an AUC of .64.

Kroner and Loza (2001) administered four risk assessment instruments (SAQ, PCL-R, GSIR and VRAG) to a group of 78 incarcerated men from five federal prisons in Canada. The mean age of the men was 28.5 and a large proportion of the sample was identified as white (87%). After an average follow-up period of approximately 20 months, information about violent convictions was collected from official sources. The results suggested that, based on the AUC, the SAQ and GSIR had the highest predictive validity (AUC = .75 and .74 respectively) for predicting future violence, followed by the PCL-R (AUC = .70) and the VRAG (AUC = .60).

Kroner and Mills (2001) prospectively administered a number of risk assessment instruments (PCL-R, LSI-R, VRAG and HCR-20) to a group of 97 predominantly violent

¹ Information about OGRS 2, OGRS 3 (under development) and OGRS SV (sexual and violence) was available. For comparability across studies only the OGRS 2 results are presented here.

federal male inmates (mean age 27) to assess the ability of these devices to predict future violent offending. After approximately two years 87 of the original sample was eligible to be followed up for official convictions for violence. The results suggested that the LSI-R significantly predicted future convictions for violence (AUC = .67, se = .08). The HCR-20 had an AUC of .62 (se = .08), the VRAG had an AUC of .60 (se = .09) and the PCL-R had an AUC of .56 (se = .08). None of these instruments significantly predicted future violence.

Loza and Green (2003) tested the effectiveness of five risk assessment instruments (VRAG, PCL-R, LSI-R, GSIR and SAQ) in predicting future violence in a group of 91 adult male offenders (mean age =30) released from federal prison in Canada. The average time served by this group of offenders was 7.8 years. All devices were administered prospectively using interviews and available file information. Five years after release the official records were searched. The SAQ (AUC = .71) and the GSIR (AUC = .74) were somewhat better in predicting future violence than the VRAG (AUC = .63), the PCL-R (AUC = .67) and the LSI – R (AUC = .67).

Loza et al. (2002) undertook a prospective study intended to examine the validity of the VRAG in predicting violence among a group of 124 predominantly violent adult males (mean age=30.1) released from Canadian federal prisons. The VRAG was completed by a number of forensic practitioners using both file information and interviews. The participants had been incarcerated for approximately 80 months and their convictions were searched two years after release. A table presenting the level of risk versus the number of violent convictions was used to calculate the AUC. The results suggested that the VRAG did not significantly predict future violence among this sample (AUC = .52, se = .06).

Mills and Kroner (2006) reported the predictive validity of the VRAG, PCL-R, LSI-R and GSIR. In this study these risk assessment instruments were completed prospectively using

files and interviews, by MA or PhD level clinicians, for a group of 209 incarcerated adult males (mean age = 29.9) sentenced to two years or more in prison. The purpose of the study was to examine the effect of disagreement between risk assessment devices on their predictive accuracy, but for the purposes of the present review only the predictive accuracy was examined. Approximately 14 months after release, official records of violence were searched. The results suggested that the VRAG (AUC=.67, se=.06), the LSI-R (AUC=.68, se=.06) and the GSIR (AUC=.67, se=.05) predicted violence at better than chance levels, but the PCL-R did not (AUC = .62, se = .06).

Nicholls et al. (2004) evaluated the predictive validity of violence risk assessments completed using the HCR-20 and the PCL:SV in a sample of 268 involuntarily hospitalized males and females. Because the information was presented separately for males and females, only the results with the 147 males (mean age = 36) are presented here. Almost all of the sample (99%) were reported to have Axis I disorders and less than half (39%) were reported to have Axis II disorders. Researchers retrospectively coded the HCR-20 and the PCL:SV. The average time of hospitalization was 108 days and those released were followed up 23 months later. Follow-up information for the PCL:SV was based on 146 men and for the HCR-20 it was based on 117 men. The AUC predicting future violence (measured using information available from both medical and correctional records) was .70 (se =.07) for the PCL:SV and .70 (se=.06) for the HCR-20.

Rice and Harris (1995) administered the VRAG (before release from a high secure psychiatric facility in Canada) to a sample of 868 violent men. Case files and interviews by forensic practitioners were used to complete the VRAG. Approximately 20% of the sample was classified as schizophrenic and 44% as having a personality disorder. Three and a half years later 799 men had been released and had their violent convictions were searched. The

VRAG had a high level of predictive validity with an AUC of .75 (se=.03).

Snowden et al. (2007) undertook a retrospective file study examining the validity of the VRAG and the OGRS for predicting violence amongst a group of 996 male patients (mean age =37.7) released from four medium secure psychiatric hospitals in England. The average time of hospitalization was 14.3 months and the violent reconvictions of the 862 men released were searched approximately 75 months after release. Both the VRAG and the OGRS had a strong relationship with future violence, with AUCs of .78 (se=.05) and .76 (se = .04) respectively.

Stadtland et al. (2005) used a German translation of the PCL-R to predict the future violence of a group of 258 offenders (mean age =34.6) assessed for their criminal responsibility. The PCL-R was coded retrospectively from case files and participants were followed up for an average period of 59 months. The results showed that the PCL-R did not significantly predict future non-violent offences (AUC= .564, 95% CI .484 - .644), but did significantly predict future violence (AUC = .715, 95% CI .615 - .815).

Tengstrom (2001) assessed the long-term predictive validity of the VRAG, the Historical part of the HCR-20 and the PCL-R using a group of violent men with schizophrenia who were detained for an average of 18 months. The risk assessment instruments were scored retrospectively using available case files and the men's criminal records were searched an average of 86 months after they were released. The results of the ROC analysis suggested that the PCL-R was the most accurate in predicting future violence (AUC=.79; 95% CI .71 – .87) and that the VRAG was somewhat less accurate (AUC=.68, 95%CI .52 – .71).

Tengstrom et al. (2000) evaluated the predictive accuracy of the PCL-R for predicting future violence using a group of 141 males referred for the first time to a court-ordered forensic psychiatric evaluation in Sweden and who were previously convicted of a violent (or sexual)

offence. Forensic psychiatric assessment files were used to complete the PCL-R and the Police National Register was used to code violent recidivism. Subjects were detained for an average of 24 months and the average time 'at risk' was 51 months. The PCL-R significantly predicted violence from one (AUC = .66, 95% CI .57 - .74) to five years (AUC = .75, 95% CI .62 - .86) after release.

Urbanik et al. (2006) used a German translation of the VRAG in a study of the violent recidivism of 79 violent and sexual offenders who had been forensically assessed by the Forensic Services of a single psychiatric hospital in Zurich. All variables used in the completion of the VRAG were coded from case files by a team of psychological experts, and an independent team coded the national and regional registered criminal records. Because of incomplete records it was not possible to determine the length of incarceration for offenders who were incarcerated or the average time 'at risk'. Overall the VRAG significantly predicted future violence according to the AUC (.72, se= .06).

Wong and Gordon (2006) used the 26-item Violence Risk Scale (VRS) to predict the future offending of a large group (n=918) of adult male offenders serving custodial sentences in prisons in Canada. This study reported on the results of several previous research projects in which files and interviews were used to score the VRS prior to the offender being released. Offenders served an average of 55 months and the research reported on several different follow-up times. After about 2 years at risk the VRS significantly predicted all convictions (AUC=.74, 95%CI .71 - .77), as well as violent convictions (AUC = .75, 95%CI .72 - .78) and non-violent (AUC= .72, 95%CI .69 - .76) convictions.

Analysis of Included Studies

In this review the AUC was used as the best measure of effect size (see e.g. Metz, Herman & Shen, 1998; Mossman, 1994). The standard error was used to calculate the

confidence intervals of this effect size (when not presented in the original report). Tables 2 to 5 show the 70 independent effect sizes classified by risk assessment instrument. Recall that the AUC is equivalent to the proportion of correct predictions in a 2 x 2 table where chance = 0.5 and perfect discrimination = 1.0 (Fergusson et al., 1977).

Table 2 shows the citation, the year of publication, the AUC and the associated confidence intervals for the 18 studies which used the VRAG. The study by Harris et al. (2002) had the strongest effect and the study by Loza et al. (2002) had the weakest effect. Overall, thirteen studies found that the VRAG significantly predicted violence and five studies did not.

Table 2. VRAG in the Prediction of Violence

Citation	Year	AUC	Lower CI	Upper CI
Harris et al	2002	.80	.72	.88
Snowden et al	2007	.78	.69	.86
Rice & Harris	1995	.75	.70	.80
Edens et al	2006	.73	.69	.77
Urbanik et al	2006	.72	.60	.84
Coid et al	2007	.70	.66	.74
Bonta & Yessine	2005	.68	.52	.84
Grann et al - 1	2000	.68	.62	.74
Tengstrom	2001	.68	.58	.78
Cooke et al	2001	.67	.59	.75
Howard	2007	.67	.66	.68
Mills & Kroner	2006	.67	.55	.79
Doyle & Dolan	2006	.66	.55	.77
Kroner & Loza	2001	.64	.42	.86
Loza and Green	2003	.63	.42	.84
Grann et al - 2	2000	.60	.50	.70
Kroner & Mills	2001	.60	.43	.76
Loza et al	2002	.52	.40	.64
Fixed Effects (18)		.68	.67	.69
Random Effects (18)		.69	.67	.72
Q = 41.3, p<.0009				

There are two main methods by which these 18 effect sizes can be combined to summarize the ability of the VRAG to predict violence. These are the fixed effects model and

the random effects model. The fixed effects model tends to produce a more conservative average effect size because it is more influenced by larger studies which tended to yield lower effect sizes. The random effects model gives more equal weight to all studies, large or small.

Combining these effect sizes together showed that the weighted standardized mean AUC of the 18 evaluations of the VRAG ranged between .68 (95% CI .67 - .69) and .69 (95% CI .67 - .72). Both of these estimates suggest that, taking all studies that used the VRAG together, this risk assessment instrument significantly predicted future violence. The heterogeneity (Q) of the effect size in the fixed effects model was significant at 41.3 ($p < .0009$). This suggests that there might be greater variation in the effect sizes than would be expected by chance. An implication of this is that the random effects model should be considered the most accurate estimate. Also, this variation could reflect the fact that the studies differed in key features such as sample size and the population under investigation (e.g. prison or psychiatric). This is investigated in a later section (Influence of Study Features).

Table 3 shows the citation, the year of publication, the AUC and the associated confidence intervals for the 18 studies which used the PCL-R or PCL: SV². The study by Douglas et al (1999) had the strongest effect and the study by Kroner and Mills (2001) had the weakest effect. Overall, 14 studies found that the PCL-R significantly predicted violence and 4 studies did not. Combining these effect sizes together showed that the weighted standardized mean AUC of the 18 evaluations of the PCL-R ranged between .66 (95% CI .65

² The PCL-R and PCL: SV were included together and will be referred to as the PCL-R because they are supposed to measure the same underlying concept (Hemphill & Hare, 2004).

Table 3. PCL-R and PCL: SV in the Prediction of Violence

Citation	Year	AUC	Lower CI	Upper CI
Douglas et al	1999	.79	.69	.89
Tengstrom	2001	.79	.71	.87
Edens et al	2006	.78	.74	.82
de Vogel et al	2004	.75	.65	.85
Grann	1999	.72	.64	.80
Hildebrand	2005	.72	.62	.82
Stadtland et al	2005	.72	.62	.81
Kroner & Loza	2001	.70	.48	.92
Nicholls	2004	.70	.56	.84
Tengstrom et al	2000	.70	.60	.80
Doyle & Dolan	2006	.69	.57	.80
Loza & Green	2003	.67	.46	.88
Cooke et al	2001	.65	.57	.73
Coid et al	2007	.64	.60	.68
Howard	2007	.64	.63	.65
Mills & Kroner	2006	.62	.50	.74
Gray et al	2004	.58	.47	.69
Kroner & Mills	2001	.56	.41	.72
Fixed Effects (18)		.66	.65	.67
Random Effects (18)		.69	.66	.73
Q = 77.6, p<.0001				

- .67) and .69 (95% CI .66 - .73). Both of these estimates suggest that, taking all studies that used the PCL-R together, this risk assessment instrument significantly predicted future violence. The heterogeneity of the effect size (Q) was significant at 77.6 (p< .0001), suggesting that the random effects estimate is more valid.

Table 4 shows the citation, the year of publication, the AUC and the associated confidence intervals for the 13 studies that included the HCR-20. Overall, 11 studies found that the HCR-20 significantly predicted violence and 2 studies did not. Combining these effect sizes together showed that the weighted standardized mean AUC of the 13 evaluations of the HCR-20 ranged between .66 (95% CI .65 - .67) and .69 (95% CI .66 - .73). Both of these estimates suggest that, taking all the evaluations that used the HCR-20 together, this risk assessment instrument significantly predicted future violence. The heterogeneity of the

Table 4. HCR-20 in the Prediction of Violence

Citation	Year	AUC	Lower CI	Upper CI
de Vogel et al	2004	.82	.74	.90
Doyle & Dolan	2006	.80	.70	.90
Douglas et al	1999	.80	.68	.92
Nicholls	2004	.75	.63	.87
Dahle	2006	.71	.65	.77
Cooke et al	2001	.69	.61	.77
Coid et al	2007	.67	.63	.71
Dolan et al	2004	.67	.47	.87
Douglas et al	2003	.67	.47	.87
Hildebrand	2005	.67	.57	.77
Howard	2007	.66	.65	.67
Kroner & Mills	2001	.62	.47	.77
Gray et al	2004	.56	.45	.67
Fixed Effects (13)		.67	.66	.68
Random Effects (13)		.70	.66	.74
Q = 36.0, p<.0003				

effect size (Q) was significant at 36.0 (p<.0003).

Table 5 shows the citation, the year of publication, the AUC and the associated confidence intervals for the 7 studies that included the LSI-R or LSI-SV³. Overall, six studies found that the LSI-R significantly predicted violence and one study did not. Combining these

Table 5. LSI-R and LSI-SV in the Prediction of Violence

Citation	Year	AUC	Lower CI	Upper CI
Girard & Wormith	2004	.68	.62	.74
Mills & Kroner	2006	.68	.56	.80
Kroner & Mills	2001	.67	.51	.82
Loza & Green	2003	.67	.46	.88
Bonta & Yessine	2005	.66	.58	.74
Dahle	2006	.65	.59	.71
Barnoski	2003	.64	.63	.65
Fixed Effects (7)		.64	.63	.66
Random Effects (7)		.64	.63	.66
Q = 2.5, n.s.				

³ There was only one study which used the LSI-SV (Bonta & Yessine, 2005). These devices were combined and are referred to as the LSI-R because they are supposed to measure similar underlying concepts.

effect sizes together showed that the weighted standardized mean AUC of the 7 evaluations of the LSI-R was .64 (95% CI .63 - .66). Because the heterogeneity of this effect size was not significant (Q = 2.5, n.s.), the fixed effects and random effects models produced the same effect size estimate.

Table 6 shows the citation, the year of publication, the AUC and the associated confidence intervals for two actuarial instruments (Offender Group Reconviction Scale and General Statistical Information on Reconviction). Four studies evaluated the predictive validity of each. Combining these effect sizes together for the OGRS showed that the weighted standardized mean AUC of the 4 evaluations ranged between .72 (95% CI .71 -

Table 6. OGRS and GSIR in the Prediction of Violence

<i>OGRS</i>				
Citation	Year	AUC	Lower CI	Upper CI
Snowden et al	2007	.76	.69	.83
Coid et al	2007	.72	.72	.72
Gray et al	2004	.71	.60	.82
Howard	2007	.66	.65	.67
Fixed Effects (4)		.72	.71	.72
Random Effects (4)		.71	.66	.75
Q = 90.0, p<.0001				
<i>GSIR</i>				
Citation	Year	AUC	Lower CI	Upper CI
Bonta & Yessine	2005	.77	.69	.85
Kroner & Loza	2001	.74	.52	.96
Loza & Green	2003	.74	.53	.95
Mills & Kroner	2006	.67	.57	.77
Fixed Effects (4)		.73	.68	.79
Random Effects (4)		.73	.68	.79
Q = 2.5, n.s.				

.72) and .71 (95% CI .66 - .75). The heterogeneity of the effect size (Q) associated with this comparison was significant at 90.0 (p<.0001). Combining the effect sizes together for the GSIR showed that the weighted standardized mean of the 4 evaluations was .73 (95% CI .68 - .79). The heterogeneity of the effect size (Q) was not significant (Q = 2.5, n.s.).

Table 7 shows the citation, the year of publication, the AUC and the associated confidence intervals for the studies which used risk assessment instruments which had two or fewer independent evaluations. These instruments were the Risk Matrix 2000 violence (RM2000V), the Self-Assessment Questionnaire (SAQ), the Offender Assessment System Sexual and Violence Predictor (OASys SV), and the Violence Risk Scale (VRS).

Table 7. RM2000V and SAQ in the Prediction of Violence

<i>RM 2000V</i>				
Citation	Year	AUC	Lower CI	Upper CI
Coid et al	2007	.69	.68	.69
Howard	2007	.67	.66	.68
Fixed Effects (2)		.68	.68	.69
Random Effects (2)		.68	.66	.69
Q = 5.5, p<.02				
<i>SAQ</i>				
Citation	Year	AUC	Lower CI	Upper CI
Kroner & Loza	2001	.75	.53	.97
Loza & Green	2003	.71	.50	.92
Fixed Effects (2)		.73	.57	.88
Random Effects (2)		.73	.57	.88
Q = 0.1, n.s.				
<i>OASys SV</i>				
Citation	Year	AUC	Lower CI	Upper CI
Howard	2007	.71	.70	.72
<i>VRS</i>				
Citation	Year	AUC	Lower CI	Upper CI
Wong & Gordon	2006	.74	.71	.77

In the case of the RM2000V the weighted standardized mean AUC of the two evaluations was .68 (95% CI .68 - .69 and .66 - .69). The heterogeneity of the effect size (Q) was significant at 5.5 (p< .02). For the SAQ, the weighted standardized mean AUC of the

two evaluations was .73 (95% CI .57 - .88). The heterogeneity of the effect size (Q) was not significant (Q = 0.1, n.s.).

The AUCs for both the VRS and the OASys SV were moderately high. However, given that only one effect size was available for each instrument, and the fact that both measures have as yet only been evaluated by the instrument developers, further research on these promising measures is required before conclusions about their ability to predict future violence can be drawn.

Which Instrument is ‘Best’?

Table 8 summarises the main findings from Tables 2 to 6, by showing the overall mean effect sizes (based on random effects models) for the risk assessment instruments that had more than two evaluations. When ordered according to the mean effect size, the GSIR appears to be the most accurate instrument and the LSI-R the least accurate. However, significant differences at the $p < .05$ level are identified by non-overlapping 95% confidence intervals. This means that the GSIR (AUC= .73, 95% CI .68 - .79) was significantly more accurate in predicting future violence than the LSI-R (AUC= .64, 95% .63 - .66). This was the only significant difference between any of the risk assessment instruments.

Table 8. Mean Effect Sizes for the Risk Assessment Instruments

Device	N	AUC	Lower CI	Upper CI
GSIR	4	0.73	0.68	0.79
OGRS	4	0.71	0.66	0.75
HCR – 20	13	0.70	0.66	0.74
PCL-R	18	0.69	0.66	0.73
VRAG	18	0.69	0.67	0.72
LSI – R	7	0.64	0.63	0.66

In order to confirm the finding of a non-statistical difference between the predictive ability of the various risk assessment instruments, the analogue of an analysis of variance (ANOVA) was used. Table 9 shows the results. As already indicated, the GSIR was

significantly more accurate than the LSI-R, but no other differences were significant. If a two-tailed ($p < .10$) criterion was used, the VRAG and the HCR-20 would also have proved significantly more accurate than the LSI-R.

Table 9 Comparison of the Predictive Ability of the Risk assessment Instruments

Instrument 1	Instrument 2	Q Between Groups	p
VRAG	PCL-R	0.04	n.s.
VRAG	HCR - 20	0.12	n.s.
VRAG	LSI - R	3.0	0.08
VRAG	OGRS	0.31	n.s.
VRAG	GSIR	0.94	n.s.
PCL-R	HCR - 20	0.07	n.s.
PCL-R	LSI - R	1.3	n.s.
PCL-R	OGRS	0.2	n.s.
PCL-R	GSIR	0.51	n.s.
HCR-20	LSI - R	2.7	0.10
HCR-20	OGRS	0.06	n.s.
HCR-20	GSIR	0.47	n.s.
LSI-R	OGRS	2.4	n.s.
LSI-R	GSIR	9.0	0.003
OGRS	GSIR	0.24	n.s.

Comparison of Effect Size and Study Features

The analysis described in Table 8, while informative, does not take into consideration the variation that existed between the evaluations that were included in this review. For example, some evaluations used samples of those released from prisons, while others used samples released from psychiatric hospitals. Furthermore, some evaluations were prospective (the measures were administered directly to the offenders before they were released), while other evaluations were retrospective (the measures were completed using file information after the offenders had been released).

In order to investigate the effects that this variability might have had on the predictive validity of the various risk assessment instruments, a coding protocol was developed to

Table 10. Features of the Included Studies

Study ID	Publication, Location	Risk Assessment Device(s)	Correctional/ Psychiatric Sample	Prospective/ Retrospective Study	Sample Size	Mean age of Sample	% of Sample White	Mean Follow-up (months)	Measure of Violence
093	Barnoski, 2003 United States	LSI-R	Correctional	Prospective	22533	N/A	N/A	24	Violent Conviction
041	Bonta and Yessine, 2005 Canada	VRAG, LSI-SV, GSIR	Correctional	Retrospective	202	35	68	40.8	Violent Conviction
045	Coid et al, 2007 England	VRAG, PCL-R, HCR-20, OGRS, RM2000V	Correctional	Prospective	1353	31	79	23.6	Violent Conviction
052	Cooke et al, 2001 Scotland	VRAG, PCL-R, HCR-20	Correctional	Prospective	190	N/A	N/A	24	Violent Conviction
043	Dahle, 2006 Germany	HCR-20, LSI-R	Correctional	Retrospective	309	30	N/A	120	Violent Conviction
092	De Vogel et al, 2004 Netherlands	PCI-R, HCR-20	Psychiatric	Retrospective	119	26	N/A	73	Violent Conviction
089	Dolan & Doyle, 2006 England	VRAG, PCL-SV, HCR-20	Psychiatric	Prospective	112	40	93	6	Violent Conviction and collateral/SR reports
001	Dolan & Khawaja, 2004 England	HCR-20 (without PCL-R)	Psychiatric	Retrospective	70	44	83	59	Violent Conviction and collateral/SR reports.

Study ID	Publication, Location	Risk Assessment Device(s)	Correctional/ Psychiatric Sample	Prospective/ Retrospective Study	Sample Size	Mean age of Sample	% of Sample White	Mean Follow-up (months)	Measure of Violence
002	Douglas et al, 2003 Canada	HCR-20	Psychiatric	Retrospective	100	N/A	N/A	43	Violent Conviction and file information about violence
036	Douglas et al, 1999 Canada	PCL-SV, HCR-20	Psychiatric	Retrospective	193	38	79	21	Violent Conviction
082	Edens et al, 2006 United States	VRAG*, PCL-SV	Psychiatric	Prospective	818	30	69	5	Violent Conviction and collateral/SR reports
055	Girard & Wormith, 2004 Canada	LSI – R	Correctional	Prospective	630	32	88	31	Violent Conviction
053	Grann et al, 2000 Sweden	VRAG	Sample 1 – Mixed Sample 2 – psychiatric	Retrospective	Sample 1 – 293 Sample 2 – 111	Sample 1 – 32 Sample 2 – 33	N/A	24	Violent Conviction
007	Grann et al., 1999 Sweden	PCL-R	Mixed	Retrospective	352	N/A*	32	24	Violent Conviction
010	Gray et al, 2004 England	PCL-SV, HCR-20, OGRS	Psychiatric	Retrospective	315	31	84	72	Violent Conviction
039	Harris et al., 2002 Canada	VRAG	Psychiatric	Retrospective*	133	N/A	N/A	60	Violent Conviction
095	Hildebrand et al, 2005 Netherlands	PCL-R, HCR - 20	Psychiatric	Retrospective	153	N/A	N/A	86	Violent Conviction
096	Howard, 2007 England and Wales	VRAG*, PCL-SV*, HCR-20*, OGRS, RM2000V, OASys	Correctional	Retrospective	26619	N/A	93	24	Violent Conviction

Study ID	Publication, Location	Risk Assessment Device(s)	Correctional/ Psychiatric Sample	Prospective/ Retrospective Study	Sample Size	Mean age of Sample	% of Sample White	Mean Follow-up (months)	Measure of Violence
031	Kroner & Loza, 2001 Canada	VRAG, PCL-R, GSIR, SAQ	Correctional	Prospective	78	29	87	20	Violent Conviction
037	Kroner & Mills, 2001 Canada	VRAG, PCL-R, HCR-20, LSI-R, LCSF	Correctional	Prospective	87	28	75	23	Violent Conviction
038	Loza & Green, 2003 Canada	VRAG, PCL-R, LSI-R, GSIR, SAQ	Correctional	Prospective	91	30	82	60	Violent Conviction
012	Loza et al, 2002 Canada	VRAG	Correctional	Prospective	124	30	83	24	Violent Conviction
013	Mills & Kroner, 2006 Canada	VRAG, PCL-R, LSI-R, GSIR	Correctional	Prospective	209	30	75	14	Violent Conviction
090	Nicholls et al, 2004 Canada	PCL-SV, HCR-20	Psychiatric	Retrospective	PCL-SV – 146, HCR-20 – 117	36	78	23	Violent Conviction
066	Rice & Harris, 1995 Canada	VRAG	Psychiatric	Prospective	799	N/A	N/A	42	Violent Conviction
069	Snowden et al. 2007 Wales	VRAG, OGRS	Psychiatric	Retrospective	VRAG – 421 OGRS – 641	38	69	75	Violent Conviction
020	Stadland et al, 2005 Germany	PCL-R	Mixed	Retrospective	258	35	81	59	Violent Conviction

Study ID	Publication, Location	Risk Assessment Device(s)	Correctional/ Psychiatric Sample	Prospective/ Retrospective Study	Sample Size	Mean age of Sample	% of Sample White	Mean Follow-up (months)	Measure of Violence
022	Tengstrom, 2001 Sweden	VRAG, PCL-R	Psychiatric	Retrospective	106	33	N/A	86	Violent Conviction
023	Tengstrom et al, 2003 Sweden	PCL-R	Psychiatric	Retrospective	131	N/A	N/A	51	Violent Conviction
024	Urbaniok et al, 2006 Switzerland	VRAG	Mixed	Retrospective	79	N/A	N/A	N/A	Violent Conviction
027	Wong & Gordon, 2006 Canada	VRS	Correctional	Prospective	758	34	55	53	Violent Conviction

* Estimation was involved in the completion of the risk instrument.

identify the key features of the included studies. Table 10 summarizes the key features of the studies that could be consistently coded. Obviously, it was not possible to obtain information about all of the potentially relevant features from all of the studies. Importantly, the level of seriousness of the sample under study was rarely available, and it was not always clear who had administered the risk assessment instrument.

1. Country where the research was conducted

Seven evaluations were conducted in the UK (one in Scotland), 15 were conducted in North America and 9 were conducted in Europe.

2. Year the study was published

The year the evaluations was published was coded in case there has been an improvement in predicting violence over time, with more recent evaluations providing more accurate estimates.

3. Age of the sample

The age of the sample was included to assess whether any instrument predicted violence better for any particular age range. The mean age of the sample was available in 24 evaluations.

4. Ethnic composition of the sample

Some risk assessment instruments might work better with some ethnic groups rather than with others. No evaluations presented results separately by ethnic group. However, an indicator that was available in most (19) evaluations was the ethnic composition of the sample. This was coded as the proportion of the sample who were identified as white, and this ranged from 55 per cent to 93 per cent.

5. Total sample size

Previous research has found that small evaluations tend to have higher effect sizes, possibly reflecting either their poorer methodological standards, their better quality control, or publication bias (Farrington & Welsh, 2003). The evaluations had sample sizes (after attrition) that ranged from 70 to 26619.

6. Psychiatric or correctional sample

Some risk assessment instruments were designed for predicting future violence primarily among psychiatric samples (e.g. HCR-20) and others primarily for correctional samples (e.g. LSI-R). Their predictive ability may be less when the instruments are used with different samples. Thirteen of the evaluations were based on correctional samples and 15 were based on psychiatric samples. Four additional evaluations were classified as mixed, because they were based on case files of those sent for forensic evaluation (e.g. Grann et al., 1999) where subsequently some offenders went to prison and others went to a psychiatric hospital.

7. Prospective or retrospective study

Some of the studies employed a prospective methodology whereby the risk assessment instrument was completed before the offenders were released, while in other studies the risk assessment instrument was completed retrospectively, after the offenders had been released. To the extent that an instrument might be more accurate if it was administered prospectively because of the additional information that could be obtained in an interview, we would expect that evaluations which used prospective administration might produce greater effect sizes. This might be the case with a measure such as the LSI-R or the HCR-20, but it would be less likely with an actuarial measure like the VRAG. However, there is the possibility that retrospective studies were not completely blind to the violent offending outcome. If this was the case then retrospective studies would appear more effective at predicting future violence, not because they actually were more effective, but because of this

research artifact.

8. Mean follow-up time

The predictive ability of risk assessment instruments may be related to the length of time that the sample was followed after release from the prison or psychiatric hospital. The length of follow-up ranged from 5 to 120 months (mean = 37, sd = 26.5).

9. Measure of violence

Given that violence leading to convictions represents only the tip of the iceberg of actual violent offending, ideally studies should collect information about additional measures of violence such as self or collateral reports of violence. Only four studies included supplemental information about violence⁴.

Correlations between Effect Sizes and Study Features

Correlations were used to investigate the relationships between the study features measured on a continuous scale (e.g. year, total sample size) and the effect sizes (AUC values) of the VRAG, PCL-R, the HCR-20 and the LSI-R⁵. Because the number of evaluations was relatively small, and information was missing in some instances, few

Table 11 Correlation of Study Features with Effect Sizes

	VRAG		PCL-R		HCR-20		LSI-R	
	N	R	N	r	N	R	N	r
Year	18	.16	18	-.23	13	-.11	7	.06
Age of Sample	13	.40	14	.30	9	.18	6	-.11
Ethnic Composition (% white)	11	-.42	13	-.25	8	.14	5	.69
Sample size	18	-.00	18	-.19	13	-.16	7	-.70
Length of Follow-up	17	.44	18	.21	13	-.17	1	-.44

statistically significant results would be expected. However, as a rule of thumb correlations with a magnitude of greater than or equal to $r = 0.2$ were considered meaningful. These

⁴ Because this information was available in such a small number of studies it was not possible to analyse the impact of this variable on the predictive accuracy of the risk assessment instruments.

⁵ There were too few effect sizes to include the OGRS and the GSIR in these analyses.

correlations are informative but do not necessarily indicate any causal effects of the study features on effect size (see Table 11).

Only seven of a possible 16 correlations met the criteria of $r \geq .2$ when the effect sizes were correlated with the study features. There was some evidence to suggest that the accuracy of the PCL-R in predicting violence may have decreased over time ($r = -.23$). There was also an indication that the violence of older samples might have been better predicted than younger samples by the VRAG ($r = .40$) and the PCL-R ($r = .30$). Both the VRAG and the PCL-R appeared to have greater predictive ability when the percentage of the sample who were white was lower ($r = -.42$ and $r = -.25$ respectively), but the reverse appeared to be the case with the LSI-R (i.e. greater predictive ability among samples that had a higher proportion of white offenders, $r = .69$). Also, for the LSI-R, evaluations with smaller samples had greater predictive accuracy ($r = -.70$).

Both the VRAG and the HCR-20 were positively correlated with the length of follow-up, but the LSI-R was negatively correlated to this variable.

Comparison of Effect Sizes with Categorical Study Features

Table 12. Comparison of Effect Size by Location

	North America		United Kingdom			Europe		Q Between Groups
	N	AUC	N	AUC	N	AUC		
VRAG	9	.71 (95% CI .66 - .75)	5	.69 (95% CI .65 - .73)	4	.67 (95% CI .65 - .73)	1.1, n.s.	
PCL-R	7	.73 (95% CI .69 - .78)	6	.65 (95% CI .62 - .68)	5	.74 (95% CI .69 - .78)	15.4, p<.0005	
HCR-20	4	.73 (95% CI .65 - .80)	7	.67 (95% CI .64 - .71)	2	.76 (95% CI .69 - .82)	5.4, n.s.	
LSI-R	6	.66 (95% CI .62 - .70)	0		1	.65 (95% CI .59 - .71)	.05, n.s.	

Table 12 shows the relationship between the location where the research was conducted and the effect size. For example, the AUC for the 9 studies which used the VRAG in North America was .71 (95% CI .66 - .75). The effect size was slightly lower among the 5

studies that used the VRAG in the United Kingdom (AUC = .69, 95% CI .65 - .73) and the 4 studies in Europe (AUC = .67, 95% CI .65 - .73). However, when these effect sizes were compared (using the analog of the ANOVA), the Q Between Groups was found to be 1.1, suggesting no significant difference in the predictive ability of the VRAG across the three locations.

However, a significant relationship was identified between the predictive accuracy of the PCL-R and the location of the study. That is, the AUC of the studies that used the PCL-R in England was lower (AUC = .65, 95% CI .62 - .68) than for those conducted in North America (AUC = .73, 95% CI .69 - .78) or Europe (AUC = .74, 95% CI .69 - .78). The Q Between Groups of 15.4 shows that this difference was significant. This can also be seen by the fact that the confidence intervals of the UK studies do not overlap with those in North America or Europe.

There were no significant differences by location for evaluations that used either the HCR-20 or the LSI-R. Furthermore, when looking for differences between risk assessment measures within locations (comparing AUCs within columns of Table 12) no significant differences were found.

Table 13. Comparison of Effect Size by Type of Study

	Prospective		Retrospective		Q Between Groups
	N	AUC	N	AUC	
VRAG	11	.70 (95% CI .66 - .73)	7	.68 (95% CI .64 - .72)	0.3, n.s.
PCL-R	8	.67 (95% CI .61 - .73)	10	.71 (95% CI .66 - .76)	0.9, n.s.
HCR-20	4	.70 (95% CI .63 - .77)	9	.70 (95% CI .65 - .75)	0.3, n.s.
LSI-R	5	.64 (95% CI .63 - .67)	2	.65 (95% CI .61 - .70)	0.2, n.s.

Table 13 shows the relationship between the type of study (prospective or retrospective) and the effect sizes. The results show that there was no significant difference

in predictive accuracy between studies that administered the risk assessment instrument before the offender was released and those that completed the instrument retrospectively. This was true for all four risk assessment instruments included in this analysis. Furthermore, no statistical differences were evident when comparisons were made between types of risk assessment instruments within the prospective and retrospective type of study (comparisons within columns).

Table 14. Comparison of Effect Size by Type of Sample

	Correctional		Psychiatric		Q Between Groups
	N	AUC	N	AUC	
VRAG	9	.67 (95% CI .63 - .69)	7	.73 (95% CI .70 - .76)	6.3, p<.01
PCL-R	7	.64 (95% CI .62 - .66)	9	.74 (95% CI .72 - .77)	24.7, p<.0001
HCR-20	5	.67 (95% CI .63 - .71)	8	.73 (95% CI .68 - .78)	3.4, p<.06
LSI-R	7	.64 (95% CI .63 - .67)			

Table 14 shows the relationship between the type of sample and the effect sizes. Because the evaluations of mixed samples (evaluations including both correctional and psychiatric samples) could not be classified as either correctional or psychiatric, they were not included in this analysis⁶. Interestingly, the results show that evaluations which used the VRAG with psychiatric samples had significantly higher effect sizes than those which used correctional samples (Q Between Groups = 6.3, p<.01). Furthermore, evaluations that used the PCL-R with psychiatric samples also had significantly higher mean effect sizes than those which used correctional samples (Q Between Groups = 24.7, p<.0001). This pattern was also evident with evaluations that used the HCR-20, but the Q Between Groups of 3.4 just failed to reach significance (p<.06). The LSI-R was only used with correctional samples.

⁶ There were 4 evaluations which used mixed samples. Two of these used the VRAG, and the mean effect size of these evaluations was .69 (95% CI .62 - .76). Two used the PCL-R with a mean effect size of .72 (95% CI .65 - .78).

Discussion

There are many risk assessment instruments which may assist practitioners to identify persons who are at risk of committing violence. This systematic review focused on those instruments that have been used with adult males in which the Area Under the Curve (AUC) could be used as the measure of accuracy in predicting violence upon release from a prison or hospital. Overall 31 studies were identified which met our inclusion criteria and these contained a total of 70 effect sizes. Eighteen of these used the VRAG, 18 used the PCL-R, 13 used the HCR-20, 7 used the LSI-R, 8 used the OGRS/GSIR actuarial devices, and 6 used four other measures (RM2000V (2), SAQ (2), VRS, OASys SV).

Using the AUC as the measure of effect size, it was found that all risk assessment instruments included in this review performed significantly better than chance in the prediction of future violence. For some instruments (those with two or fewer effect sizes) it was not possible to undertake further analyses, but the other instruments were compared based on their predictive ability.

For the most part, the results suggested that the instruments performed comparably in predicting future violence. This is perhaps not surprising, since many of them contain similar items. However, the evaluations that used the LSI-R appeared to have somewhat lower effect sizes than those that used the VRAG and HCR-20. In fact, this difference in effect sizes was significant between evaluations of the LSI-R and the GSIR. The difference between the predictive accuracy of the LSI-R, the VRAG and the HCR-20 was probably because the VRAG and HCR-20 were used more with psychiatric as opposed to correctional populations. Unfortunately, because of the small number of effect sizes available for the GSIR it was not possible to investigate further the difference between this instrument and the LSI-R.

However, both the GSIR and the LSI-R were only used with correctional samples, suggesting that the sample used was not the explanation.

It was possible to examine the relationship between some of the features of the evaluations and the effect sizes produced by the VRAG, PCL-R, HCR-20 and the LSI-R in predicting violence. Interestingly, evaluations that administered the risk assessment instruments prospectively did not differ in mean effect size from those studies that administered the risk assessment instrument retrospectively. This was true for all four measures. While further research is needed to support this finding, if this was accurate it would mean that interviews and offender compliance may not be required to obtain the most accurate prediction of future violence.

Perhaps the key finding of this research was that evaluations that used the VRAG and PCL-R with psychiatric samples had significantly higher effect sizes than those that used correctional samples. The HCR-20 showed a similar, and nearly-significant, relationship. This suggests that these instruments might be more useful at predicting violence among psychiatric populations. This might be because the VRAG was designed and validated on this type of population (e.g. Harris, Rice & Quinsey, 1993), but the PCL-R was designed for correctional populations.

Another potentially interesting finding was that studies that used the PCL-R in the UK had significantly lower effect sizes than those in Europe and North America. This might mean that the PCL-R is less accurate in predicting future violence in the UK. It is not because the PCL-R was more likely to be used with correctional populations. Three of the six studies that used the PCL-R in the UK were conducted using correctional populations and 3 were conducted using psychiatric populations. This was similar to the situation in North America where 4 of the 7 studies using the PCL-R were conducted using correctional samples.

Based on a negative correlation between the effect sizes and the year of the study, there was some suggestion that the PCL-R might be getting less predictive. However, this may be attributed to the fact that the PCL-R has been used recently in large studies of correctional samples in the UK. Both evaluations using correctional samples and evaluations in the UK had lower effect sizes.

The effect sizes of the evaluations that used the VRAG and the PCL-R were somewhat higher among evaluations that used older offenders. While this might mean that these instruments might be more accurate in predicting violence among older individuals, this again may reflect the fact that evaluations that used older offenders also tended to use psychiatric samples. In support of this, the average age in the psychiatric evaluations was 34.5, while the average age in the correctional evaluations was 30.0. This difference was statistically significant ($t = 4.1, p < .0001$).

The differential use of psychiatric versus correctional samples did not explain the possible difference with respect to the ethnic composition of the sample. The results suggest that the VRAG and the PCL-R might be more accurate with samples containing a lower percentage of white subjects. For the LSI-R, however, the finding was the reverse. Obviously, the relationship between the effectiveness (and applicability) of various risk assessment instruments and ethnic origin requires future research. Ideally research reports should disaggregate results by ethnicity to facilitate this.

In regard to the LSI-R, there was a relationship between the effect sizes of the evaluations and the sample size. This suggested that smaller studies produced more accurate predictions of future violence. This result might be driven by the effect size of one particularly large study (Barnoski, 2003), or it might reflect the fact that smaller studies had

greater quality control when completing the LSI-R. Also, the LSI-R had greater predictive efficiency for shorter follow-up periods.

What can we conclude about the relative advantages and disadvantages of the different instruments? First, the LSI-R had the lowest predictive efficiency and also had lower accuracy for larger samples and longer follow-up periods. These are worrying conclusions. The LSI-R has been commonly used to assist in decisions about risk. However, the LSI-R was only used with correctional samples, where predictive efficiency was generally low, and it is now being replaced by the LS/CMI. Further research will be needed to ascertain fully whether the LS/CMI performs better than the LSI-R. The most accurate instruments for correctional samples were the actuarial OGRS and GSIR.

It may be true that actuarial instruments based on static predictors are more accurate than dynamic instruments including changeable risk (and need) factors, just as it has been argued that statistical prediction is superior to clinical prediction (e.g. Grove and Meehl, 1996). However, the use of dynamic instruments is desirable to measure change, and the use of structured clinical judgement is desirable to identify unusual cases (see Webster and Hucker, 2007).

The VRAG, PCL-R and HCR-20 had comparable predictive efficiency, but were more accurate with psychiatric samples. The PCL-R was less accurate in the United Kingdom than in North America or Europe. Therefore, for the United Kingdom, the VRAG and HCR-20 seem to be associated with more accurate estimates of risk. However, it is important to be cautious in interpreting this finding. Psychopathy is a complex construct to assess and it is possible that, in psychiatric samples, a formal diagnosis might be difficult to achieve when there is a comorbid mental disorder. While the VRAG is static, the HCR-20 contains both static and dynamic items. The advantage of dynamic items is that they can measure within-

individual change over time. In addition, they are sensitive to individual risk as opposed to group results (e.g. Hart, Michie & Cooke, 2007). On balance, we conclude that, on the basis of currently available evidence, the HCR-20 seems the most useful of these instruments.

More research is clearly needed in the future. In particular, there is a need to establish the predictive validity of individual items (ideally based on theories of violence). This should help delineate risk processes and the relevance of those risks to the person being assessed. In addition, it is important to develop more dynamic prediction instruments to allow for change to be identified. There is already some progress in this area with tools such as the Short-Term Assessment of Risk and Treatability or START (Webster et al., 2004). More evaluation research, involving prospective follow-ups of large samples, is needed to evaluate the risk assessment instruments used in Scotland. For the present, it seems useful to build on simple static instruments such as OGRS and dynamic instruments such as the HCR-20. However, a final point is that Part 1 of our research was confined to identifying those instruments that contained items that were predictive. The field of risk assessment is fast evolving and it is widely acknowledged that assessing the “right risk factors” is only part of the process of risk assessment. We consider a range of other practical issues in part 2.

References

- Andrews, D. A. & Bonta, J. (1995). The Level of Service Inventory – Revised. Toronto: Multi-Health Systems.
- Andrews, D. A. & Bonta, J. (2006). The Psychology of Criminal Conduct. New York: Anderson Publishing.
- Andrews, D.A., Bonta, J., & Wormith, S.J. (2004). The Level of Service/Case Management Inventory. Toronto: Multi-Health Systems.
- Barnoski, R., & Aos, S. (2003). Washington's Offender Accountability Act: An Analysis of the Department of Corrections' Risk Assessment. Olympia, WA: Washington State Institute for Public Policy (December, 2003).
- Bonta, J. & Yessine, A. K. (2005). The national flagging system: Identifying and responding to High-Risk Violent Offenders. Ottawa: Public Safety Canada (User Report 2005 – 04).
- Campbell, M.A., French, S., & Gendreau, P. (2008). Assessing the Utility of Risk Assessment Tools and Personality Measures in the Prediction of Violent Recidivism for Adult Offenders. Ottawa: Public Safety Canada (User Report 2007-04, revised April 2008).
- Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., Farrington, D. P. & Rogers, R. D. (2008). A comparison of five structured risk assessment instruments to predict violence. Psychological Medicine (in press).
- Cooke, D. J., Mitchie, C. & Ryan, J. (2001). Evaluating Risk for Violence: A Preliminary Study of the HCR – 20, PCL-R and VRAG in a Scottish Prison Sample. Scottish Prison Service Occasional Paper Series 5/2001.
- Copas, J., & Marshall, P. (1998). The offender group reconviction scale: A statistical reconviction score for use by probation officers. Applied Statistics, 47, 159 – 171.
- Dahle, K. P. (2006). Strengths and limitations of actuarial prediction of criminal reoffence in a German sample: A comparative study of LSi-R, HCR-20 and PCL-R. International Journal of Law and Psychiatry, 29, 431 – 442.
- De Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B. & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. International Journal of Forensic Mental Health, 3, 149 – 165.
- Dolan, M. & Khawaja, A. (2004). The HCR-20 and post-discharge outcome in male patients discharged from medium security in the UK. Aggressive Behavior, 30, 469 – 483.
- Douglas, K. S., Ogloff, J. R. P. & Hart, S. D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients. Psychiatric Services, 54, 1372 – 1379.

- Douglas, K. S., Ogloff, J., R., P., Nicholls, T., L. & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the psychopathy checklist: screening version. Journal of Consulting and Clinical Psychology, 67, 917 – 930.
- Doyle, M. & Dolan, M. (2006). Predicting community violence from patients discharged from mental health services. British Journal of Psychiatry, 189, 520 – 526.
- Edens, J. F., Campbell, J. S. Weir, J. M. (2007). Youth psychopathy and criminal recidivism: A meta-analysis of the psychopathy checklist measures. Law and Human Behavior, 31, 53 – 75.
- Edens, J. F., Skeem, J. L., & Douglas, K. S. (2006). Incremental validity analyses of the Violence Risk Appraisal Guide and the Psychopathy Checklist: Screening Version in a civil psychiatric sample. Assessment, 13, 368-374.
- Farrington, D. P. & Petrosino, A. (2001). The Campbell Collaboration Crime and Justice Group. Annals of the American Academy of Political and Social Science, 578, 35-49.
- Farrington, D. P. and Welsh, B. C. (2002) Effects of Improved Street Lighting on Crime: A Systematic Review. London: Home Office. (Research Study No. 251).
- Fergusson, D. M., Fifield, J. K. & Slater, S. W. (1977). Signal detectability theory and the evaluation of prediction tables. Journal of Research in Crime and Delinquency, 14, 237-246.
- Gendreau, P., Goggin, C. & Smith, P. (1999). Predicting recidivism: LSI-R vs. PCL-R. Canadian Psychology Abstracts, 40, 2a.
- Gendreau, P., Goggin, C. & Smith, P. (2002). Is the PCL-R really the 'unparalleled' measure of offender risk? A lesson in knowledge accumulation. Criminal Justice and Behavior, 29, 397 – 426.
- Girard, L., & Wormith, J. S. (2004). The predictive validity of the Level of Service Inventory – Ontario Revision on general and violent recidivism among various offender groups. Criminal Justice and Behavior, 31, 150-181.
- Grann, M., Belfarge, H., & Tengstrom, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and historical part of the HCR-20. Criminal Justice and Behavior, 27, 97 – 114.
- Grann, M., Langstrom, N., Tengstrom, A. & Kullgren, G. (1999). Psychopathy (PCL-R) predicts violent recidivism among criminal offenders with personality disorders in Sweden. Law and Human Behavior, 23, 205 – 217.
- Gray, N. S., Snowden, R. J., MacCulloch, S., Phillips, H., Taylor, J., MacCulloch, M. J., (2004). Relative efficacy of criminological, clinical and personality measures of future

risk of offending in mentally disorderd offenders: A comparative study of HCR-20, PCL:SV and OGRS. Journal of Consulting and Clinical Psychology, 72, 523 – 530.

- Grove, W. & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. Psychology, Public Policy and Law, 2, 293-323.
- Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. Journal of Interpersonal Violence, 20, 212 – 217.
- Hanson, R. K. & Morton-Bourgon, K. E. (2007). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis. Ottawa: Public Safety Canada (User Report 2007-01).
- Harris, G. T., Rice, M. E. & Cormier, C. A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. Law and Human Behavior, 26, 377 – 394.
- Harris, G. T., Rice, M. E. & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. Criminal Justice and Behavior, 20, 315 – 335.
- Hart, S. D., Cox, D. N., & Hare, R. D. (1995). The Hare Psychopathy Checklist – Screening Version (PCL:SV). Toronto: Multi-Health Systems.
- Hart, S. D., Michie, C. & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments. British Journal of Psychiatry, 190 (suppl. 49), S60-S65.
- Hemphill, J. F., & Hare, R. D. (2004). Some misconceptions about the Hare PCL-R and risk assessment: A reply to Gendreau, Goggin, and Smith. Criminal Justice and Behavior, 31, 203-243.
- Hemphill, J. F., Hare, R. D. & Wong, S. (1998). Psychopathy and recidivism: A review. Legal and Criminological Psychology, 3, 139 – 170.
- Hildebrand, M., Hesper, B. L., Spreen, M. & Nijman, H. L. I. (2005). The importance of structured risk assessment and the diagnosis of psychopathy: A study in the reliability and predictive validity of he HCR-20, HKT-30 and the PCL-R. Utrecht, the Netherlands: Expertisecentrum Forensische Psychiatrie.
- Howard, P. (2007). OASys General Reoffending Predictor and OASys Violent/Sexual Predictor. Unpublished manuscript.
- Kester, A., D., M. & Buntinx, F. (2000). Meta-analysis of ROC curves. Medical Decision Making, 20, 430 – 439.
- Kroner, D. & Loza, W. (2001). Evidence for the efficacy of self-report in predicting nonviolent and violent criminal recidivism. Journal of Interpersonal Violence, 16, 168 – 177.

- Kroner, D. G. & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. Criminal Justice and Behavior, 28, 471 – 489.
- Lipsey, M. W. & Wilson, D. B. (2001). Practical Meta-Analysis. Thousand Oaks, CA: Sage.
- Loza, W. & Dhaliwal, G. K. (1997). Psychometric evaluation of the Risk Appraisal Guide (RAG): A tool for assessing violent recidivism. Journal of Interpersonal Violence, 12, 779 – 793.
- Loza, W. & Green, K. (2003). The Self-Appraisal Questionnaire. Journal of Interpersonal Violence, 18, 781 – 797.
- Loza, W., Loza-Fanous, A. (2000). Predictive validity of the Self-Appraisal Questionnaire. Journal of Interpersonal Violence, 15, 1183 – 1191.
- Loza, W., Villeneuve, D. B. & Loza-Fanous, A. (2002). Predictive validity of the Violence Risk Appraisal Guide: A tool for assessing violent offender's recidivism. International Journal of Law and Psychiatry, 25, 85 – 92.
- Metz, C. E., Herman, B. A. & Shen, J-H. (1998). Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously distributed data. Statistics in Medicine, 17, 1033-1053.
- Mills, J. F. & Kroner, D. G. (2006). The effect of discordance among violence and general recidivism risk estimates on predictive accuracy. Criminal Behaviour and Mental Health, 16, 155 – 166.
- Mossman, D. (1994). Assessing predictors of violence: Being accurate about accuracy. Journal of Consulting and Clinical Psychology, 62, 783-792.
- Nicholls, T. L., Ogloff, J. R. P. & Douglas, K. S. (2004). Assessing risk for violence among male and female civil psychiatric patients: The HCR-20, PCL:SV, and VSC. Behavioral Sciences and the Law, 22, 127 – 158.
- Nuffield, J. (1982). Parole Decision Making in Canada: Research Towards Decision Guidelines. Ottawa: Solicitor General of Canada.
- Quinsey, V. L., Harris, G. T., Rice, M. E. & Cormier, C. A. (1998). Violent Offenders: Appraising and Managing Risk. Washington, DC: American Psychological Association.
- Salekin, R. T., Rogers, R. & Sewell, K. W. (1996). A review and meta-analysis of the psychopathy checklist and psychopathy checklist-revised: Predictive validity of dangerousness. Clinical Psychology: Science and Practice, 203 – 215.
- Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. Law and Human Behavior, 31, 449 – 462.

- Serin, R. C. (1996). Violent recidivism in criminal psychopaths. Law and Human Behavior, 20, 207 – 217.
- Snowden, R. J., Gray, N. S., Taylor, J. MacCulloch, M. J. (2007). Actuarial prediction in violent recidivism in mentally disordered offenders. Psychological Medicine, 37, 1539 – 1549.
- Stadtland, C., Kleindienst, N., Kroner, C., Eidt, M. & Nedopil, N. (2005). Psychopathic traits and risk of criminal recidivism in offenders with and without mental disorders. International Journal of Forensic Mental Health, 4, 89 – 97.
- Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. Psychological Bulletin, 99, 100-117.
- Tengstrom, A. (2001). Long-term predictive validity of historical factors in two risk assessment instruments in a group of violent offenders with schizophrenia. Nordic Journal of Psychiatry, 55, 243 – 249.
- Tengstrom, A., Grann, M., Langstrom, N., Kullgren, G. (2000). Psychopathy (PCL-R) as a predictor of violent recidivism among criminal offenders with schizophrenia. Law and Human Behavior, 24, 45 – 58.
- Thornton, D. Scoring Guide for the Risk Matrix 2000.9/SVC. [http://psg275.bham.ac.uk/forensic_centre/External%20Documents/SCORING%20GUIDE%20FOR%20RISK%20MATRIX\(ver-Feb%202007\).pdf](http://psg275.bham.ac.uk/forensic_centre/External%20Documents/SCORING%20GUIDE%20FOR%20RISK%20MATRIX(ver-Feb%202007).pdf)
- Urbaniok, F., Noll, T., Grunewold, S., Steinbach, J. & Endrass, J. (2006). Prediction of violent and sexual offences: A replication study of the VRAG in Switzerland. Journal of Forensic Psychiatry and Psychology, 17, 23 - 31.
- Webster, C. D., Douglas, K., S., Eaves, D. & Hart, S. D. (1997). HCR-20: Assessing the Risk for Violence (Version 2). Vancouver: Mental Health Law and Policy Institute, Simon Fraser University.
- Webster, C.D. & Hucker, S. (2007). Violence Risk Assessment and Management. Chichester: Wiley.
- Webster, C.D., Martin, M-L., Brink, J., Nicholls, T.L. & Middleton, C. (2004). Short-Term Assessment of Risk and Treatability (START), version 1.0 (consultation edition). Hamilton, Ontario: St Joseph's Healthcare and Forensic Psychiatric Services Commission.
- Wong, S. & Gordon, A. (2001). The violence risk scale. Forensic Update, 67, 19 – 23.
- Wong, S. C. P. & Gordon, A. (2006). The validity and reliability of the violence risk scale: A treatment-friendly violence risk assessment tool. Psychology, Public Policy and Law, 12, 279 – 309.

Technical Appendix For Part 1

Excluded Studies⁷

	Reference	Reason for Exclusion
1	Douglas, K. S., Hart, S. D. & Knopp, P. R. (2001). Validity of the Personality Assessment Inventory for forensic assessments. <i>International Journal of Offender Therapy and Comparative Criminology</i> , 45, 183 – 197.	- Not a prospective investigation.
2	Doyle, M., Dolan, M. & McGovern, J. (2002). The validity of North American risk assessment tools in predicting in-patient violent behaviour in England. <i>Legal and Criminological Psychology</i> , 7, 141 – 154.	- Only institutional violence
3	Fujii, D. E. M., Tokioka, A. B., Lichton, A. I. & Hishinuma, E. (2005). Ethnic differences in prediction of violence risk with the HCR-20 among psychiatric in-patients. <i>Psychiatric Services</i> , 56, 711 – 716.	- Only institutional violence
4	Gagliardi, G. J., Lovell, D., Peterson, P. D. & Jemelka, R. (2004). Forecasting recidivism in mentally ill offenders released from prison. <i>Law and Human Behavior</i> , 28, 133 – 155	- No risk assessment instrument
5	Glover, A. J. J., Nicholson, D. E., Hemmati, T., Bernfield, G. A. & Quinsey, V. L. (2002). A comparison of predictors of general and violent recidivism among high-risk federal offenders. <i>Criminal Justice and Behavior</i> , 29, 235 – 249.	- No AUC available
6	Gray, N. S., Hill, C., McGleish, A., Timmons, D., MacCulloch, M. J., Snowden, R. J. (2003). Prediction of violence and self-harm in mentally disordered offenders: A prospective study of the efficacy of HCR-20, PCL-R, and psychiatric symptomatology. <i>Journal of Consulting and Clinical Psychology</i> , 71, 443 – 451.	- N too low (34)
7	Heilbrun, K., Hart, S. D., Hare, R. D., Gustafson, D., Nunez, C., & White, A. J. (1998). <i>Journal of Interpersonal Violence</i> , 13, 514 – 522.	- No AUC available
8	Mills, J. F., Loza, W., & Kroner, D. G. (2003). Predictive validity despite social desirability: Evidence for the robustness of self-report among offenders. <i>Criminal Behavior and Mental Health</i> , 13, 140 – 150.	- No AUC available
9	Raynor, P. (2007). Risk and need assessment in British probation: The contribution of LSI-R. <i>Psychology, Crime & Law</i> , 13, 125 – 138.	- Only general reoffending
10	Rice, M. & Harris, G. T. (1992). A comparison of criminal recidivism among schizophrenic and nonschizophrenic offenders. <i>International Journal of Law and Psychiatry</i> , 15, 397 – 408.	- No AUC available - No risk assessment instrument
11	Rice, M. & Harris, G. T. (1995). Psychopathy, schizophrenia, alcohol abuse, and violent recidivism. <i>International Journal of Law and Psychiatry</i> , 18, 333 – 342.	- No AUC available
12	Simourd, D. J. (2004). Use of dynamic risk/need assessment instruments among long-term incarcerated offenders. <i>Criminal Justice and Behavior</i> , 31, 306 – 323.	- Only general reoffending
13	Simourd, D. J. & Van De Ven, J. (1999). Assessment of criminal	

⁷ 1990 onwards

	attitudes: Criterion-related validity of the criminal sentiments scale-modified and pride in delinquency scale. <i>Criminal Justice and Behavior</i> , 26, 90 – 106.	- No AUC available
14	Strand, S., Belfrage, H., Fransson, G. & Levander, S. (1999). Clinical and risk management factors in risk prediction of mentally disordered offenders – more important than historical data? <i>Legal and Criminological Psychology</i> , 4, 67 – 76.	- Not predictive study
15	Villeneuve, D. B., Oliver, N., Loza, W. (2003). Cross-validation of the Self-Appraisal Questionnaire with a maximum-security psychiatric population. <i>Journal of Interpersonal Violence</i> , 18, 1325 – 1334.	- No AUC available
16	Walters, G. D. (2006). Risk-appraisal versus self-report in the prediction of criminal justice outcomes: A meta-analysis. <i>Criminal Justice and Behavior</i> , 3, 279 – 304.	- Review article
17	Loza, W. & Loza-Fanous, A. (2000). Predictive Validity of the Self-Appraisal Questionnaire. <i>Journal of Interpersonal Violence</i> , 15, 1183 – 1191.	- No AUC available
18	Loza, W. & Loza-Fanous, A. (2002). The effectiveness of the Self-Appraisal Questionnaire as an offender; classification measure. <i>Journal of Interpersonal Violence</i> , 17, 3 – 13.	- Only institutional violence
19	Loza, W. & Loza-Fanous, A. (2001). The effectiveness of the Self-Appraisal Questionnaire in predicting offenders' post-release outcome: A comparison study. <i>Criminal Justice and Behavior</i> , 28, 105 – 121.	- No AUC available
20	Loza, W. & Loza-Fanous, A. (2003). More evidence for the validity of the Self-Appraisal Questionnaire for predicting violent and nonviolent recidivism. <i>Criminal Justice and Behavior</i> , 30, 709 – 721.	- No AUC available
21	Mills, J. F. , Jones, M. N. & Kroner, D. G. (2005). An examination of the generalizability of the LSI-R and VRAG probability bins. <i>Criminal Justice and Behavior</i> , 32, 565 – 585.	- Only previously published data
22	Catchpole, R. E. H. & Gretton, H. M. (2003). The predictive validity of risk assessment with violent young offenders. <i>Criminal Justice and Behavior</i> , 30, 688 - 708.	- Young offenders
23	Belfrage, H., Fransson, G. & Strand, S. (2000). Prediction of violence using the HCR-20: A prospective study in two maximum security correctional institutions. <i>Journal of Forensic Psychiatry</i> , 11, 167 – 175.	- Only institutional violence
24	Daffern, M., Ogloff, J. R.P., Ferguson, M. & Thomson, L. (2005). Assessing risk for aggression in a forensic psychiatric hospital using the Level of Service Inventory-Revised:Screening Version. <i>International Journal of Forensic Mental Health</i> , 4, 201 – 206.	- Only institutional violence
25	Dernevik, M., Grann, M. & Johansson, S. (2002). Violent behaviour in forensic psychiatric patients: Risk assessment and different risk-management levels using the HCR-20. <i>Psychology, Crime & Law</i> , 93 – 111.	- Only institutional violence
26	Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Farrington, D. P., Rogers, & R. D. (2008). Most items in risk assessment instruments do not predict violence. Unpublished Manuscript.	- Item analysis
27	Coid, J., Yang, M., Ullrich, S., Zhang, T., Roberts, A., Roberts, C., Rogers, R. D. & Farrington, D. (2007). <i>Predicting and Understanding Risk of Re-offending: The Prisoner Cohort Study</i> . (Research Summary 6). London: Ministry of Justice.	- Only previously published data
28	McNeil, D., Gregory, A., Lam, J., Binder, R. & Sullivan, G. (2003). Utility of decision support tools for assessing acute risk of violence. <i>Journal of Consulting and Clinical Psychology</i> , 71, 945 – 953.	- Not blind to outcome. (selection of viol v nonviol
29	De Vogel, V. & de Ruiter (2005). The HCR-20 in personality	

	disordered female offenders: A comparison with a matched sample of males. <i>Clinical Psychology and Psychotherapy</i> ,12, 226 – 240	- N too low (42 males)
30	Fujii, D., Lichton, A. & Tokioka, A. (2007). Structured professional judgement versus actuarial data in violence risk prediction using the Historical Clinical Risk Management – 20. (under review)	- Only institutional violence
31	Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. <i>Criminal Justice and Behavior</i> , 32, 479-510.	- Not predictive study
32	Harris, G. T., & Rice, M. E. (2007). Adjusting actuarial violence risk assessments based on aging or the passage of time. <i>Criminal Justice and Behavior</i> , 34, 297-313.	- Only previously published data
33	Hemphill, J. F., & Hare, R. D. (2004). Some misconceptions about the Hare PCL-R and risk assessment: A reply to Gendreau, Goggin, and Smith. <i>Criminal Justice and Behavior</i> , 31, 203-243.	- Review article
34	Kroner, D. G., Mills, J. F., Reitzel, L. R., Dow, E., Aufderheide, D. H., & Railey, M. G. (2007). Directions for violence and sexual risk assessment in correctional psychology. <i>Criminal Justice and Behavior</i> , 34, 906- 918.	- Review Article
35	Skeem, J. L. Mulvey, E. P., Appelbaum, P., Banks, S., Grisso, T., Silver, E., & Clark Robbins, P. (2004). Identifying subtypes of civil psychiatric patients at high risk for violence. <i>Criminal Justice and Behavior</i> , 31, 392-437.	- No risk assessment instrument
36	Tengström, A., Hodgins, S., Grann, M., Långström, N., & Kullgren, G. (2004). Schizophrenia and criminal offending: The role of psychopathy and substance use disorders. <i>Criminal Justice and Behavior</i> , 31, 367-391.	- Not a predictive study
37	Gardner W, Lidz CW, Mulvey EP, & Shaw EC. (1996). Clinical versus actuarial predictions of violence in patients with mental illnesses. <i>Journal of Consulting and Clinical Psychology</i> . 64, 602-609.	- No risk assessment instrument
38	Menzies R.,Webster CD. (1995). Construction and Validation of Risk Assessments in a 6-Year Follow-up of Forensic Patients - a Tridimensional Analysis. <i>Journal of Consulting and Clinical Psychology</i> .63, 766-778.	- No risk assessment instrument
39	Skeem J.L. & Mulvey E.P. (2001). Psychopathy and Community Violence Among Civil Psychiatric Patients: Results From the MacArthur Violence Risk Assessment Study. <i>Journal of Consulting and Clinical Psychology</i> . 69, 358-374.	- Only previously published data
40	Gretton, H., Hare, R. D., & Catchpole, R. (2004). Psychopathy and offending from adolescence to adulthood: A ten-year follow-up. <i>Journal of Consulting and Clinical Psychology</i> , 72, 636-645.	- Young offenders
41	Ferris LE, Sandercock J, Hoffman B, Silverman M, Barkun H, Carlisle J et al. (1997). Risk assessments for acute violence to third parties: A review of the literature. <i>Canadian Journal of Psychiatry</i> , 42,1051-1060.	- Review article
42	Gardner W, Lidz CW, Mulvey EP, Shaw EC. (1996). A comparison of actuarial methods for identifying repetitively violent patients with mental illnesses. <i>Law and Human Behavior</i> . 20, 35-48.	- Review article
43	Leistico, A. R., Salekin, R. T., DeCoster, J., & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. <i>Law and Human Behavior</i> , 32, 28-45.	- Review article
44	Johnston S.J. (2002). Risk assessment in offenders with intellectual disability: the evidence base. <i>Journal of Intellectual</i>	- Review article

- Disability Research*. 46, 47-56.
- 45 Skilling TA, Harris G.T., Rice M.E., & Quinsey V.L. (2002). Identifying persistently antisocial offenders using the Hare psychopathy checklist and DSM antisocial personality disorder criteria. 8. *Psychological Assessment*. 14, 27-33. - Not a predictive study.
- 46 Guy, L. S., & Douglas, K. S. (2006). Examining the utility of the PCL:SV as a screening measure using competing factor models of psychopathy. *Psychological Assessment*, 18, 225-230. - Not a predictive study
- 47 Loza, W. & Dhaliwal GK. (1997). Psychometric evaluation of the risk appraisal guide (RAG) - A tool for assessing violent recidivism. *Journal of Interpersonal violence*.12:779-793. - Not a predictive study
- 48 Mills, J.F. & Kroner, D.G. (1997). The Criminal Sentiments Scale: Predictive validity in a sample of violent and sex offenders. *Journal of clinical psychology*. 53:399-404. - No AUC available
- 49 Dolan, M. & Doyle, M. (2000). Violence risk prediction: Clinical and actuarial measures and the role of the Psychopathy Checklist. *British Journal of Psychiatry*, 177, 303-311. - Review article.
- 50 Monahan, J., Steadman, H.J., Appelbaum, P.S., Robbins, P.C., Mulvey, E.P., Silver, E., et al. (2000). Developing a clinically useful actuarial tool for assessing violence risk. *British Journal of Psychiatry*, 176, 312-319. - No AUC available.
- 51 Abushua'leh, K., & Abu-Akel, A. (2006). Association of psychopathic traits and symptomatology with violence in patients with schizophrenia. *Psychiatry Research*, 143, 205-211. - Not predictive study.
- 52 Douglas, K. S., Strand, S., Belfrage, H., Fransson, G., & Levander, S. (2005). Reliability and validity evaluation of the Psychopathy Checklist: Screening Version (PCL: SV) in Swedish correctional and forensic psychiatric samples. *Assessment*, 12, 145-161. - Not predictive study.
- 53 Edens, J. F., & Cahill, M. A. (2007). Psychopathy in adolescence and criminal recidivism in young adulthood: Longitudinal results from a multiethnic sample of youthful offenders. *Assessment*, 14, 57-64. - Young offenders
- 54 Grann, M., Långström, N., Tengström, A., & Stalenheim, E. G. (1998). Reliability of file-based retrospective ratings of psychopathy with the PCL-R. *Journal of Personality Assessment*, 70, 416-426. - Not predictive study
- 55 Hare, R. D., Clark, D., Grann, M., & Thornton, D. (2000). Psychopathy and the predictive validity of the PCL-R: An international perspective. *Behavioral Sciences and the Law*, 18, 623-645. - Review article.
- 56 Kroner, D. G., Mills, J. F., & Reddon, J. R. (2005). A coffee can, factor analysis, and prediction of antisocial behavior: The structure of criminal risk. *International Journal of Law and Psychiatry*, 28, 360-374. - Only general reoffending.
- 57 Laurell, J., & Dåderman, A. M. (2005). Recidivism is related to psychopathy (PCL-R) in a group of men convicted of homicide. *International Journal of Law and Psychiatry*, 28, 255-268. - Not predictive study.
- 58 Sreenivasan, S., Kirkish, P., Eth, S., Mintz, J., Hwang, S., van Gorp, W., & Van Vort, W. (1997). Predictors of recidivistic violence in criminally insane and civilly committed psychiatric inpatients. *International Journal of Law and Psychiatry*, 20, 279-291. - Not predictive study.
- 59 Austin, J., Coleman, D., Peyton, J, & Johnson, K. D. (2003). *Reliability and Validity Study of the LSI-R Risk Assessment Instrument*. The Institute on Crime, Justice and Corrections. George Washington University. Washington, DC. - Only general reoffending.
- 60 Mills, J. F., Kroner, D. G. & Hemmati, T. (2003). Predicting violent

	behavior through a static-stable variable lens. <i>Journal of Interpersonal Violence</i> , 18, 891 – 904.	- Not predictive study.
61	Lovell, D., Gagliardi, G. & Phipps, P. (2005). <i>Washington's dangerous mentally-ill offender law: Was community safety increased?</i> (Report #05-03-10901). Olympia, WA: Washington State Institute of Public Policy.	- No AUC available
62	Firestone, P., Bradford, J. M., Greenburg, D. M., Curry, S. & Larose, M. R. (1998). Predictions of recidivism in extrafamilial child molesters based on court related assessments. <i>Sexual Abuse: A Journal of Research and Treatment</i> , 12, 203 – 222.	- Sex offenders
63	Firestone, P., Bradford, J. M., McCoy, M., Greenburg, D. M., Curry, S. & Larose, M. R. (1998). Recidivism factors in convicted rapists. <i>Journal of the American Academy of Psychiatry and the Law</i> , 26, 185 – 200.	- Sex offenders
64	Firestone, P., Bradford, J. M., Greenburg, D. M. & Larose, M. R. (1998). Prediction of recidivism in incest offenders. <i>Journal of Interpersonal Violence</i> , 14, 511 – 531.	- Sex offenders
65	Forth, A. E., Hart, S. D. & Hare, R. D. (1990). Assessment of psychopathy in male young offenders. <i>Psychological Assessment</i> , 2, 342 – 344.	- Young offenders
66	Harris, G. T., Rice, M. E. & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. <i>Criminal Justice and Behavior</i> , 20, 315 – 335.	- No AUC available
67	Ilacqua, G. E., Coulson, G. E., Lombardo, D. & Nutbrown, V. (1999). Predictive validity of the Young Offender Level of Service Inventory for criminal recidivism of male and female young offenders. <i>Psychological Reports</i> , 84, 1214 – 1218.	- Young offenders
68	Kirkpatrick, B. L. (1999). Exploratory research of female risk prediction and LSI-R. <i>Correctional Compendium: The National Journal for Corrections</i> . 24, 1 – 3.	- Female offenders
69	Lowenkamp, C. T., Holsinger, A. M. & Latessa, E. J. (2001). Risk/needs assessment, offender classification, and the role of childhood abuse. <i>Criminal Justice and Behavior</i> , 28, 543 – 563.	- Not predictive study
70	Quinsey, V. L., Rice, M. E. & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. <i>Journal of Interpersonal Violence</i> , 10, 85 – 105.	- Sex offenders
71	Shields, I. (1993). The use of the Young Offender – Level of Service Inventory (YO-LSI) with adolescents. <i>IARCA Journal on Community Corrections</i> , 5, 10 – 26.	- Young offenders
72	Toupin, J., Mercier, H., Dery, M., Cote, G. & Hodgins, S. (1995). The validity of the PCL-R for adolescents. <i>Issues in Criminological and Legal Psychology</i> , 24, 143 – 145.	- Young offenders
73	Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. <i>Criminal Justice and Behavior</i> . 28, 490-521.	- Sex offenders
74	Beech, A., Friendship, C., Erikson, M., & Hanson, R.K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of UK child abusers, <i>Sexual Abuse: A Journal of Research and Treatment</i> , 14, 155-167.	- Sex offenders
75	Bonta J, Law M, Hanson K. 1998. The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. <i>Psychological Bulletin</i> . 123:123-42.	- Review article
76	Douglas, K.S. & Webster, C.D. (1999). The HCR-20 violence risk assessment scheme - Concurrent validity in a sample of incarcerated offenders. <i>Criminal Justice and Behavior</i> , 26, 3-19.	- Not predictive study
77	Ferguson, G. E., Eidelson, R. J., & Witt, P.H. (1998). New	

	Jersey's sex offender risk assessment scale: Preliminary validity data. <i>Journal of Psychiatry and Law</i> , 26, 327-51.	- Sex offenders
78	Fuller, J. & Cowan, J. (1999). Risk assessment in a multi-disciplinary forensic setting: clinical judgement revisited. <i>Journal of Forensic Psychiatry</i> , 10, 276-89.	- No risk assessment instrument
79	Hood, R. (2002). <i>Reconviction Rates of Serious Sex Offenders and Assessments of their Risk</i> . (Home Office Research Findings No 164) London, Home Office .	- No risk assessment instrument.
80	Loza W, Dhaliwal G, Kroner DG, Loza-Fanous A. 2000. Reliability, Construct, and Concurrent Validities of the Self-Appraisal Questionnaire: A Tool for Assessing Violent and Non-violent Recidivism. <i>Criminal Justice and Behavior</i> ; 27, 356-74.	- No AUC available
81	Rice, M.E. & Harris, G.T. (1997). Cross-validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. <i>Law and Human Behavior</i> , 21, 231-241.	- Sex offenders.
82	Salekin, R.T., Rogers, R., & Sewell, K.W. (1996). A review and meta-analysis of the psychopathy checklist and psychopathy checklist-revised: Predictive validity of dangerousness. <i>Clinical Psychology: Science and Practice</i> , 3, 203-215.	- Review article
83	Sjostedt, G & Langstrom, N. (2002). Assessment of risk for criminal recidivism among rapists: A comparison of four different measures. <i>Psychology Crime and Law</i> , 8, 25-40.	- Sex offenders
84	Hemphill, J. F., Hare, R. D. & Linden, W. (1998). Psychopathy and recidivism: A review. <i>Legal and Criminological Psychology</i> , 3, 139 – 170.	- Review article.
85	Hare, R. D. (1999). Psychopathy as a risk factor for violence. <i>Psychiatric Quarterly</i> , 70, 181-197.	- Review article
86	Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. <i>Legal and Criminological Psychology</i> , 3, 121-137.	- Review article
87	Hemphill, J. F., Hare, R. D., & Wong, S. (1998). Psychopathy and recidivism: A review. <i>Legal and Criminological Psychology</i> , 3, 139-170.	- Review article
88	Långström, N., & Grann, M. (2002). Psychopathy and violent recidivism among young criminal offenders. <i>Acta Psychiatrica Scandinavica</i> , 106(Suppl. 412), 86-92.	- Young offenders
89	Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. <i>Psychology, Public Policy, and Law</i> , 7, 409-443.	- Review article
90	Loza, W., Conley, M., & Warren, B. (2004). Concurrent cross-validation of the Self-Appraisal Questionnaire: A tool for assessing violent and nonviolent recidivism and institutional adjustment on a sample of North Carolina offenders. <i>International Journal of Offender Therapy and Comparative Criminology</i> , 48, 85-95.	- Not a predictive study
91	Loza, W., Cumbleton, A., Shahinfar, A., Neo, L. H., Evans, M., Conley, M., & Summers, R. (2004). Cross-validation of the Self-Appraisal Questionnaire (SAQ): An offender risk and need assessment measure on Australian, British, Canadian, Singaporean, and American offenders. <i>Journal of Interpersonal Violence</i> , 19, 1172-1190.	- Not a predictive study
92	Loza, W. & Loza-Fanous, A. (1999). Anger and prediction of violent and nonviolent offender's recidivism. <i>Journal of Interpersonal Violence</i> , 14, 1014-1028.	- Not a study of risk assessment instrument
93	Tengström, A., & Hodgins, S. (2002). Criminal behavior of forensic and general psychiatric patients with schizophrenia: Are they different? <i>Acta Psychiatrica Scandinavica</i> , 106 (Suppl. 412), 62-66.	- No AUC available

94	Walters, G. D. (2003). Predicting criminal justice outcomes with the Psychopathy Checklist and Lifestyle Criminality Screening Form: A meta-analytic comparison. <i>Behavioral Sciences and the Law</i> , 21, 89-102.	- Review article
95	Walters, G. D. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factor scores: A meta-analysis. <i>Law and Human Behavior</i> , 27, 541-558.	- Review article
96	Harris, G. T., Rice, M. E. & Camilleri, J. A. (2004). Applying forensic actuarial assessment (the Violence Risk Appraisal Guide) to nonforensic patients. <i>Journal of Interpersonal Violence</i> , 19, 1063 – 1074.	- Only previous published data.

PART 2. OPINIONS OF PRACTITIONERS

Background: Both the professional and research literatures have articulated standards, principles and guidelines for carrying out violence risk assessments. This study examined the extent to which the tools currently used by the various multi-agencies across Scotland are perceived by users as meeting these criteria. This study therefore acknowledges that having predictive items is only part of the process.

Methods: Questionnaire and interview methodology was used to obtain data – quantitative and qualitative – on users' views on a set of standards and principles articulated in the extant literature. Over 300 multi-agency, multi-disciplinary practitioners involved in the practice of violence risk assessment were invited to take part. Twenty-nine practitioners responded. Participants represented criminal justice social work, forensic mental health, the Scottish Prison Service, the Parole Board for Scotland, multi-agency public protection arrangements, and academia.

Results: The results showed that a range of risk assessment tools were used across Scotland. Actuarial and structured clinical judgement approaches were utilised. The most commonly used tool was the RAG-F followed by the HCR-20, then the LSI-R and/or LSCMI,⁸ and then psychopathy assessments. None of the tools met every standard and/or principle in risk assessment but the HCR-20 was rated as the most satisfactory instrument and, despite its widespread use, the RAG-F was rated most poorly.

Discussion: The practice of violence risk assessment across Scotland is not uniform: different agencies use different approaches. The level of satisfaction and perceived utility of the different instruments is varied. The implications of the findings for violence risk assessment in Scotland are discussed.

⁸ Because the LS/CMI is not widely used, the responses usually concerned the LSI-R.

Introduction

There are many tools purporting to assess risk for violence. An overview of each of the available instruments is provided in the Risk Management Authority (RMA) Scotland publication entitled Risk Assessment Tools Evaluation Directory (RATED) (RMA Scotland, 2006, 2007). Risk assessment tools generally fall into one of two categories: (1) the structured clinical judgment approach refers to a multi-modal, multi-informant assessment of a minimum set of risk factors derived from research evidence and best practice principles. The risk factor ratings are used by expert raters to organise and structure their professional judgement about the presence and relevance of the risk factors; (2) the actuarial approach refers to an assessment of a set of risk factors based on statistically derived models to predict the likelihood of reoffending. In this category there is a further approach termed the ‘adjusted actuarial’ approach. This refers to a model of risk assessment whereby actuarial tools are used as the basis for judgments but the conclusions are then adjusted according to practitioner opinions.

In the Scottish context, it is not clear if and how different agencies and professionals have come to rely more or less on one specific model of risk assessment– actuarial or structured clinical judgment. Part of the remit of the RMA is to promote best practice in violence risk assessment; the RMA exists to “provide a centre of excellence in risk assessment and risk management, enabling and promoting best practice and regulating the delivery of services to help manage and minimise the risk of serious harm caused by sexual and violent offending” (RMA Scotland, 2006). The RMA has therefore developed standards and guidelines for conducting risk assessments (RMA Scotland, 2006) and for producing and implementing risk management plans (RMA Scotland, 2007). In devising their guidelines, the RMA has acknowledged that the task of producing an appropriate risk evaluation and

management intervention is complex and thus the RMA guidance is detailed and comprehensive. Key stages involve the selection of appropriate tools, gathering a comprehensive information source, analysing the information and rating risk factors, formulating the risks that are present and using all of the information to delineate risk management strategies.

At the present time, what remains untested is the extent to which the current models of risk assessment used by the various multi-agencies in Scotland are perceived by users as conforming to best practice and if and how they facilitate the entire process of risk assessment. As indicated in Part 1, our main research focus has been on investigating the predictive utility of a tool. While this is important to justify the inclusion of a risk factor or set of risk factors in an instrument, this is not the whole picture. There are problems with using group-based studies to understand individual risk and also problems such as wide confidence intervals associated with actuarial tools (see Hart, Michie, & Cooke, 2007). Thus, the purpose of the present study was to examine the extent to which the tools currently used by the various multi-agencies across Scotland are considered satisfactory by multi-agency, multi-disciplinary practitioners.

Methods

In order to ascertain the extent to which the currently used risk assessment tools are perceived as achieving best practice in risk assessment we considered the work of the RMA Scotland (2006) and other authoritative writings in the field of risk assessment (e.g. Webster et al., 1997; Hart et al., 2003) to articulate a set of standards and best practice principles for risk assessment tools. These are described below.

Standards and Principles

1. User Manual

It has been argued that a coherent, descriptive and effective manual should increase reliability, transparency and validity in risk assessment (Hare et al., 2003). The rationale for this may be based on findings from the offender treatment literature where it has been shown that the existence of a programme manual promotes integrity in practice where integrity refers to the process whereby “the programme [or intervention] is conducted in practice as intended in theory and design” (Hollin, 1995, p. 196). The same principles are likely to apply in risk assessment, although, we believe that this has not been empirically tested.

2. Specified User Qualifications, Competencies and Skill Levels

The second issue is the need for specified user qualifications and competencies. As with all interventions, users must work within their own professional and experiential competencies. Given the complex nature of violence risk assessment, and the difficult clinical decisions that are involved in this process, risk assessment tools should be explicit about necessary user qualifications and competencies.

3. Accessible Training

Risk assessment requires complex and varied skills, and familiarity with guidelines and theory. In order to ensure the appropriate use of tools, training needs to be accessible.

4. Ease of Use

Instruments that are easy to understand, apply and use are desirable. Ease of use is therefore an important aspect of any instrument. Ease of use refers to the appropriateness of language, application and procedure of assessment.

5. Uses a Comprehensive Information Source

It is important that multiple sources of information are used in order to ensure a comprehensive assessment. Information may come from background files, interviews and specialist assessments with the client, collateral interviews with relatives, acquaintances and

(where appropriate) victims, formal assessments, etc. A comprehensive information source is therefore a key part of a good risk assessment.

6. Appropriate Administration Time

Given the variety of problems and backgrounds that offenders can present, we considered that risk assessment tools should be appropriate in terms of their administration time i.e., be responsive to the complexity (or straightforward nature) of the case. We considered that it was not possible to be prescriptive about a specified time that risk assessment should require; we were more concerned with user perceptions about the appropriateness of administration time and therefore included this as an important aspect of violence risk assessment.

7. Social History

In order to provide a context to offender risk and risk assessment, it is important that social background is examined. The RMA Scotland has articulated the standard whereby evaluators “must review a range of relevant documents concerning the offender’s social, criminal and medical context in order to inform their assessment” (RMA Scotland, 2006, p.16). It is therefore important in the practice of risk assessment that a comprehensive social history is obtained.

8. Offence Analysis

The RMA Scotland has stated that, in a risk assessment report, there should be “a detailed analysis of offending behaviour. Assessors should cover both current and past offending and set out the criminogenic factors specific to the offender...This section must include an analysis of why the offender has committed his or her offences, and will obviously draw from the offender background information...The style of this section will be analytical rather than descriptive” (RMA Scotland, 2006, p.30). For the purposes of our study, we

considered it important for risk assessment tools to facilitate a detailed evaluation of the person's offence analysis.

9. Appropriate Number of Risk Factors

It is important that the number of risk factors included in the tool is appropriate. Risk assessments that include too many risk factors may be cumbersome and difficult to administer whereas those with too few may fail to facilitate a comprehensive assessment. It is therefore desirable that the number of risk factors is a good fit to the complexity of the case and allows for a thorough assessment but does not result in a large amount of redundancy.

10. Appropriateness of Emphasis on Static Factors

In risk assessment, static risk factors refer to those features of a person's past functioning and behaviour that are historical in nature and therefore unchanging or only slowly changing (e.g., previous violence, young age at first violence, history of substance use, relationship instability). It has been demonstrated in the literature that static, historical variables have a significant relationship with future violence and so it is important that these risk factors are considered in violence risk assessment. For the purposes of this study, we therefore specified that there should be an appropriate emphasis on static factors.

11. Appropriateness of Emphasis on Dynamic Factors

It is important to consider dynamic risk factors in risk assessment. Dynamic factors can be stable or acute. The RMA provided the following definition of stable dynamic risk factors. "Stable dynamic risk factors refer to those areas of an individual's circumstances that are enduring over a period of time and contribute to the risk of further offending, over months or years, but are amenable to change and, if changed, may be expected to reduce the likelihood or seriousness of further offending, for example, a pattern of excessive alcohol use over the past year would be a stable dynamic risk factor" (RMA, 2007, p. 33). Acute dynamic risk

factors are defined by the RMA Scotland as “those elements that change quickly, perhaps over days or hours, and whose emergence indicates a period of critical risk in which serious offending is more likely than not to occur. Examples include an increase in level of substance misuse, heightened emotional distress, or environmental changes that increase the potential for offending. So while alcohol use may be a stable dynamic risk factor, intoxication may be an acute dynamic risk factor” (p. 33). Evidently, dynamic risk factors are important in determining current and future risk. Risk assessment instruments should have an appropriate emphasis on dynamic risk factors.

12. Well-defined Risk Factors

It is likely that the validity and reliability of a risk factor will be dependent on the clarity and consistency of the definition. Well-defined risk factors that are easy to understand are therefore important.

13. Ease of Scoring/Rating Procedures

Scoring procedures should be easy to understand and follow so that they can be reliable across different assessors and across time.

14. Describes and Identifies Protective Factors

There is a growing awareness in the practice of risk assessment and risk management that, as well as identifying problem risk factors, protective factors should also be identified and included in the evaluation. The RMA Scotland has provided a useful definition of protective factors, i.e., “Protective factors are aspects of an individual’s personal, inter-personal and environmental context that diminish or inhibit the occurrence of further offending. To be considered as contributing to the minimisation of risk, protective factors require to be present and active...The absence of a risk factor is not necessarily in itself a protective factor, for example, the absence of an anti-social peer group is the absence of a

risk factor, whereas the presence of an actively pro-social peer group may be a protective factor” (p. 22). The ability of the tools to describe and identify protective factors is an important feature of risk assessment.

15. Assists in Risk Formulation

Since the purpose of risk assessment is to facilitate risk management and minimisation, it is important to consider the ability of the various tools to inform the risk management plan. Risk formulation is a key stage in informing risk management. Risk formulation has been defined by RMA Scotland as “a working model or explanation of an individual’s risk in key areas (e.g. sexual violence, domestic violence) based on what are believed to be the most relevant risk and protective factors for the behaviours to be prevented in the future. It also importantly considers the interaction of those risk and protective factors to explain the risk of an individual” (RMA Scotland, 2007, p.34). For the purposes of our research, we considered the utility of a risk tool to facilitate risk formulation as being an important feature.

16. Assists in Identifying Plausible Risk Scenarios

Scenario planning is the process of risk assessment whereby the evaluator considers all the relevant information in order to identify the most plausible future eventualities of risk. In the RMA guidelines on risk management (RMA Scotland, 2007) and other up-to-date risk assessment guides (e.g. Risk for Sexual Violence Protocol, Hart et al., 2003) evaluators are prompted to consider risk scenarios. A repeat scenario is said to occur when the offender commits a similar offence to the current or most recent incident; an improvement scenario is where the violence decreases and the person engages in a less serious act; a worst case scenario is where the violence increases and the offence committed is of a more serious nature. In each scenario, facets of the violence such as likelihood, nature, severity,

imminence, frequency and duration are considered. It has been postulated that, by considering these more detailed aspects of plausible risk scenarios, risk management interventions are more likely to be targeted appropriately. We therefore included the ability of a risk assessment instrument to help identify plausible risk scenarios as an important aspect of the risk assessment process.

17. Assists in the Development of a Risk Management Plan

The risk and protective factors assessed should meaningfully inform the risk management plan and should not be focused merely on predictor variables. For example, being male is a statistically strong correlate of violence but clearly not amenable to any form of risk management whereas substance use, also a significant predictor, is potentially manageable via treatment, supervision, monitoring etc. We therefore included the need to assist in the development of a risk management plan as an important principle.

18. Assists in the Identification of Treatment Needs for Risk Management

The purpose of risk management is not only to contain the risk but also to minimise it. Therefore, risk factors that are amenable to treatment should be targeted for intervention. We therefore considered it appropriate to evaluate tools in accordance with their ability to inform treatment needs.

19. Assists in the Identification of Monitoring Needs for Risk Management

Similar to 18 above, risk management involves monitoring, and therefore the risk tool should helpfully and meaningfully inform the monitoring needs of the individual. Risk assessment tools were rated according to their ability to assist in identifying monitoring needs.

20. Concordance of Conclusions with Other Risk Assessment Tools/Guidelines

If risk assessment evaluations are valid, they should have concordant findings with other procedures designed to assess the same risk. We therefore considered the extent to which users considered that conclusions reached by the various tools showed agreement with other procedures.

21. Confidence in Risk Judgments

Risk assessments should aid the practitioner in reaching an accurate decision about the person's risk and presentation. A tool which causes a lack of confidence and uncertainty is not likely to be useful.

Questionnaires and Interviews

In order to obtain relevant information from practitioners regarding their perceptions of the extent to which the tools they used achieved these standards and principles, two different methods were employed; a questionnaire and/or an interview. The rationale for selecting the questionnaire methodology was that this approach offers a time effective and economic method of data collection given that questionnaires can be distributed widely and returned easily via electronic mail systems. Because we were not aware of a questionnaire that was appropriate for the purposes of the study, we developed a specific questionnaire (see Technical Appendix for Part 2). This questionnaire included questions with rating options and space for any qualitative information. The content of this questionnaire was deliberately focused on the functionality of the tools as opposed to the experience, background (etc.) of the rater.

The questionnaire and letter of invitation and information regarding the study was sent to potential participants. In order to sample widely, the questionnaire was sent electronically as a Microsoft Word attachment with a covering letter explaining the purpose of the research

(see Technical Appendix) to a contact person at the following agencies with a request that the questionnaire be forwarded to all relevant personnel:

1. All Heads of Department in each of the 8 Criminal Justice Authorities across Scotland.
2. All MAPPA Co-ordinators (Multi-Agency Public Protection Authorities) (N = 8) across Scotland.
3. All clinicians and practitioners (in psychiatry, psychology, nursing, occupational therapy, and social work) employed within the Directorate of Forensic Mental Health, NHS Greater Glasgow and Clyde (N= approximately 250). This service is the only setting in Scotland that cares for patients needing medium, low and out-patient care for mentally disordered offenders (adults, learning disabilities, and female).
4. The Professional Lead for Psychology Services in the Scottish Prison Service (SPS) and all forensic psychologists working in SPS (N=5). The Lead person is responsible for overseeing and leading on violence risk assessment practices in all of the SPS Institutions.
5. The Scottish Clinical Forensic Psychologist Group (N= 45). This group includes all clinical psychologists working in forensic settings (such as community, prison, secure health care, etc.) who would typically have risk assessment as part of their routine duties.

Key practitioners and researchers who did not complete and return questionnaires were approached by one of the researchers and invited to take part in an interview which was based on the questionnaire format. The following people were contacted: the Professional Lead for risk assessment in the Scottish Prison Service (SPS), the Psychiatric Advisor to the First Minister, and a professor of clinical and forensic psychology who is also an international expert on the practice of violence risk assessment. Two interviews were conducted: one

assessment was completed over the phone (Lead for risk assessment in SPS) and the other (professor of clinical forensic psychology) via a face-to-face interview. The Psychiatric Advisor to the First Minister was not available in time to arrange for an interview prior to the research study submission date.

Participants

It was impossible to be fully accurate in determining how many participants were invited to complete questionnaires. However, it is believed that, even if the questionnaires were not distributed widely at each of the organisations, a minimum of 310 people were invited to take part. According to the information received at each setting, the questionnaires were forwarded by contact persons to their colleagues. For example, according to the information available, the Heads of Department at the Criminal Justice Authorities sent group emails to their staff with the letter of invitation and questionnaire attached. The Professional

Table 1: Participants

Organisation/Location	Designations	Number/Percentage
Criminal Justice Social Work	Senior Social Worker	3 (10%)
	Social Worker	4 (14%)
	Lead Officer	1 (3%)
	Team Leader	4 (14%)
	Service Manager	1 (3%)
	TOTAL	44%
Forensic Mental Health	Consultant Forensic Psychiatrist	1 (3%)
	Professional Lead for Psychology Services	1 (3%)
	Consultant Clinical Forensic Psychologist	1 (3%)
	Consultant Clinical Psychologist	1 (3%)
	Forensic Mental Health Nurse/Ward Manager	5 (17%)
	Lead Nurse in Risk Assessment	1 (3%)
	Occupational Therapist	1 (3%)
	TOTAL	35%
Parole Board	Consultant Clinical Forensic Psychologist	1 (3%)
TOTAL	3%	
Scottish Prison Service	Professional Lead for Psychology Services (N=1)	1 (3%)
TOTAL	3%	
Academic	Professor of Clinical Forensic Psychology (N=1)	1 (3%)
TOTAL	3%	
MAPPA	MAPPA Co-ordinators	1 (3%)
TOTAL	3%	
Other/Not Stated	Other/Not Stated	1 (3%)
TOTAL	3%	

Lead for Psychology Services (SPS) confirmed that he had sent a group email to 5 chartered forensic psychologists responsible for undertaking violence risk assessments in their

respective institutions. Of the total questionnaires distributed, 29 (approximately 9%) were returned. This response rate is comparable with typical questionnaire-based studies. Questionnaires were returned by representatives from each of the key agencies and professionals involved in violence risk assessment. The sample composition is shown in Table 1.

Results

In order to address the research questions, the results were analysed to determine: (1) the type and frequency of risk assessment tools used across the relevant agencies within Scotland and; (2) the extent to which the tools were perceived by users as achieving the standards and principles listed above. The findings are described below.

(1) Type and frequency of tools used

The data indicated that the following tools were used to conduct violence risk assessments⁹: (1) the Historical Clinical Risk 20 (HCR-20; Webster, Douglas, Eaves & Hart, 1997); (2) the Risk Assessment Guide Framework incorporating RA1-4 (RAG-F; Scottish Executive, 2000), (3) the Psychopathy Checklist Revised (PCL-R, Hare, 1991; 2002), (4) the Psychopathy Checklist Screening Version (PCL:SV, Hart, Cox & Hare, 1995); (5) the Level of Service Inventory (LSI-R, Andrews & Bonta, 1995), (6) the Level of Service Case Management Inventory (LS/CMI, Andrews, Bonta & Wormith, 2004), (7) the Violence Risk Appraisal Guide, (VRAG, Quinsey et al., 1998), (8) the Initial Risk of Harm Test (unknown), (9) the Violence Risk Scale (VRS, Wong & Gordon, 2000) and (10) the Dynamic Risk Assessment and Management System (DRAMS, Lindsay et al., 2004). A brief overview of the instruments is given below in Table 2.

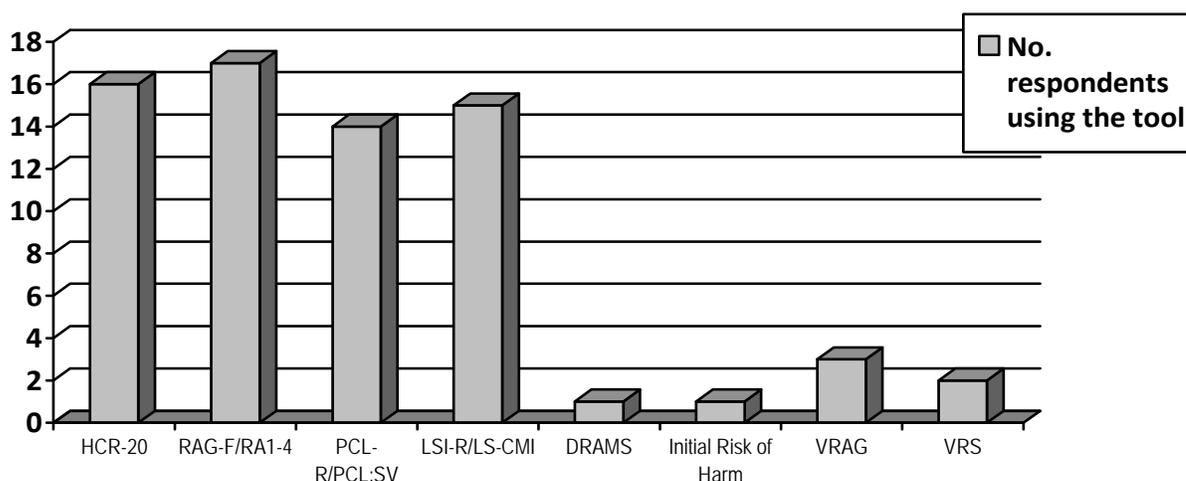
⁹ Some respondents provided information on risk assessment for specific types of violence, e.g. spousal, child abuse, and sexual violence. These were not included in the data analysis.

Table 2: Instruments for Assessing Violence Risk

Tool/Instrument	Brief Description
Historical Clinical Risk 20 (HCR-20) Webster, Douglas, Eaves & Hart (1997)	The HCR-20 is a structured clinical judgment approach to risk assessment. It was designed for use with mentally disordered adult offenders. It includes 20 risk factors, 10 Historical (including clinical constructs such as personality disorder, psychopathy and mental illness), 5 Clinical and 5 Risk Management items.
RAG-F (incorporating RA1-4) Scottish Executive (2000)	The RAG-F is described as a structured professional judgment approach to risk assessment that predicts reconviction. The tool has 4 parts: RA1 is designed to predict reconviction; RA2 is designed to consider dynamic risk factors relevant to the risk of reconviction; RA3 and RA4 screen and then fully assess for the risk of serious harm, identifying the triggers and likely scenarios where harm might occur.
Psychopathy Checklist Revised (PCL-R) Hare, (1991; 2002)	This instrument is not specifically designed for risk assessment but high scores on this instrument are correlated with violence. The PCL-R is a personality disorder assessment for offenders over the age of 18 years. It requires an evaluation of interpersonal, affective and behavioural features of this disorder.
Psychopathy Checklist: Screening Version (PCL:SV) Hart, Cox & Hare (1995)	The PCL:SV is a 12 item screening version of the PCL-R. This instrument has combined certain items of the PCL-R, has less emphasis on criminality and criminal history, and as such is appropriate for assessing psychopathy in a clinical population. Similar to the PCL-R the PCL:SV was not developed as a risk assessment tool but is a personality assessment. It is however correlated with violence.
Level of Service Inventory-Revised (LSI-R) Andrews & Bonta (1995)	This tool evaluates risk of re-offending and needs assessment. A primary aim is to assist in the development of supervision plans for offenders on probation but it has been shown to be correlated with recidivism. It is designed for use with offenders aged 16 and above.
Level of Service/Case Management Inventory (LS/CMI) Andrews, Bonta & Wormith (2004)	This instrument was designed for general application and not violence risk assessment per se but it has been shown to be related to violence (Bonta 2004, Girard & Wormith, 2004). The instrument is derived from the LSI-R and includes 8 sections designed to identify dynamic areas of risk/needs which are specifically relevant to the treatment offenders. The tool provides a score for the level of risk of reoffending and then in section 2 there is further consideration of factors related to criminogenic potential, responsivity, personal concerns and the history and nature of perpetration. The instrument contains a professional override to arrive at a final risk/need level between very high and very low.
Violence Risk Appraisal Guide (VRAG) Quinsey et al., (1998)	This is an actuarial instrument designed to assess violence risk among adult mentally disordered offenders. The VRAG has 12 factors which are used to allocate the individual to one of nine categories of risk.
Violence Risk Scale (VRS, Wong & Gordon, 2000, 2001)	The VRS is for use with offenders aged 18 and above. It includes 6 static and 20 dynamic risk factors. Scores are used to allocate offenders into different risk categories.
Dynamic Risk Assessment and Management System (Lindsay et al., 2004)	DRAMS is an assessment for dynamic and proximal risk factors for offending and challenging behaviour in people with intellectual disabilities.
Initial Test of Harm	No information available

In Figure 1, the frequency of use of each of the tools is presented. Because of the similarities between LSI-R and LS/CMI these tools were combined. Similarly, the PCL-R and PCL:SV were combined and the RAG-F incorporates the RA1-4. As is shown in Figure 1, the HCR-20, LSI-R and LS/CMI, PCL-R and PCL: SV, and the RAG-F were rated as the most commonly used assessment models. Because of the small number of responses regarding the other instruments (VRAG, VSR, DRAMS and Initial Test of Harm), these were excluded from further data analysis.

Figure 1: Frequency of Risk Assessment Tools Used by Participants



Perceived Utility of Risk Tools

The type of data collected and the sample size meant that descriptive statistics were most appropriate for analysing the results. The findings are detailed below according to conceptually similar headings with regard to the phases of risk assessment. The information is presented according to the frequency of endorsement of each rating option. In each cell, the number of missing responses is also listed. Missing data occurred when the rater either omitted to score the question or provided only qualitative information. Table 3 shows the ratings on whether the tool had a user manual, specified user qualifications and the skills level (expert, general or none) required to use the instrument.

Table 3: Frequency of Ratings on Questions Regarding User Manual, Qualifications and Skills Levels

Standard/Principle	HCR-20 N=16	RAG-F N=17	PCL-R N=14	LSI-R N=15
User Manual <i>Options</i> 1=Yes 0=No 9=Don't know	Yes=11 No=2 Don't know=0 (Missing=3)	Yes=7 No=0 Don't know=5 (Missing=5)	Yes=10 No=1 Don't Know=1 (Missing=2)	Yes=14 No=0 Don't Know=1 (Missing=0)
User Qualifications <i>Options</i> 1=Yes 0=No 9=Don't know	Yes=12 No=1 Don't know=0 (Missing=3)	Yes=2 No=4 Don't Know=6 (Missing=5)	Yes=8 No=0 Don't Know=3 (Missing=3)	Yes=10 No=0 Don't Know=4 (Missing=1)
Skills Level <i>Options</i> Expert=2 General=1 None=0	Expert=10 General=2 None=0 (Missing=4)	Expert=2 General=5 None=0 (Missing=10)	Expert=7 General=1 None=1 (Missing=6)	Expert=3 General=8 None=0 (Missing=4)

Table 4 shows the extent to which participants rated the accessibility of training on each of the tools.

Table 4: Frequency of ratings on Accessibility of Training

Standard/Principle	HCR-20 N=16	RAG-F N=17	PCL-R N=14	LSI-R N=15
Accessibility of Training <i>Options</i> 2=Very 1=Moderately 0=Not at all	Very access.=6 Mod. access.=6 Not access=0 (Missing=4)	Very access.=1 Mod. access.=5 Not access.=3 (Missing=8)	Very access.=3 Mod. access.=4 Not access=1 (Missing=6)	Very access.=3 Mod. access.=9 Not access.=2 (Missing=1)

Table 5 is concerned with administration issues. Participants' ratings on ease of use, the use

Table 5: Frequency of ratings on Administration

Standard/Principle	HCR-20 N=16	RAG-F N=17	PCL-R N=14	LSI-R N=15
Ease of Use <i>Options</i> 2=Very 1=Moderately 0=Not at all	Very easy=5 Mod. easy=8 Not easy=0 (Missing=3)	Very easy=4 Mod. easy=5 Not easy=1 (Missing=7)	Very easy=3 Mod. easy=7 Not easy=1 (Missing=3)	Very easy=5 Mod. easy=8 Not easy=1 (Missing=0)
Multi-informant data <i>Options</i> 2=Yes 1=No	Yes=13 No=0 (Missing=3)	Yes=9 No=2 (Missing=6)	Yes=9 No=1 (Missing=4)	Yes=10 No=5 (Missing=0)
Administration time 3>About right 2=Too long 1=Too short	About right=9 Too long=3 Too short=0 (Missing=4)	About right=4 Too long=3 Too short=1 (Missing=9)	About right=6 Too long=2 Too short=0 (Missing=6)	About right=10 Too long=0 Too short=2 (Missing=3)

of a comprehensive and multi-modal information set and an appropriate administration time are shown.

Table 6 is concerned with the principles regarding consideration of the offender's social history and the events leading to and surrounding their offence (offence analysis). The extent to which the tools were perceived as achieving this aspect of risk assessment is shown.

Table 6: Frequency of Ratings on Offender's Background Information and Offence Analysis

Standard/Principle	HCR-20 N=16	RAG-F N=17	PCL-R N=14	LSI-R N=15
Social History <i>Options</i> 3=Very 2=Moderately 1=Not at all	Very=7 Mod=5 No=0 (Missing=4)	Very=5 Mod=4 No=2 (Missing=6)	Very =6 Mod.=2 No =2 (Missing=4)	Very =0 Mod.=11 No =1 (Missing=3)
Offence Analysis <i>Option</i> 2=Yes 1=No	Yes=8 No=3 (Missing=4)	Yes=9 No=0 (Missing=6)	Yes=3 No=7 (Missing=6)	Yes=5 No=8 (Missing=1)

Table 7 displays the ratings for the number and appropriateness of type of risk factors included in the instruments and the clarity and usefulness of their respective definitions.

Table 7: Frequency of Ratings on Appropriateness of Risk Factors

Standard/Principle	HCR-20 N=16	RAG-F N=17	PCL-R N=14	LSI-R N=15
No. of risk factors <i>Options</i> 3=about right 2=too few 1=too many	About right=11 Too few=0 Too many=0 (Missing=5)	About right=8 Too few=1 Too Many=0 (Missing=8)	About right=4 Too few=3 Too Many=1 (Missing=6)	About right=9 Too few=2 Too Many=1 (Missing=3)
Emphasis on static factors <i>Options</i> 3=about right 2=too few 1=too many	About right=11 Too many=0 Too few=1 (Missing=4)	About right=7 Too few=1 Too Many=1 (Missing=8)	About right=6 Too few=0 Too Many=3 (Missing=5)	About right=5 Too few=3 Too Many=3 (Missing=3)
Emphasis on dynamic factors <i>Options</i> 3=about right 2=too few 1=too many	About right=8 Too many=2 Too few=1 (Missing=5)	About right=6 Too few=3 Too Many=0 (Missing=8)	About right=2 Too few=4 Too Many=2 (Missing=6)	About right=5 Too few=6 Too Many=0 (Missing=3)
Definition of risk factors <i>Options</i> 2=very well 1=moderately well 0=not well	Very well=9 Mod. well=2 Not at all=0 (Missing=5)	Very well=5 Moderately well=3 Not at all well=1 (Missing=8)	Very well=0 Moderately well=5 Not at all well=2 (Missing=7)	Very well=1 Mod. well=2 Not at all well=1 (Missing=11)

Ease of risk ratings <i>2=very</i> <i>1=moderately</i> <i>0=not at all</i>	Very easy=10 Mod. easy=2 Not at all=0 (Missing=4)	Very well=6 Moderately well=3 Not at all well=0 (Missing=8)	Very well=2 Moderately well=3 Not at all well=2 (Missing=7)	Very well=0 Mod. well=3 Not at all=0 (Missing=12)
Identification of Protective Factors <i>2=very</i> <i>1=moderately</i> <i>0=not at all</i>	Very well=0 Mod. well=5 Not at all=7 (Missing=4)	Very well=3 Mod. well=1 Not at all=6 (Missing=7)	Very well=1 Mod. well=1 Not at all =2 (Missing=10)	Very well=0 Mod. well=3 Not at all=0 (Missing=12)

The data shown in table 8 presents the ratings on the processes between risk ratings and risk management (i.e., risk formulation and scenario planning). Ratings for the utility of the tools on facilitating risk management and its component parts are also shown.

Table 8: Frequency of Ratings on Risk Formulation, Scenarios and Management

Standard or Principle	HCR-20 N=16	RAG-F N=17	PCL-R N=14	LSI-R N=15
Risk Formulation <i>Options</i> <i>2=very</i> <i>1=moderately</i> <i>0=not at all</i>	Very well=6 Moderately well=4 Not at all well=2 (Missing=4)	Very well=1 Moderately well=2 Not at all well=5 (Missing=9)	Very well=1 Moderately well=6 Not at all well=0 (Missing=7)	Very well=1 Moderately well=3 Not at all well=0 (Missing=11)
Risk Scenarios <i>Options</i> <i>2=very</i> <i>1=moderately</i> <i>0=not at all</i>	Very well=4 Moderately well=4 Not at all well=4 (Missing=4)	Very well=2 Moderately well=2 Not at all well=6 (Missing=7)	Very well=1 Moderately well=6 Not at all well=0 (Missing=7)	Very well=0 Moderately well=2 Not at all=2 (Missing=11)
Risk Management <i>Options</i> <i>2=very</i> <i>1=moderately</i> <i>0=not at all</i>	Very well=5 Moderately well=7 Not at all well=0 (Missing=4)	Very well=2 Moderately well=4 Not at all well=4 (Missing=7)	Very well=1 Moderately well=5 Not at all well=1 (Missing=7)	Very well=1 Moderately well=2 Not at all =1 (Missing=11)
Treatment <i>Options</i> <i>2=very</i> <i>1=moderately</i> <i>0=not at all</i>	Very well=5 Moderately well=2 Not at all well=5 (Missing=4)	Very well=2 Moderately well=3 Not at all well=4 (Missing=8)	Very well=0 Moderately well=4 Not at all well=3 (Missing=7)	Very well=1 Moderately well=3 Not at all well=0 (Missing=11)
Monitoring <i>Options</i> <i>2=very</i> <i>1=moderately</i> <i>0=not at all</i>	Very well=7 Moderately well=3 Not at all well=2 (Missing=4)	Very well=3 Moderately well=2 Not at all well=4 (Missing=8)	Very well=0 Moderately well=7 Not at all well=0 (Missing=7)	Very well=0 Mod well=4 Not at all well=0 (Missing=11)

Table 9 shows the perceived level of agreement between tools by users who utilise more than one instrument and also the level of confidence users have in their judgments using the instruments.

Table 9: Frequency of Ratings on Conclusions and Confidence in Judgement

Standard/Principle	HCR-20 N=16	RAG-F N=17	PCL-R N=14	LSI-R N=15
Concordance of conclusions <i>Options</i> 2=very 1=moderately 0=not at all	Very well=6 Moderately well=4 Not at all well=0 (Missing=6)	Very well=5 Moderately well=4 Not at all well=0 (Missing=8)	Very well=2 Moderately well=2 Not at all well=0 (Missing=10)	Very well=1 Moderately well=0 Not at all well=0 (Missing=14)
Confidence in judgment <i>Options</i> 2=very 1=moderately 0=not at all	Very=7 Moderately=5 Not at all=0 (Missing=4)	Very=2 Moderately=4 Not at all=3 (Missing=7)	Very=2 Moderately=3 Not at all=1 (Missing=8)	Very=0 Moderately=1 Not at all=0 (Missing=14)

As indicated above, on each questionnaire and during each interview, participants were asked to provide any additional comments with respect to the risk assessment standards and practice in violence risk assessment. The responses given are detailed in Table 10.

Table 10: Qualitative Comments on Each of the Standards

Standard or Principle	Comments/Feedback
User Manual	<i>Referring to HCR-20 “[manuals exist] but are not held on the ward” (P22)</i> <i>Referring to RAI-4 “relatively little guidance” (P4)</i>
User Qualifications	<i>“I presume as a clinical psychologist I can use the VRAG” (P7)</i> <i>“the RAG-F requires no formal training or standard competency [for use. It is] a very subjective tool...” (P17)</i> <i>“...the user qualifications/competencies for use of all these tools should be more clearly specified” (P19)</i> <i>“the user requires expertise in conducting individual assessments and knowledge in the study of violence/psychopathy (depending on the tool used)”.(P22)</i> <i>Referring to the HCR-20 “[it is] difficult for social workers to receive the psychological oversight...required to prove competency” (P24)</i>
Skills Level	<i>Referring to HCR-20 “expert skills to sign off and general skills to be part of the clinical team/ discussions” (P16)</i> <i>Referring to the HCR-20 “The user should have training specific to the tool they are employing. [Nurse] assessors should be supervised by identified psychology lead” (p22)</i>
Accessible Training	<i>“It is always an effort to access training from Aberdeen” (P7)</i> <i>Referring to HCR-20 training “not [accessible] so far” (P4)</i> <i>“There is a need for different levels of training for those competent/ qualified to conduct the assessments themselves vs. Those who can only contribute, i.e., not trained to conduct personality or clinical assessments” (P19)</i>

	<p>Referring to HCR-20 “Scottish/ Canadian events are very expensive. We brought in our own consultant to train 30 social workers” (P24)</p> <p>Referring to LSI-R “[we had] 2 days training from the Cognitive Centre. We believe we could deliver a higher standard of training in-house but are not permitted to by licence conditions” (P24)</p>
Ease of Use	<p>“You have to take great care with the HCR-20 and PCL-R. Use with caution” (P7)</p> <p>“Some of the scoring terms are confusing on the structured clinical judgement tools, e.g. ‘?’ does not mean ‘don’t know’ but [partially present]. All the numbers should go. [There is also] confusion over the meaning of “provisional diagnosis...I have concerns over the complicated nature of the PCL scores [sliders up and down] and the validity of each item being equally weighted regardless of the amount of evidence there might be for each” (P19)</p> <p>Referring to HCR-20 “easy if you are trained” (P29)</p> <p>Referring to PCL “not at all easy even if you are trained” (P29)</p>
Multi-informant data	<p>“it is very appropriate to obtain collateral sources of information” (P22)</p>
Administration time	<p>Referring to RAI-4 “it can be time consuming” (P4)</p> <p>“It is difficult to answer this question. It depends on the purpose of the assessment. The administration tool is only one part of the assessment and the length of time taken for this will depend on the complexity of the case, availability of information, etc” (P19)</p>
Comprehensive social history	<p>“if information is available” (P16)</p> <p>“varies case by case” (P17)</p> <p>“it is important to tell the person’s story [history] in an assessment report which often means going beyond what is covered by the tools” (P19)</p> <p>“..I find [LSI-R] gathers factual data information well but the ability to decipher the embeddedness and multivariate correlations appears to be somewhat lacking” (P24)</p> <p>Referring to the PCL-R “not at all comprehensive, it is only one facet of risk that is assessed” (P29)</p>
Comprehensive offence analysis	<p>“The tools do not help enough with this – although the relevant information is sought, the analysis is a complex task and insufficient attention given to it in the manuals” (P19)</p> <p>“The lack of ability of actuarial tools to assess more complex behaviours which present in different ways in different circumstances keep them from being useful as guides to practice” (P21)</p> <p>“The tool [HCR-20] is a helpful aid in the formulation of risk” (P22)</p> <p>“No, this is not what they are for” (P29)</p>
Number of risk factors	<p>“Referring to HCR-20 and PCL-R “about right” (P7)</p> <p>“many of the factors detailed in the PCL-R are repeated in the HCR-20 so it often feels as if the individual is penalised twice for the same behaviours” (P9)</p> <p>So long as there are no problems with methodology, I am of the opinion that the more variable they are, the more likely we are to be able to understand the people we are testing” (P21)</p> <p>Referring to LSI-R “only measures general risk of reconviction and criminogenic need”</p> <p>Referring to HCR-20 “about right, they allow for case-specific risk factors” (P29)</p>
Emphasis on static factors	<p>Referring to PCL-R “I would not use as a stand-alone risk assessment” (P7)</p> <p>“While static factors may be predictive from an actuarial perspective, they cannot facilitate change. The more we standardise people the more we neglect individual strengths and the things that are most likely to decrease recidivism one person at a time” (P21)</p>
Emphasis on dynamic factors	<p>“It is a step forward that we are beginning to look at this increasingly but we have a long way to go and many dynamic factors are very individualised such that they cannot be standardised into further actuarial tools” (P21)</p>
Definition of risk factors	<p>“[The risk factors in RAI-4 [could be better explained. Need more definition, particularly of Risk of Harm” (P18)</p> <p>“such definitions appear to be fitting people to problems not to solutions” (P21)</p> <p>“risk factor definitions [in HCR-20 and PCL-R] are helpful in guiding the assessor although can be open to interpretation depending on the assessor” (P22)</p> <p>Referring to PCL-R “definitions are still difficult, The screening version has clearer definitions” (P29)</p>
Ease of risk ratings/scoring	<p>For HCR-20 and PCL-R, the participant indicated they rated it easy “after training” (P7)</p> <p>“some people are still “scoring” and others not” (P16)</p>

	<p>Referring to RAI-4 “you have to pull it together yourself without a score. All quite subjective” (P18)</p> <p>“most of the tools I have training in or have reviewed are quite self-explanatory” (P21)</p> <p>Referring to LSI-R “some sections too complicated to be used easily” (P23)</p>
Identification of Protective Factors	<p>Referring to HCR-20 “you need to reverse the risk factor to achieve this” (P7)</p> <p>“far too little emphasis on protective factors” (P7)</p> <p>Referring to LSI-R “by implication if not scored [item is] protective” (P14)</p> <p>“much more emphasis is required” (P19)</p> <p>Referring to LSI “Protective factors, when using an actuarial tool, often appears to take a cookie-cutter approach to identifying factors. Protective factors are typically specific to individuals. What these tools appear to do is allow for law enforcement to determine a level of response but not guide any form of social work or clinical practice” (P21)</p> <p>[None of the tools achieve this, [in the literature we] still can't decide what is a protective factor” (P29)</p>
Risk Formulation	<p>Referring to LSI “I am a strong proponent of a psychosocial assessment of any person. All of these tools leave out the factors that can make a person more or less dangerous. There is no chance to look into the embeddedness of interpersonal traits and decipher how they might influence the factors included in any actuarial tool” (P21)</p> <p>Referring to LSI “does not help at all as does not cover dynamic risk factors or protective factors” (P23)</p> <p>“LSI-R does not assess for risk of violence, only risk of reconviction” (P27)</p> <p>Referring to HCR-20 and PCL-R “not as they are constituted” (P29)</p>
Risk Scenarios	<p>“you have to be creative [with HCR-20 and PCL-R]” (P7)</p> <p>Referring to LSI “...these tools, at best, identify few specifics and more general categories of risk” (P21)</p> <p>“the tools [HCR-20 and PCL-R] help to do this” (P22)</p> <p>“LSI-R does not take account of one off serious offences, for example, first offence murder” (P27)</p>
Risk Management	<p>Referring to PCL-R “not as stand alone” (P7)</p> <p>Referring to LSI-R “Andrews & Bonta developed this tool for a system (Canada) with different levels of care than those available in Scotland. It is meant to help guide the level of care provided but many of the suggested [levels] are not available here, e.g. halfway houses” (P21)</p> <p>Referring to LSI-R “...I find actuarial tools and any tool that attempts to create categories/types of people counter to the values of the social work profession. Nonetheless, as our role is dualistic...I can see the benefit of these tools...nonetheless, risk management plans which are created through the input of these tools are likely to be law enforcement based” (P21)</p> <p>“Both tools [LSI-R and RAI-4] give a broad overview in terms of what constitutes risk and what needs to be addressed via intervention” (P27)</p> <p>Referring to PCL-R “not at all good, it is only one facet” (P29)</p>
Treatment	<p>“we use a team approach to define the management plan” (P7)</p> <p>“LSI-R can be helpful” (P21)</p>
Monitoring	<p>“the LSI-R can identify factors that need monitoring but is not wholly suited to this” (P18)</p> <p>“[RAI-4] can identify risk factors that need monitoring” (P18)</p> <p>Referring to LSI-R “I believe the aforementioned tools can be helpful”</p>
Concordance of conclusions with other tools	<p>Referring to HCR-20 “it converges with team opinions as well as tools used by other professions” (P7)</p> <p>“I don't use other tools because nothing works as well as HCR-20 or PCL-R ...actuarial tools are misleading” (P29)</p>
Confidence in judgement	<p>“not so when assessing females” (P18)</p> <p>“my confidence would be greater if greater attention was paid to offence analysis” (P19)</p> <p>Referring to LSI-R “while these tools have significant statistical validity, what that means in human terms is that they are pretty accurate in 3 of 4 cases. As a person who prides themselves in trying to help anyone I am working with, I cannot tolerate these numbers and to think that courts are using the results from some of these tools to sentence people is even more disconcerting” (P21)</p> <p>“Not at all. The LSI-R does not give a comprehensive assessment of risk issues” (P23)</p> <p>Referring to HCR-20 “good for violent offenders. Very time consuming to complete” (P24)</p>

	Referring to PCL-R “good for identifying psychopathy” (P24) Referring to LSI-R “good for identifying risk of general criminal conviction” (P24)
Other comments	<p>“The training in risk assessment could involve more staff types...it’s a top down process” (P6)</p> <p>“I’d really like to see the RMA doing less research and more actual training on the tools identified above and to encourage the use of risk assessment processes that are not prohibitively expensive in respect of either training or ongoing costs (LSI-R)” (P23)</p> <p>“I have concerns that we (SPS) might miss people who should have risk assessments” (P28)</p> <p>“I have concerns regarding the use of risk assessment across the system in Scotland particularly the use of actuarial tools and the lack of understanding about how misleading they are. There is a danger that risk management will fall into disrepute because social workers are required to do these assessment without resources. There is a need for more research on specific violence risk such as knife crime”. (P29)</p>

Table 11: Summary Table

Standard/Principle	HCR-20	RAG-F/RA1-4	LSI/LS-CMI	PCL-R/PCL:SV
User Manual	✓	✓	✓	✓
User Qualifications	✓	x	✓	✓
Skills Level	✓	✓	✓	✓
Accessible Training	✓	✓	✓	✓
Ease of Use	✓	✓	✓	✓
Multi-informant data	✓	✓	✓	✓
Administration time	✓	✓	✓	✓
Comprehensive social history	✓	✓	✓	✓
Comprehensive offence analysis	✓	✓	✓	✓
Number of risk factors	✓	✓	✓	✓
Emphasis on static factors	✓	✓	✓	✓
Emphasis on dynamic factors	✓	✓	x	x
Definition of risk factors	✓	✓	✓	x
Ease of risk ratings/scoring	✓	✓	✓	x
Identification of Protective Factors	x	x	x	x
Risk Formulation	✓	x	✓	✓
Risk Scenarios	✓	x	✓	✓
Risk Management	✓	✓	✓	✓
Treatment	✓	✓	✓	✓
Monitoring	✓	✓	✓	✓
Concordance of conclusions with other tools	✓	✓	✓	✓
Confidence in judgement	✓	✓	✓	✓
Total	21/22	18/22	20/22	18/22

Table 11 provides an overall summary of the participants’ ratings. Those standards that are marked with ✓ are those which have been rated by most users as either moderately or fully achieving the standard/principle. Those marked with x are those which were marked by most respondents as not meeting the standard. Most instruments met most standards, but the least satisfactory instruments were the RAG-F and PCL-R.

Discussion

The purpose of this study was to assess practitioners' views on the utility of the various violence risk assessment instruments used across Scotland. The most commonly used devices were evaluated according to a list of standards and principles articulated in the available literature and best practice guidelines. The general findings and the implications of the results are discussed below.

Based on the available data, the RAG-F, which is an adjusted actuarial approach to violence risk assessment was the most commonly used method of risk assessment among the respondents sampled. This is an important finding for the Scottish context since structured clinical judgment approaches to violence risk assessment are considered the most appropriate methodology for assessing violence risk (RMA Scotland, 2006). The implications of this finding are important. When considered in light of the other data obtained in this study, which showed that the RAG-F was perceived to be comparatively poorer on standards and principles for risk assessment, it can be concluded that practitioners who are required to conduct risk assessments are using a method which is not considered optimal by either the research or best practice literatures or indeed by those who use the instrument. In addition, the widespread application of actuarial tools has been an issue of concern. One of the respondents in the current study indicated that there are many misunderstandings about the application of actuarial tools to the extent that they considered it possible that "risk management will fall into disrepute" (see Table 10 above) because of the widespread use of these tools.

The psychopathy assessments were also rated comparatively poorly on important aspects of risk assessment. It has been noted elsewhere (Hare, 2000, cited in RMA Scotland, 2007) that psychopathy assessments (i.e., PCL-R and PCL:SV) were not designed for the

purposes of risk assessment; the inclusion of these tools is largely the result of the strong association between the personality disorder traits and violence. The data from this study are consistent with the notion that these instruments should therefore be used only as part of the entire risk assessment process. The usefulness of psychopathy is supposed to rest its ability to provide comprehensive information about static and historical factors relevant to understanding the person, rather than risk assessment per se.

Based on the quantitative information, the HCR-20, followed by the LSI-R/LS/CMI, were rated most favourably by practitioners in their ability to fulfil the various phases of a risk assessment evaluation. However, the qualitative responses raised several concerns regarding the LSI-R/LS/CMI as a violence risk assessment measure. Participants indicated that this instrument was more suited to providing information regarding criminogenic as opposed to violence-specific needs. In addition, concerns were also raised about using an instrument that carried with it the risk of false positives (see Table 10 above).

Overall, the HCR-20 was rated most favourably in both the quantitative and qualitative data. However, all tools considered in this study were rated by practitioners as being poorly equipped to identify protective factors. This was even true of LS/CMI, which explicitly takes account of 'strengths' or protective factors. However, it is important to be cautious in interpreting this finding. Qualitative information revealed a level of confusion about the meaning of the term 'protective factor' with two respondents indicating that the absence of a risk factor was equal to the presence of a protective factor. This is an erroneous assumption. This finding highlights that there is a training need to help assessors understand the difference between risk and protective factors and that risk assessment tools should ensure that protective factors are appropriately defined and given due consideration.

In terms of accessibility of training, most tools were rated as having very or moderately

accessible training but it was noted that some of the courses were expensive to access, e.g. the HCR-20 and PCL-R. This is an important consideration for agencies who are working towards best practice in risk assessment. Because risk assessment is a very complex endeavour it is critical that assessors are highly trained and (as the data revealed) there are many advantages and disadvantages to using external training consultants as opposed to internally delivered training. For example, external consultants may be more expensive and have restricted availability but they might also be highly knowledgeable and expert in an area in which they specialise and have experience at the academic and practical level. If their training fails to meet the needs of the organisation it may be possible to find an alternative. Internally based trainers may have the advantage and ability to design and deliver context/service specific training needs, they are on hand, and they are likely to cost less.

The findings from this study show that training is an important aspect of developing a framework for practice. In considering the need for accessible training, it may be productive for agencies to consider securing different types of training (e.g., education versus skills acquisition) from different types of people (e.g., internal trainers versus external expert consultants) for different purposes (e.g., to promote multi-disciplinary team or agency assessments versus training for expert opinions). In addition, some sort of governance of standards might be useful.

Limitations

It is important to note the limitations to this study. Despite sampling widely, only a small proportion of possible participants provided data. The generalisability of the findings is therefore difficult to establish. It is possible that, because the questionnaire was sent to busy practitioners and managers, potentially useful informants were unable to find the time to complete and return it. The small sample size should be borne in mind when interpreting the

conclusions of this report. A further issue is that some of the people who were invited to take part might not be fully up to date with some of the risk assessment principles being asked about. For example, scenario planning and risk formulation are relatively new developments in this field and are better articulated in the sexual violence risk assessment literatures (e.g., Hart et al., 2003). In addition, because of the small numbers only descriptive statistics were possible. Differences between tools were not examined for statistical significance.

Summary and Recommendations

We conclude by making recommendations about a framework for assessing the risk of violence and what the RMA might do next to develop a screening instrument. Our findings indicate that, while a range of risk assessment tools are used across various Scottish agencies involved in the practice of violence risk assessment, the RAG-F is most commonly employed. Notwithstanding its widespread use, the RAG-F was viewed least positively by participants on its ability to achieve the standards and principles of risk assessment articulated for this study. The main recommendation that might be drawn from this finding is that the practice of violence risk assessments should be based on the structured clinical judgment model of risk assessment, following the HCR-20. Quantitative and qualitative responses regarding the use of this instrument indicated that it was considered most suitable for the purposes of risk assessment. However, limitations to this tool were also identified, specifically the need to address protective factors.

It can also be concluded from this study that training should be accessible and user competency should be appropriate to the task in hand and the tool being used. In addition, it would be helpful for researchers to explore the utility of other aspects of risk assessment evaluations, such as risk formulation, the success of risk management plans, etc. However, given the methodological limitations in this study it would be useful to confirm the findings by

using a larger sample size and/or an alternative methodology such as focus groups.

Reflecting on the findings from Parts 1 and 2, it is apparent that a comprehensive risk assessment demands high level skills and is resource intensive. For example, a full HCR-20 requires a specialist assessment of mental disorder (psychopathic personality disorder, mental illness, personality disorder) by an appropriately qualified mental health professional such as a chartered clinical or forensic psychologist or a forensic psychiatrist. It is possible (perhaps likely) that many violent offenders will not have access to such expertise and therefore will be unable to benefit from the opportunity to have their risk fully assessed and managed. It therefore becomes apparent that short-term decisions about risk such as the need for hospital admission, the need to be taken to a place of safety, suitability for bail as opposed to remand, the need for any urgent risk management interventions (e.g. medication or victim-safety planning) are still likely to be dependent on unstructured opinions which are problematic. In addition, referral to specialist services for an opinion on risk might occur on an ad hoc basis and thus contribute to patchy service provision. It would be beneficial to have referral decisions evidenced and systematised. In sum, there does appear to be a role and indeed a need for a brief risk assessment tool. However, there are significant challenges associated with developing such an approach.

An overriding concern underpinning the practice of risk assessment is the potential consequences of a false positives and false negatives. On the one hand is the risk of someone falling victim to violence that could have been prevented. On the other hand, people may be deprived of their liberty (full or partial) because they were wrongly deemed to present a significant risk of harm. We believe that our review of the empirical bases and practitioners' views provides an important starting point for identifying some key issues that would need to

be addressed in developing a screening instrument and about what features a brief risk assessment tool might need to possess. These are listed below:

1. Purpose and Function: Any risk assessment should state unequivocally what it is designed to do (and what it does not do). For example, a brief risk assessment tool might be used most appropriately to inform urgent short-term decisions or to prioritise someone for a more detailed assessment.
2. Manualised, Training, User-Competencies: We recommend that any risk assessment tool must link scientific evidence, theory, practice and professional standards to provide an efficient, reliable, valid, and ethical model of accomplishing an evaluation. Any such tool should have an administration manual and training programme to facilitate its reliable use and appropriate application.
3. Risk and Protective Factors: We recommend that risk factors addressed in a brief risk assessment should be derived from empirical evidence and have good empirical support. However, we also acknowledge the need to include dynamic risk factors and protective factors. While the evidence base is not as well established for these, the qualitative information confirmed that they are important. Any risk assessment should be based on sound evidence and professional guidelines.
4. Individualised: Any instrument developed for the purpose of a brief assessment should be sensitive to the individual case. Individuals who commit serious harm are a heterogeneous group. For example, motives might be instrumental, reactive, psychotic or contextual. The structured clinical judgment approach is therefore the most appropriate model to identify these and facilitate an individualised assessment.
5. Format: One of the most demanding and time-intensive parts of preparing a risk assessment is the need to write a detailed report. It is recommended that any brief

assessment has corresponding worksheets such as those developed for the Risk for Sexual Violence Protocol (Hart et al., 2003), but with additional sections for information such as risk formulation. These worksheets might provide useful initial communications about risk without the need for a full written report.

- Triage Model: It is unlikely that any brief risk assessment will fully and properly delineate the true nature of a person's risks or needs. These instruments might however help identify those who are most in need of a full risk assessment (e.g. those with past serious harm, escalating violence, violent ideation) and those who do not appear to merit intensive assessment (e.g. those with a history of minor verbal aggression).
- Audit and Evaluation: There is a need to ensure that methodologies for establishing the utility of a risk assessment tool are sensitive to the wide gamut of aims that the process must fulfil. Validation research must go beyond predictive validity. We know that group data do not necessarily generalise to individuals and that the confidence intervals of actuarial tools are so large that some experts have seriously called into question their use (Hart, Michie & Cooke, 2007). Specifically, group-based risks of reoffending may not apply to unusual individual cases.

Case Example

It is worthwhile to acknowledge that some services have already attempted to tackle these issues, albeit at a local level. For example, the Directorate of Forensic Mental Health, Glasgow, has endorsed a policy on risk assessment which includes different levels of risk assessment. They have distinguished a Level 1 from a Level 2 risk assessment.

The function of a Level 1 risk assessment is to act as a screening instrument for all types of clinical risks potentially relevant to the patient. This is designed to inform immediate risk management and admission decisions in the context of the overall clinical assessment. It

is not viewed as a robust assessment of risk. Level 1 risk assessments are completed by any member of the clinical team and are expected to achieve the following outcomes: 1) inform the short-term management of the patients; 2) highlight the nature of risks that require to be more fully assessed; 3) facilitate access to any identified therapies or interventions that are shown to be needed (so that risk management is not delayed); and 4) identify and implement other risk management interventions deemed necessary, e.g. admission, disclosure, medication, victim-safety planning, etc. In this context, patients who have a prior history (formal or informal) of serious violence or a pattern of violent conduct and whose Level 1 risk screen indicates that a full risk assessment is required are then identified as in need of a Level 2 risk assessment.

A Level 2 risk assessment is based on the structured clinical judgment model. It is coordinated by a fully trained and experienced assessor but relies on multi-disciplinary input and is expected to provide a robust answer to the following: What risks (e.g. instrumental, psychotic, reactive violence) are present; what risk factors are present; what risk factors are relevant; what are the main risk scenarios to be managed; what management action needs to be taken or planned in the immediate, short and longer term; what additional assessments or evaluations need to be completed, by whom, and by what date should it be complete and also dates for review.

This model was accepted because it was viewed by the management and multi-disciplinary team staff as providing an appropriate method of assessing and managing risk and providing a triage model of intervention. However, this policy has not been subjected to any formal evaluation or auditing. Therefore, it is not known whether it achieves its goals but it might at least provide a useful framework for the RMA in developing a screening instrument.

REFERENCES

- Andrews, D., Bonta, J., & Wormith, S. (2004). LS/CMI: The Level of Service/Case Management Inventory: An Offender Assessment System. Toronto: Multi-Health Systems.
- Andrews, S., and Bonta, J. (1995). LSI-R: The Level of Service Inventory-Revised. Toronto: Multi-Health Systems.
- Hare, R.D. (1991). The Psychopathy Checklist Revised. Toronto: Multi-Health Systems.
- Hare, R.D. (2000). Presentation to Risk Assessment and Risk Management: Implications for the Prevention of Violence. A Report of the Committee on Serious Violence and Sexual Offenders, Annex 4. Edinburgh: Scottish Executive.
- Hare, R.D. (2003). The Psychopathy Checklist Revised, 2nd Edition. Toronto: Multi-Health Systems.
- Hart, S.D., Cox, D., & Hare, R.D. (1995). The Psychopathy Checklist: Screening Version. Toronto: Multi-Health Systems.
- Hart, S.D., Kropp, R., Laws, R., Klaver, J., Logan, C., & Watt, K.A. (2003). The Risk for Sexual Violence Protocol. Burnaby, BC Mental Health and Law Policy Institute, Simon Fraser University, Pacific Psychological Assessment Corporation and the British Columbia Institute Against Family Violence.
- Hart, S.D., Michie, C. & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments. British Journal of Psychiatry, 190 (suppl. 49), S60-S65.
- Hollin, C. (1995). The meaning and implications of "programme integrity". In J. McGuire (Ed.), What Works: Reducing Reoffending – Guidelines from Research and Practice (pp. 195-208). Chichester: Wiley.
- Lindsay, W., Murphy, L., Smith, G., Murphy, D., Edwards, Z., Chittock, C., Grieven A., &

- Young, S.J., (2004). The dynamic risk assessment and management system: An assessment of immediate risk of violence for individuals with offending and challenging behaviour. Journal of Applied Research in Intellectual Disabilities,17, 267- 274.
- Quinsey, V., Harris.,G., Rice,M., & Cormier, C., (1998). Violent Offenders: Appraising and Managing Risk. Washington, DC: American Psychological Association.
- RMA Scotland (2006). Risk Assessment Tools Evaluation Directory, Version 1. Edinburgh: Risk Management Authority, Scotland.
- RMA Scotland (2007a). Risk Assessment Tools Evaluation Directory, Version 2. Edinburgh: Risk Management Authority, Scotland.
- RMA Scotland (2007b). Standards and Guidelines: Risk Management of Offenders Subject to an Order for Lifelong Restriction. Version 1. Edinburgh: Risk Management Authority, Scotland.
- Scottish Executive (2000). Risk Assessment Guide Framework. Edinburgh: Scottish Executive.
- Webster, C., Douglas, K., Eaves, D., & Hart, S.D., (1997). Historical Clinical Risk-20: Assessing Risk for Violence, Version 2. Burnaby, BC: Mental Health, Law and Policy Institute, Simon Fraser University.
- Webster, C.D., Martin, M.L., Brink, J., Nicholls, T.L., & Middleton, C. (2004). Short-Term Assessment of Risk and Treatability (START). Hamilton, Ontario: St Joseph's Healthcare and Forensic Services Commission.
- Wong, S. & Gordon, A. (2000). Violence Risk Scale. Saskatoon: Research Unit, Regional Psychiatric Centre.
- Wong, S., & Gordon, A., (2001). The Violence Risk Scale. Bulletin of the International Society for Research on Aggression, 23 (2).

Technical Appendix for Part 2

Covering letter and questionnaire

Dear Colleague

**RE: Risk Management Authority Research Study
Assessing Risk for Violence: A Framework for Practice**

We have been commissioned by the Risk Management Authority Scotland to undertake research into the effectiveness of available violence risk assessment tools. We would be extremely grateful if you would take a few minutes to complete and return the enclosed questionnaire so that we can consider your views.

Once we have gathered views from the range of disciplines involved in the practice of risk assessment, we will analyse the data and produce a report. The findings from this stage of the research will compliment a larger quantitative analysis.

Please be assured that, as with all research, participation is voluntary and your information will remain anonymous.

If you are willing to take part, please can you complete the questionnaire electronically and forward it to Lorraine.Johnstone@ggc.scot.nhs.uk by **31st October 2007**. We apologise for the short timescale but we are working to a tight deadline.

We trust this research will be of interest to you. However, if you would like additional information or you would like to discuss any aspect of this request, please contact Dr Lorraine Johnstone (details above).

Thank you for your assistance.

We look forward to hearing from you in due course.

Yours faithfully

**Prof. David P. Farrington
Professor of Criminology**

**Dr Darrick Jolliffe
Research Fellow**

**Dr Lorraine Johnstone
Consultant Clinical Forensic
Psychologist**

Enc.

ASSESSING RISK FOR VIOLENCE: A FRAMEWORK FOR PRACTICE

INTRODUCTION: This questionnaire has been designed for the purposes of gathering information from practitioners on the extent to which the various tools used for or as part of violence risk assessments meet and/or facilitate the standards for risk assessment articulated by the Risk Management Authority and other standards that are likely to be relevant to the effective use of risk assessment tools. The focus of this research is on the process and outcome of the use of these tools as opposed to their evidence base (this is addressed in the RMA Scotland publication entitled RATED). We would be very grateful if you would spend some time to provide the following information and completing the following ratings to our list of questions. We would also appreciate any other comments you may have.

PERSONAL DETAILS	Place of work:	Population:
Name:	Designation:	Is violence risk assessment part of your role? Yes/No

RISK ASSESSMENT TOOLS. Some of the most commonly used risk assessment tools and assessment procedures for evaluating violence risk are listed below. Please tick those which you use and detail any others used by you/your organisation.

	Please tick if used
HCR-20: Historical Risk Clinical 20 (Webster et al, 1997)	
VRAG: Violence Risk Appraisal Guide	
VRS: Violence Risk Assessment Scale	
PCL-R: Psychopathy Checklist Revised (Hare, 1991; 2003)	
PCL:SV: Psychopathy Checklist Screening Version (Hart & Hare 1995)	
LSI: Level of Service Inventory (Andrews & Bonta, 1995)	
LSI/CMI: Level of Service/Case Management Inventory, (Andrews et al., 2004)	
COVR: Classification of Violence Risk (Monahan et al., 2000)	
OASyS: Offender Assessment System (NOMS, 1999)	
OGRS: Offender Group Re-Conviction Scale (Home Office, 1996)	
CARDS: Clinical Assessment of Risk Decision Support (Watts et al., 2004)	
RAGF: Risk Assessment Guide Framework, (Scottish Executive, 2000)	
DRAMS: Dynamic Risk Assessment Management System (Lindsay et al., 2005)	
Other, please specify:	
Other, please specify:	

Please answer the following questions and provide any additional comments or observations about each of the standards in the space provided

STANDARD 1	Tool	Rating	Please add any comments about the availability and effectiveness of the user manuals:	
Is there a user manual?	HCR-20	1 = Yes, 0 = No, 9 = Don't know		
	VRAG	1 = Yes, 0 = No, 9 = Don't know		
	VRAS	1 = Yes, 0 = No, 9 = Don't know		
	PCL-R	1 = Yes, 0 = No, 9 = Don't know		
	LSI-R	1 = Yes, 0 = No, 9 = Don't know		
	LSI-CMI	1 = Yes, 0 = No, 9 = Don't know		
	Other, specify	1 = Yes, 0 = No, 9 = Don't know		
STANDARD 2	Tool	Rating	Please add your comments or views about the user qualifications and competencies required to use this tool?	
	Are there specified user qualifications and competencies for using this tool?	HCR-20		1 = Yes, 0 = No, 9 = Don't know
		VRAG		1 = Yes, 0 = No, 9 = Don't know
		VRAS		1 = Yes, 0 = No, 9 = Don't know
		PCL-R		1 = Yes, 0 = No, 9 = Don't know
		LSI-R		1 = Yes, 0 = No, 9 = Don't know
		LSI-CM		1 = Yes, 0 = No, 9 = Don't know
Other, specify		1 = Yes, 0 = No, 9 = Don't know		

	Other, specify	1 = Yes, 0 = No, 9 = Don't know	
<i>If yes, what level of skills are required?</i>	HCR-20	2 = Expert, 1 = General, 0 = None	
	VRAG	2 = Expert, 1 = General, 0 = None	
	VRAS	2 = Expert, 1 = General, 0 = None	
	PCL-R	2 = Expert, 1 = General, 0 = None	
	LSI-R	2 = Expert, 1 = General, 0 = None	
	LSI-CM	2 = Expert, 1 = General, 0 = None	
	Other, specify	2 = Expert, 1 = General, 0 = None	
	Other, specify	2 = Expert, 1 = General, 0 = None	
STANDARD 2A	Tool	Rating	Please state your vies about the accessibility and usefulness of training in these tools:
How easily accessible is the training to teach skills in the use of this tool?	HCR-20	2=Very, 1=Moderately, 0=Not at all	
	VRAG	2=Very, 1=Moderately, 0=Not at all	
	VRAS	2=Very, 1=Moderately, 0=Not at all	
	PCL-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-CM	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
STANDARD 3	Tool	Rating	Please add any comments about the ease (or otherwise) of this tool:
How easy (i.e., appropriateness of language, application, procedure of assessment) is this tool to use?	HCR-20	2=Very, 1=Moderately, 0=Not at all	
	VRAG	2=Very, 1=Moderately, 0=Not at all	
	VRAS	2=Very, 1=Moderately, 0=Not at all	
	PCL-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-CM	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
STANDARD 4	Tool	Rating¹⁰	Please give your comments on the appropriateness of the information sources needed to administer this tool:
Does this tool require consideration of a wide range of information from a wide range of sources, in other words is a multi-modal, multi-information information source needed?	HCR-20	2= Yes, 1= No	
	VRAG	2= Yes, 1= No	
	VRAS	2= Yes, 1= No	
	PCL-R	2= Yes, 1= No	
	LSI-R	2= Yes, 1= No	
	LSI-CM	2= Yes, 1= No	
	Other, specify	2= Yes, 1= No	
	Other, specify	2= Yes, 1= No	
STANDARD 5	Tool	Rating	Please add your comments about the time taken to administer this tool
How appropriate is the time required to administer this assessment?	HCR-20	3 = About right, 2 = Too long 1 = Too short	
	VRAG	3 = About right, 2 = Too long 1 = Too short	
	VRAS	3 = About right, 2 = Too long 1 = Too short	
	PCL-R	3 = About right, 2 = Too long 1 = Too short	
	LSI-R	3 = About right, 2 = Too long 1 = Too short	
	LSI-CM	3 = About right, 2 = Too long 1 = Too short	
	Other, specify	3 = About right, 2 = Too long 1 = Too short	
	Other, specify	3 = About right, 2 = Too long 1 = Too short	

¹⁰ These ratings were collapsed into about right (2) or too many/too few (1) for data analysis

STANDARD 6	Tool	Rating	Please add your comments regarding the usefulness of the tools in identifying key aspects about the offender's personal history
How comprehensive is the information gathered about the offender's personal background (i.e. their social history, relationship history, educational and work history etc)?	HCR-20	3=Very, 2=Moderately, 1=Not at all	
	VRAG	3= Very, 2=Moderately, 1=Not at all	
	VRAS	3= Very, 2=Moderately, 1=Not at all	
	PCL-R	3= Very, 2=Moderately, 1=Not at all	
	LSI-R	3= Very, 2=Moderately, 1=Not at all	
	LSI-CM	3= Very, 2=Moderately, 1=Not at all	
	Other, specify	3= Very, 2=Moderately, 1=Not at all	
Other, specify	3= Very, 2=Moderately, 1=Not at all		
STANDARD 7	Tool	Rating	Please add your comments on the usefulness of the tools on helping delineate offending behaviour patterns
Does the information obtained allow for a comprehensive analysis of offending behaviour i.e. the circumstances leading to the offending?	HCR-20	2= Yes, 1= No	
	VRAG	2= Yes, 1= No	
	VRAS	2= Yes, 1= No	
	PCL-R	2= Yes, 1= No	
	LSI-R	2= Yes, 1= No	
	LSI-CM	2= Yes, 1= No	
	Other, specify	2= Yes, 1= No	
	Other, specify	2= Yes, 1= No	

STANDARD 8	Tool	Rating	Please add your comments about the appropriateness of the number of risk factors covered in the tools, e.g. are any key risk factors missing?
How would you rate the number of risk factors covered in this tool?	HCR-20	1 =Too many, 2=Too few, 3=About right	
	VRAG	1 =Too many, 2=Too few, 3=About right	
	VRAS	1 =Too many, 2=Too few, 3=About right	
	PCL-R	1 =Too many, 2=Too few, 3=About right	
	LSI-R	1 =Too many, 2=Too few, 3=About right	
	LSI-CM	1 =Too many, 2=Too few, 3=About right	
	Other, specify	1 =Too many, 2=Too few, 3=About right	
	Other, specify	1 =Too many,2=Too few, 3=About right	
STANDARD 9	Tool	Rating	Comments
How would you rate the emphasis on static factors?	HCR-20	1 = too much, 2= too little, 3=about right	
	VRAG	1 = too much, 2=too little, 3=about right	
	VRAS	1 = too much, 2=too little, 3=about right	
	PCL-R	1 = too much, 2=too little, 3=about right	
	LSI-R	1 = too much, 2=too little, 3=about right	
	LSI-CMI	1 = too much, 2=too little, 3=about right	
	Other, specify	1 = too much, 2=too little, 3=about right	
	Other, specify	1 = too much, 2=too little, 3=about right	
STANDARD 10	Tool	Rating	Comments
How would you rate the emphasis on dynamic factors?	HCR-20	1=too much, 2=too little, 3=about right	
	VRAG	1 = too much, 2= too little, 3 = about right	
	VRAS	1 = too much,2=too little, 3=about right	
	PCL-R	1 = too much, 2= too little, 3=about right	
	LSI-R	1 = too much, 2=too little, 3=about right	
	LSI-CM	1 = too much, 2=too little, 3=about right	
	Other, specify	1 = too much, 2=too little, 3=about right	
	Other, specify	1 = too much, 2=too little, 3=about right	
STANDARD 11	Tool	Rating	Please add your comments about the appropriateness and usefulness of the risk factor definitions
How well defined and explained are the risk factors?	HCR-20	2=Very, 1=Moderately, 0=Not at all	
	VRAG	2=Very, 1=Moderately, 0=Not at all	
	VRAS	2=Very, 1=Moderately, 0=Not at all	
	PCL-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-CM	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
STANDARD 12	Tool	Rating	Please add your comments on the user-friendliness of the tools
How easy to follow are the scoring/rating guidelines for the risk factors?	HCR-20	2=Very, 1=Moderately, 0=Not at all	
	VRAG	2=Very,1=Moderately, 0=Not at all	

	VRAS	2=Very, 1=Moderately, 0=Not at all	
	PCL-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-CM	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
STANDARD 13	Tool	Rating	Please let us know what you think of the emphasis on the protective factors in this tool?
How well does this tool describe and identify protective factors?	HCR-20	2=Very, 1=Moderately, 0=Not at all	
	VRAG	2=Very, 1=Moderately, 0=Not at all	
	VRAS	2=Very, 1=Moderately, 0=Not at all	
	PCL-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-CM	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
STANDARD 14	Tool	Rating	Please give your views about the utility of the tools in developing a risk formulation
How well does this tool assist in the preparation of a risk formulation (i.e. the process of knitting together the various risk factor to provide a narrative (or story) about the person's risk?	HCR-20	2=Very, 1=Moderately, 0=Not at all	
	VRAG	2=Very, 1=Moderately, 0=Not at all	
	VRAS	2=Very, 1=Moderately, 0=Not at all	
	PCL-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-R	2=Very, 1=Moderately, 0=Not at all	
	LSI-CM	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	
	Other, specify	2=Very, 1=Moderately, 0=Not at all	

STANDARD 15	Tool	Rating	Please give your opinions on the extent to which the tools help identify plausible risk scenarios
How well does this tool assist in identifying plausible risk scenarios?	HCR-20	2 = Very ,1=Moderately, 0 = Not at all	
	VRAG	2 = Very, 1=Moderately, 0 = Not at all	
	VRAS	2 = Very, 1=Moderately, 0=Not at all	
	PCL-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-CM	2 = Very, 1=Moderately, 0 =Not at all	
	Other, specify	2 = Very, 1=Moderately, 0=Not at all	
	Other, specify	2 = Very, 1=Moderately, 0=Not at all	
STANDARD 16	Tool	Rating	Please describe your views on the usefulness of the tools you use in helping develop a risk management plan
Overall, how good would you rate the usefulness of this tool in facilitating the developing of a risk management plan?	HCR-20	2 = Very ,1=Moderately, 0 = Not at all	
	VRAG	2 = Very, 1=Moderately, 0 = Not at all	
	VRAS	2 = Very, 1=Moderately, 0=Not at all	
	PCL-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-CM	2 = Very, 1=Moderately, 0 =Not at all	
	Other, specify	2 = Very, 1=Moderately, 0=Not at all	
	Other, specify	2 = Very, 1=Moderately, 0=Not at all	
STANDARD 17	Tool	Rating	Please provide your comments on how well the tools you use help identify key treatment targets
How well does this tool help identify treatment needs for risk management?	HCR-20	2 = Very ,1=Moderately, 0 = Not at all	
	VRAG	2 = Very, 1=Moderately, 0 = Not at all	
	VRAS	2 = Very, 1=Moderately, 0=Not at all	
	PCL-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-CM	2 = Very, 1=Moderately, 0 =Not at all	
	Other, specify	2 = Very, 1=Moderately, 0=Not at all	
	Other, specify	2 = Very, 1=Moderately, 0=Not at all	
STANDARD 18	Tool	Rating	Please detail what you think about the tools utility in identifying monitoring needs
How well does this tool help identify monitoring needs for risk management?	HCR-20	2 = Very ,1=Moderately, 0 = Not at all	
	VRAG	2 = Very, 1=Moderately, 0 = Not at all	
	VRAS	2 = Very, 1=Moderately, 0=Not at all	
	PCL-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-R	2 = Very, 1=Moderately, 0=Not at all	

		all	
	LSI-CM	2 = Very, 1=Moderately, 0 =Not at all	
	Other, specify	2 = Very, 1=Moderately, 0=Not at all	
	Other, specify	2 = Very, 1=Moderately, 0=Not at all	
STANDARD 19	Tool	Rating	If you use more than one tool to conduct a risk assessment, please comment on your experience of convergence between the conclusions reached
How well do the conclusions about risk generated from this tool correspond to conclusions from other tools that you use?	HCR-20	2 = Very ,1=Moderately, 0 = Not at all	
	VRAG	2 = Very, 1=Moderately, 0 = Not at all	
	VRAS	2 = Very, 1=Moderately, 0=Not at all	
	PCL-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-CM	2 = Very, 1=Moderately, 0 =Not at all	
	Other, specify	2 = Very, 1=Moderately, 0=Not at all	
Other, specify	2 = Very, 1=Moderately, 0=Not at all		
STANDARD 20	Tool	Rating	Please explain your rating and give any other comments
Overall, how confident are you in the risk judgements reached by you when using this tool?	HCR-20	2 = Very ,1=Moderately, 0 = Not at all	
	VRAG	2 = Very, 1=Moderately, 0 = Not at all	
	VRAS	2 = Very, 1=Moderately, 0=Not at all	
	PCL-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-R	2 = Very, 1=Moderately, 0=Not at all	
	LSI-CM	2 = Very, 1=Moderately, 0 =Not at all	
	Other, specify	2 = Very, 1=Moderately, 0=Not at all	
Other, specify	2 = Very, 1=Moderately, 0=Not at all		
PLEASE ADD ANY OTHER COMMENTS BELOW:			

Would you be willing to take part in a further interview? Yes / No

THANK YOU FOR YOUR TIME