Violence risk prediction

Clinical and actuarial measures and the role of the Psychopathy Checklist

M. DOLAN and M. DOYLE

Background Violence risk prediction is a priority issue for clinicians working with mentally disordered offenders.

Aims To review the current status of violence risk prediction research.

Method Literature search (Medline). Key words: violence, risk prediction, mental disorder.

Results Systematic/structured risk assessment approaches may enhance the accuracy of clinical prediction of violent outcomes. Data on the predictive validity of available clinical risk assessment tools are based largely on American and North American studies and further validation is required in British samples. The Psychopathy Checklist appears to be a key predictor of violent recidivism in a variety of settings.

Conclusions Violence risk prediction is an inexact science and as such will continue to provoke debate. Clinicians clearly need to be able to demonstrate the rationale behind their decisions on violence risk and much can be learned from recent developments in research on violence risk prediction.

Declaration of interest None.

Violence risk assessment and management are key components of clinical practice (Monahan, 1992). In the UK, however, the adequacy and accuracy of risk prediction has been questioned in several inquiries into serious incidents involving mentally disordered patients (Reed, 1997), and there is a growing emphasis on systematising protocols for risk assessment and management. More recently, the Government highlighted the issue of risk in those with ‘dangerous severe personality disorders’ (DSPD) and outlined new proposals for dealing with this challenging but ill-defined group. Risk prediction is once again high on the public, political and clinical agenda. Here we review progress on violence risk assessment (particularly, violent recidivism) in the mentally disordered. We discuss recent developments in systematic violence risk assessment, focusing on the Psycho-pathy Checklist (PCL) as a predictor, and examine data from key meta-analytic studies in the field.

HISTORY OF VIOLENCE PREDICTION

Predicting future risk of violent behaviour has a long and difficult history. Before 1966 relatively little attention was paid to how well clinicians assessed risk. The Baxstrom v. Herald (1966) ruling in the USA (which resulted in the release or transfer from maximum security hospitals of 966 patients to the community or to lower security) was a notable landmark in risk assessment history. Steadman & Coccoza (1974) reported on the 4-year outcomes of this cohort and found that only 20% had been reconvicted, the majority for non-violent offences. Throughout the 1970s several other studies reported in the literature fuelled the notions that clinicians had little expertise in predicting violent outcomes (e.g. Coccoza & Steadman, 1976; Thornberry & Jacoby, 1979).

Monahan (1984) reviewed these ‘first generation’ studies and concluded that “the upper bound level of accuracy that even the best risk assessment technology could achieve was of the order of 0.33”. He reported that the best predictors of violence among the mentally disordered were the same demographic factors that predicted violence among non-disordered people, and that the poorest predictors were psychological factors such as diagnosis or personality traits. Subsequent studies, however, challenged these conclusions, particularly those demonstrating links between rates of violent offending and specific clinical diagnoses (e.g. Taylor, 1982; Binder & McNeil, 1988). The recent MacArthur Violence Risk Assessment Study (VRAS; Monahan et al, 2000) also highlights the significance of clinical factors such as substance misuse and psychopathy as assessed by Hare’s (1991) criteria, in the prediction of violent outcomes in non-forensic psychiatric patients discharged from hospital (see Steadman et al, 1994; 1998).

GENERAL ISSUES IN VIOLENCE RISK PREDICTION

Clinical v. research perspectives

Clinicians have traditionally assessed violence risk on an individual basis, using a case formulation approach, i.e. ‘unaided clinical judgement’. Until recently, however, research tended to focus on the accuracy of risk prediction variables in large, often heterogeneous, populations using relatively static actuarial predictors. These divergent approaches have resulted in debate over the merits of clinical v. actuarial approaches and their relevance to risk prediction for groups v. individuals. Furthermore, they have resulted in different perceptions about the relative contribution of clinical items in risk prediction scales in forensic and non-forensic settings. The clinical v. actuarial debate, however, has also led to the development of violence risk prediction instruments which adopt a combined approach and recognise the importance of both static actuarial variables and the clinical/risk management items that clinicians normally take into account in risk assessments of individuals. The latter approach appears to be a first step in bridging the gap between clinical and actuarial measures, and between group
and individual risk assessment approaches. In the following sections some of the key issues pertaining to clinical, actuarial and structured clinical assessments in violence risk prediction are discussed.

**Approaches to violence risk prediction**

**Unaided clinical risk assessment**

In clinical practice, assessments of the risk of dangerousness or violence in an individual are usually based solely on unaided clinical judgement. The unstructured clinical judgement approach to risk assessment has been criticised on a number of grounds, including low interrater reliability, low validity and a failure to specify the decision-making process (Monahan & Steadman, 1994; Webster et al, 1997a), and inferior predictive validity compared to actuarial predictions (Meehl, 1954; Lidz et al, 1993; Mossman, 1994). Others, however, consider that clinical approaches offer the advantages of flexibility and an emphasis on violence prevention (Snowden, 1997; Hart, 1998a). Buchanan (1999) also suggests that clinical approaches, if they focus on mechanisms through which violence occurs, may enhance the validity of risk assessment.

Clinicians may be better than was believed in the immediate aftermath of Baxstrom studies (Cocozza & Steadman, 1976). Gardner et al (1996), for example, showed that while actuarial measures were better than clinical ratings, clinical ratings were better than chance. Studies also showed that the accuracy of prediction can be enhanced when clinicians consider the context in which violence occurs in their patients (Mulvey & Lidz, 1985). Recently, Fuller & Cowan (1999) showed that multi-disciplinary team consensus predictions of risk were comparable with actuarially based schedules over similar time-scales.

**Actuarial methods**

Actuarial methods allow assessors to make decisions based on data which can be coded in a predetermined manner (Meehl, 1954). Decisions are made according to rules, and focus on relatively small numbers of risk factors that are known, or are thought, to predict violence across settings and individuals. For diverse samples and contexts, these factors tend to be static (e.g. demographic variables). Actuarial approaches undoubtedly improve the consistency of risk assessment, but Hart (1988a,b) argues that they tend to ignore individual variations in risk, overfocus on relatively static variables, fail to prioritise clinically relevant variables and minimise the role of professional judgement.

Despite these criticisms, actuarial risk assessment tools have been utilised for some time in US penal settings to help in making decisions about parole. Examples include the Base Expectancy Score (Gottfredson & Bonds, 1961), the Level of Supervision Inventory (Andrews, 1982), the Salient Factor Score (revised) (Hoffman, 1983), and the Statistical Information on Recidivism (SIR) scale (Nuffield, 1989). In the UK, similar measures have been developed to produce ‘risk of reconviction’ scores for prisoners before the parole board (Copas et al, 1996).

**Structured clinical judgement**

Structured clinical judgement represents a composite of empirical knowledge and clinical/professional expertise. Webster et al (1997a), who are the leading proponents of this model, argue that clinical violence risk prediction can be improved significantly if:

(a) assessments are conducted using well-defined published schema;
(b) agreement between assessors is good, through their training, knowledge and expertise;
(c) prediction is for a defined type of violent behaviour over a set period;
(d) violent acts are detectable and recorded;
(e) all relevant information is available and substantiated;
(f) actuarial estimates are adjusted only if there is sufficient justification.

Several instruments have been developed along these lines to assess risk of violence in clinical contexts. These include the Historical/Clinical/Risk Management 20-item (HCR–20) scale (Webster et al, 1997b) the Spousal Assault Risk Assessment guide (Kropp et al, 1995) and the Sexual Violence Risk (SVR–20) scale (Boer et al, 1997) (see Douglas & Cox (1999) for an in-depth review of these instruments).

Hart (1998a,b) suggests that structured clinical instruments like the above promote systematic data collection based on sound scientific knowledge, yet allow flexibility in the assessment process. He also argues that, unlike strict actuarial measures, they encourage clinicians to use professional discretion.

**Violence risk prediction in clinical settings**

A number of violence risk prediction tools have been developed and introduced into clinical settings in North America. Among these, the Dangerous Behaviour Rating Scale (DBRS: Menzies et al, 1985a,b), the Violence Risk Appraisal Guide (VRAG: Harris et al, 1993) and the HCR–20 (Webster et al, 1997b) have received most attention. The last two instruments contain an item assessing psychopathy, based on the Psychopathy Checklist (revised) (PCL–R; Hare, 1991). The PCL–R itself, however, has also been shown to have reasonable predictive validity in determining future violence, and will also be discussed in some detail.

Before describing these tools and their predictive validity it may be useful to describe one of the more recent statistical measures which is frequently cited in the literature on the accuracy of violence risk prediction.

**Statistical measures for assessing predictive accuracy**

There are several measures available to evaluate the predictive accuracy of different tools in studies on violence risk prediction in large cohorts (see Appendix). Receiver operator characteristics (ROCs), which yield an area under the curve (AUC) measure, however, appear to be the preferred method, and much of the recent literature on predictive accuracy quotes ROC–AUC data. ROCs are particularly useful as they provide data which are fairly independent of the base rates of violence in a given population (Mossman, 1994). The ROC–AUC parameter, which can range from 0 to 1, provides information which is similar to that yielded by the more commonly used effect size estimate (such as Cohen’s d; see Cohen, 1988; 1992) and can be used to compare accuracy between instruments. Figure 1 shows an example of a ROC curve. The straight line on the ROC curve corresponds to the line of no information, i.e. no better than random prediction (AUC=0.5). Instruments or clinicians which distinguish violent from non-violent patients with nearly perfect accuracy would have ROC–AUCs approaching 1.0. In general, Cohen's
d > 0.30 or ROC–AUCs > 0.75 are considered large effect sizes. ROC curves also give an indication of the trade-offs between specificity and sensitivity at different decision thresholds or cut-off scores on measures.

CLINICAL RISK ASSESSMENT TOOLS AND THEIR PREDICTIVE VALIDITY

The Dangerous Behaviour Rating Scale (DBRS)

The DBRS (Menzies et al, 1985a) was initially developed in conjunction with clinicians from a model devised by Megaree (1976). The item list comprised 18 ratings of personality, situation, lifestyle-related variables, and interview-specific factors believed to relate to risk. Reports on the predictive validity of the DBRS, however, indicate that it has met with little success. For example, in Menzies and Webster’s (1995) 6-year follow-up study of 162 Canadian mentally disordered persons tracked across institutions and the community, comparatively little association was found between actuarial and clinical risk factors and follow-up violence outcome data. Professionals were no more accurate than non-clinical raters, and the DBRS items showed less predictive power than attributes such as age, violent history or employment status. The DBRS is now rarely used, and Webster et al (1997a) argue that the limitations of this instrument reflect the limited literature on which it was based.

The Violence Risk Appraisal Guide (VRAG)

The VRAG (Harris et al, 1993) incorporates 12 items, which are scored on the basis of a weighting procedure developed on a calibration sample of 618 males charged with severe violent crimes. The items are listed in Table 1. The variable with the heaviest weighting is the PCL–R score. Using ROCs, Rice & Harris (1995) analysed the data from several populations of offenders independent of the calibration sample, and found that the VRAG predicted violent recidivism with AUCs of 0.75, 0.74 and 0.74 for 3.5, 6 and 10 years, respectively. Using more restrictive definitions of violent recidivism, the relevant normalised ROC gave a mean AUC of 0.73. Later reports, however, suggest that the VRAG is less valuable in predicting violent sexual recidivism in paedophile sex offender populations (Rice & Harris, 1997). The VRAG has been criticised because of its reliance on relatively static factors, and Webster et al (1994) now recommend that it be supplemented with a clinical checklist to produce a ‘violence prediction scheme’.

Psychopathy Checklist (Revised) (PCL–R)

The 20-item PCL–R (Hare, 1991), which is scored on a three-point scale, was originally devised as a research tool for operationalising psychopathy (see Table 1). Scores range from 0 to 40, with a cut-off of > 30 reflecting a prototypical psychopath. The PCL–R has been shown to have good psychometric properties (Hare, 1998). It has a stable factor structure (Hare, 1991), in which factor 1 taps interpersonal/affective traits, while factor 2 reflects the behavioural components of psychopathy. Cooke & Mitchie (1998), however, have recently presented a three-factor model of psychopathy using confirmatory factor analytic procedures. A number of studies demonstrate its utility as a risk assessment tool, in identifying recidivists and predicting violence in North American forensic and prison samples (Hart, 1998a). As yet, there are few data on its predictive validity in European samples, although recent work by Grann et al (1999) suggests that the PCL–R scores were the best predictor of violent recidivism 2 years after release from confinement in Swedish offenders with personality disorder (AUC=0.72).

The PCL–R has been supplemented by the 12-item screening version (PCL–SV: Hart et al, 1995) (Table 1). It has similar psychometric properties to the PCL–R, with scores ranging from 0–24 (cut-off at 18). The PCL–SV has been shown to have good predictive validity for institutional violence (Hill et al, 1996; Grann, 1998) and community violence (MacArthur study, Monahan et al, 2000).

As some psychopathy checklist items may be linked to outcome variables of interest (such as violence), researchers have used different methods to control for this potential confounder, including statistical control for past criminal activity or removing potentially confounding items from the checklist in the analysis.

The Historical/Clinical/Risk Management 20-item (HCR–20) scale

The HCR–20 (Webster et al, 1997b) contains 10 historical (H-10) items (two of which address the issue of personality dysfunction), five clinical (C-5) items, and five risk management (R-5) items (Table 1). It is scored in a similar manner to the PCL–R and shows good interrater reliability (Webster et al, 1997a). When the personality disorder variable is removed, H-10 items show significant correlations with on-ward violence (unpublished, 1996; details available from the first author upon request). In two studies, the H-10 items showed stronger correlations with violent outcome than the C-5 scales (see Douglas, 1996; unpublished, 1996); this may reflect the lack of inclusion of interview data in these retrospective studies.

Table 2 lists some key HCR–20 studies examining the predictive validity using ROC–AUC data. While the studies are limited to a small group of North American researchers, the data generally show ‘better than chance’ relationships between HCR–20 scores and violent outcomes. As yet, no studies have been published of the reliability and validity of this instrument in UK samples, although such work is in progress (details available from the first author upon request).

Comparison of actuarial risk scales

The PCL/PCL–R has generally been found to be superior to other classical actuarial risk scales on indices of recidivism or violent recidivism (Harris et al, 1993; Rice & Harris, 1995; Zamble & Palmer, 1996; see also
<table>
<thead>
<tr>
<th>PCL-R (Hare, 1991)</th>
<th>PCL-SV (Hart et al., 1995)</th>
<th>HCR–20† (V2) (Webster et al., 1997b)</th>
<th>VRAG (Harris et al., 1993)</th>
<th>MacArthur VRAS (Monahan et al., 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Predictors of community</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>violence</td>
</tr>
<tr>
<td>1</td>
<td>Glibness/superficial charm (I)</td>
<td>1 Superficial (I)</td>
<td>H1 Previous violence</td>
<td>1 PCL–SV score</td>
</tr>
<tr>
<td>2</td>
<td>Grandiose sense of self-worth (I)</td>
<td>2 Grandiose (I)</td>
<td>H2 Young age at first violent incident</td>
<td>2 Elementary school maladjustment</td>
</tr>
<tr>
<td>3</td>
<td>Need for stimulation/proneness to boredom (2)</td>
<td>3 Manipulative (I)</td>
<td>H3 Relationship instability</td>
<td>3 DSM–III diagnosis of personality disorder</td>
</tr>
<tr>
<td>4</td>
<td>Pathological lying (I)</td>
<td>4 Lacks remorse (I)</td>
<td>H4 Employment problems</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Conning/manipulative (I)</td>
<td>5 Lacks empathy (I)</td>
<td>H5 Substance misuse problems</td>
<td>4 Age at index offence</td>
</tr>
<tr>
<td>6</td>
<td>Lack of remorse or guilt (I)</td>
<td>6 Does not accept</td>
<td>H6 Major mental illness</td>
<td>5 Lived with both parents to age 16</td>
</tr>
<tr>
<td>7</td>
<td>Shallow affect (I)</td>
<td>7 Impulsive (2)</td>
<td>H7 Psychopathy (PCL–R/PCL–SV)</td>
<td>6 Failure on prior conditional release</td>
</tr>
<tr>
<td>8</td>
<td>Callous/lack of sympathy (I)</td>
<td>8 Poor behaviour controls (2)</td>
<td>H8 Early maladjustment</td>
<td>7 Non-violent offence score</td>
</tr>
<tr>
<td>9</td>
<td>Parasitic lifestyle (2)</td>
<td>9 Lacks (goals) (2)</td>
<td>H9 Personality disorder</td>
<td>8 Marital status</td>
</tr>
<tr>
<td>10</td>
<td>Poor behavioural controls (2)</td>
<td>10 Irresponsible (2)</td>
<td>H10 Prior supervision failure</td>
<td>9 DSM–III diagnosis of schizophrenia</td>
</tr>
<tr>
<td>11</td>
<td>Promiscuous sexual behaviour</td>
<td>11 Adolescent antisocial</td>
<td>C1 Lack of insight</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Early behavioural problems (2)</td>
<td>12 Adult antisocial behaviour (2)</td>
<td>C2 Negative attitudes</td>
<td>10 Victim injury (index offence)</td>
</tr>
<tr>
<td>13</td>
<td>Lack of realistic long-term goals (2)</td>
<td>13 Adult antisocial behaviour (2)</td>
<td>C3 Active symptoms of major mental illness</td>
<td>11 History of alcohol misuse</td>
</tr>
<tr>
<td>14</td>
<td>Impulsivity (2)</td>
<td></td>
<td>C4 Impulsivity</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Irresponsibility (2)</td>
<td></td>
<td>C5 Unresponsive to treatment</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Failure to accept responsibility</td>
<td></td>
<td>R1 Plans lack feasibility</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Many short-term marital relationships</td>
<td></td>
<td>R2 Exposure to destabilisers</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Juvenile delinquency (2)</td>
<td></td>
<td>R3 Lack of personal support</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Revocation of conditional release (2)</td>
<td></td>
<td>R4 Non-compliance with remediation attempts</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Criminal versatility</td>
<td></td>
<td>R5 Stress</td>
<td></td>
</tr>
</tbody>
</table>

1. H1–10 relate to history. CI–5 to clinical and RI–5 to risk.
(I), factor 1 loading; (2), factor 2 loading.
Hemphill et al. (1998). Rice and Harris (1995) compared the SIR scale and VRAG, and found significantly better prediction rates with the VRAG, although the SIR scale (contrary to initial perceptions: Nufield, 1982) also showed reasonable ROC-AUCs (0.69, 0.67 and 0.66 at 3.5-, 6- and 10-year follow-up). Zamble and Palmer (1996) compared the PCL–R, parole board decisions and the SIR scale, in 106 male offenders released from Canadian federal penitentiaries and found the PCL–R to be the most accurate at predicting reconviction or revocation of parole at a mean follow-up time of 30 months. Hemphill and Hare (1996) also compared the predictive validity of the PCL/PCL–R and several actuarial measures, and found that they performed similarly for general recidivism prediction, but that the PCL–R was significantly better for violent recidivism prediction.

Using ROCs, Grann (1998) compared the H–10 scale of the HCR–20 and VRAG in predicting reconviction for violence within 2 years of release, in a retrospective study of 293 violent offenders with personality disorders and 111 with schizophrenia. He reported that both scales performed better in the personality-disorder group but the H–10 did better than the VRAG in both groups of offenders. It is possible that historical/static variables may be relatively good predictors of violent recidivism in subjects with personality disorder, but clinical and risk management variables may be better predictors in populations with schizophrenia (Webster et al., 1997a; Grann, 1998; Strand et al., 1999).

General reviews and meta-analysis of studies of violence and recidivism

There have been four relatively recent meta-analytic studies of recidivism, including violent recidivism, and each differs in the studies included in it and the method of effect size determination.

Mossman (1994) extracted 58 data sets from 44 published studies dating from 1972 to 1993 on violence risk prediction, and examined prediction accuracy using ROCs. The studies included a broad range of subjects, settings, population sizes and clinical criteria for assessing violence, and Mossman acknowledges that conclusions can only be tentative. The median ROC–AUCs for all 58 datasets was 0.73, suggesting, overall, that clinicians were predicting violence more accurately than chance. However, short-term (1–7 days, AUC = 0.68) predictions were no more accurate than long-term (>12 months, AUC = 0.64). First-generation studies (before 1986) (AUC = 0.74) were less accurate than second-generation studies (after 1986) (mean AUC = 0.83), but the samples were extremely heterogeneous. Mossman suggests that clinicians were able to distinguish violent from non-violent patients with a “modest, better than chance level of accuracy”. Since this work was published, other reviews have concentrated on the issue of recidivism, particularly violent recidivism, which is generally perceived as a ‘harder’ outcome measure.

Bonta et al. (1998) conducted a meta-analysis of predictive longitudinal studies (1959–1995), to examine whether predictors of recidivism, including violent recidivism, for mentally disordered offenders were different from those for non-disordered offenders. Using 64 separate samples with 27 predictors for violent recidivism, they showed that criminal history variables were better predictors than clinical variables, using adjusted and transformed Pearson’s correlations to assess effect size (Zr). For violent recidivism, criminal history variables had the largest effect size (Zr = 0.15, p < 0.001), followed by personal demographics (Zr = 0.12, p < 0.001), deviant lifestyle (Zr = 0.08, p < 0.001) and clinical variables (Zr = −0.03, p < 0.01). A diagnosis of antisocial personality disorder was the most significant clinical predictor.

Role of the PCL/PCL–R and PCL–SV in risk prediction

The PCL–R and PCL–SV are currently believed to be some of the most reliable tools for assessing personality constructs likely to be relevant to violent risk prediction (Hart, 1998b). For this reason the PCL–SV was included in the MacArthur VRAS, where it was shown to have reasonable predictive validity for community violence (Monahan et al., 2000). Hemphill & Hare (1996) have also shown that the PCL/PCL–R, entered into a hierarchical multiple regression analysis with other demographic/clinical history variables, such as the HCR–20, have been shown to be the best predictors of violent recidivism. (Hart, 1998b).

Table 2: HCR–20 risk predictive validity studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Patient sample</th>
<th>Measures</th>
<th>Outcomes</th>
<th>Instrument</th>
<th>ROC–AUCs</th>
<th>Instrument</th>
<th>ROC–AUCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Douglas et al.</td>
<td>Non-forensic psychiatric</td>
<td>Community violence</td>
<td>Violent crime</td>
<td>HCR–20</td>
<td>0.80</td>
<td>PCL–SV</td>
<td>0.78</td>
</tr>
<tr>
<td>Nicholls et al.</td>
<td>Non-forensic psychiatric (males)</td>
<td>Community violence</td>
<td>Any violence</td>
<td>HCR–20</td>
<td>0.76</td>
<td>PCL–SV</td>
<td>0.68</td>
</tr>
<tr>
<td>Nicholls et al.</td>
<td>Non-forensic psychiatric (females)</td>
<td>Community violence</td>
<td>Physical violence</td>
<td>HCR–20</td>
<td>0.76</td>
<td>PCL–SV</td>
<td>0.73</td>
</tr>
<tr>
<td>Unpublished</td>
<td>Civil psychiatric patients</td>
<td>Community violence</td>
<td>Any violence</td>
<td>HCR–20</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grann (1998)</td>
<td>Forensic</td>
<td>Community violence</td>
<td>Violent arrest</td>
<td>HCR–20</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unpublished</td>
<td>Non-forensic psychiatric</td>
<td>In-patient violence</td>
<td>Violent arrest</td>
<td>HCR–20</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unpublished</td>
<td>Non-forensic psychiatric</td>
<td>In-patient violence</td>
<td>Violent arrest &amp; people with schizophrenia</td>
<td>H scale</td>
<td>0.71</td>
<td>VRAG</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>H scale</td>
<td>0.66</td>
<td>VRAG</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>H/C scales</td>
<td>0.57–0.65</td>
<td>PCL–SV</td>
<td>0.60–0.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>H/C scales</td>
<td>0.63</td>
<td>PCL–SV</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>H/C scales</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Details available from the first author upon request.

HCR–20, Historical/Clinical/Risk Management 20-item scale; PCL–SV, Psychopathy Checklist – Screening Version; ROC, receiver operator characteristics; AUC, area under curve.
adds significant incremental validity to the prediction of violence.

**Meta-analytic studies using the PCL/PCL–R/PCL–SV in risk prediction**

Salekin *et al.* (1996) examined all 18 available (published and unpublished) studies using the PCL/PCL–R between 1974 and 1995 and conducted a meta-analytic review, using an adaptation for effect size calculation from Rosenthal (1991). Separate analyses were conducted for violent recidivism. Despite the variation in cut-off scores on these instruments, Salekin *et al.* (1996) reported moderate to strong effect sizes (Cohen’s $d=0.55$ for criminality, $r=0.37$ and $d=0.79$ for violent recidivism; see Table 3). The largest effect sizes were reported in the study of institutional violence by Hill *et al.* (1996) using the PCL–SV. Although the study by Salekin *et al.* (1996) included a small number of postdictive studies (comparing assessment measures with previous violence) which may have inflated their reported mean effect sizes, they found no significant difference between postdictive (0.75) and predictive (0.79) effect sizes, on a separate analysis.

**Table 3** Studies utilizing Psychopathy Checklists (PCLs) to predict violent behaviour

<table>
<thead>
<tr>
<th>Study</th>
<th>Version</th>
<th>Sample</th>
<th>Validity</th>
<th>Outcome</th>
<th>Effect size (ES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salekin <em>et al.</em> (1996) review</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Forth *et al.* (1990)$^1$ | PCL | Maximum security youth detention centre | Postdict | Violent recidivism | $r=0.37$
| Forth *et al.* (1990)$^1$ | PCL | Maximum security youth detention centre | Prediction | Violent recidivism | $d=0.56$
| Forth *et al.* (1990)$^1$ | PCL | Maximum security youth detention centre | Postdict | Institutional violence | $d=1.04$
| Hare & McPherson (1984)$^1$ | PCL | Federal medium security prison | Postdict | Violent recidivism | $d=0.54$
| Harris *et al.* (1991)$^1$ | PCI | Therapeutic community programme | Prediction | Violent recidivism | $d=0.93$
| Heilbrun *et al.* (1998)$^1$ | PCL | Forensic psychiatric patients | Prediction | Institutional violence and/or aggression | $d=0.63$
| Hill *et al.* (1996)$^1$ | PCL–SV | Forensic psychiatric patients | Prediction | Institutional violence and/or aggression | $d=1.92$
| Kosson *et al.* (1990)$^1$ | PCL–R | Maximum security youth detention centre | Postdict | Violent recidivism | $d=0.42$
| Miller *et al.* (1994)$^1$ | PCL–R | Forensic treatment centre for sexual offenders | Postdict | Violent recidivism | $d=1.18$
| Quinsey *et al.* (1995)$^1$ | PCL–R | Forensic treatment centre for sexual offenders | Prediction | Violent recidivism | $d=0.7$
| Rice & Harris (1992)$^2$ | PCL | Forensic psychiatric patients | Prediction | Violent recidivism | $d=0.42$
| Unpublished (1992)$^2$ | PCL–R | Therapeutic community programme | Prediction | Violent recidivism | $d=0.72$
| Rice *et al.* (1990)$^1$ | PCL | Forensic treatment centre for sexual offenders | Prediction | Violent recidivism | $d=0.74$
| Serin (1991) | PCL | Federal medium security prison | Postdict | Violent recidivism | $d=0.71$
| Serin & Amos (1995)$^3$ | PCL–R | Federal medium security prison | Prediction | Violent recidivism | $d=0.58$
| Hemphill *et al.* (1998) review | | | | | |
| Hemphill *et al.* (1992)$^2$ | PCL | Criminal psychopaths | Prediction | Violent recidivism | $r=0.27$
| Harris *et al.* (1993)$^2$ | PCL | Forensic psychiatric patients | Prediction | Violent recidivism | $r=0.06$
| Heilbrun *et al.* (1998)$^2$ | PCL | Forensic psychiatric patients | Prediction | Violent recidivism | $r=0.34$
| Ross *et al.* (1992)$^2$ | PCL | French prison parolees | Prediction | Violent recidivism | $r=0.16$
| Serin & Amos (1995)$^3$ | PCL–R | Federal medium security prison | Prediction | Violent recidivism | $r=0.17$
| Studies outside North America | | | | | |
| Grann *et al.* (1999) | PCL–R | Forensic psychiatric evaluations (prison, hospital and probation) | Prediction | Violent recidivism | AUC of ROC at 2 years $=0.75$
| Unpublished (1999)$^3$ | PCL–SV | Medium security forensic | Prediction | In-patient violence | AUC of ROC at 3 months $=0.75$

1. Included in Salekin *et al.* (1996). ES = mean PCL/PCL–R score for recidivists-non-recidivists/pooled s.d. of two groups (mean Cohen’s $d=0.79$, $r=0.37$).
2. Included in Hemphill *et al.* (1998). ES = raw PCL mean/recidivism correlation coefficient to Fisher Z's averaging these values, then converting back to rs (mean Cohen’s $d=0.56$, $r=0.27$).
3. Further details available from the first author upon request.

SV, Screening Version; R, Revised; ROC, receiver operator characteristics, AUC, area under the curve.
Hemphill et al (1998) also conducted a meta-analysis of PCL/PCL–R studies in prediction of general/violent recidivism, but included only predictive studies and those with independent samples. The 1996 review by Salekin et al had included several same sample studies from the Oak Ridge group. Based on the five studies shown in Table 3 (1374 offenders) and more restrictive criteria, Hemphill et al (1998) reported a slightly lower mean effect size for violent recidivism ($r=0.27$, Cohen’s $d=0.56$). Overall the predictive validity of the PCL–R is moderately high (Hart, 1998a).

PUTTING SYSTEMATIC RISK ASSESSMENT INTO PRACTICE

Gardner et al (1996) suggest that clinicians may be averse to actuarial or structured clinical prediction instruments because they are impractical and too costly, and the analyses too complex. They developed a 'regression tree’ (i.e. structured sequences of yes/no answers that lead to classification of a case as high or low risk) and a two-stage screening process, which they showed was as accurate as traditional actuarial measures. Monahan et al (2000) also developed an Iterative Classification Tree (ICT) which successfully classified 77.6% of their sample as high or low risk, based on the variables shown in Table 2. ROCs for the ICT method were high (AUC = 0.82). In higher risk cases, Serin and Amos (1995) suggest a three-stage decision tree, which includes an analysis of ‘group base rates’ of violence, ‘individual base rate risk’ and ‘risk management variables’.

Decision or classification trees appear to be a useful means of streamlining violence risk assessments in large populations with relatively low base rates of violence. In smaller samples of high-risk patients or offenders, however, more in-depth batteries of relevant tools such as the PCL–R and HCR–20 will be required to assess future risk of violent recidivism.

SUMMARY

This review indicates that structured clinical judgement and systematic risk assessment scales should be used cautiously and judiciously. The assessment tools chosen, and how to interpret the scores, will largely be influenced by the populations or settings and the questions we want answered. The MacArthur project group have developed a classificatory tree method for assessing risk in community samples. For clinicians working in forensic or penal settings, much can be learned from the studies demonstrating the predictive accuracy of tools such as the PCL–R and HCR–20. Future British studies should aim to establish the validity of North American risk assessment tools in a range of populations and settings. Efforts should also be made to enhance the predictive validity of these tools by the addition of physiological measures and assessments of neurocognitive function and how individuals process emotional information. Violence prediction will never be entirely accurate, given that violence itself is a complex concept. Clinicians need to be aware of the benefits and limitations of current assessment tools and how scores on these measures might be used or interpreted by other agencies (see Hare (1998) for a commentary on the use and misuse of the PCL–R).

APPENDIX – TERMINOLOGY

- **True positive, TP**: predicted risk, outcome violent
- **False positive, FP**: predicted risk, outcome not violent
- **False negative, FN**: predicted no risk, outcome violent
- **True negative, TN**: predicted no risk, outcome not violent
- **Base rate, BR = (TP+FN)/(TP+FP+FN+TN)**: proportion of violent individuals in a population
- **Selection ratio, SR = (TP+FP)/(TP+FP+FN+TN)**: cut-off scores used to classify individuals as violent
- **Correct fraction, CF = (TP+TN)/(TP+FP+FN+TN)**
- **Sensitivity = true positive rate, TPR = TP/(TP+FN)**
- **Specificity = true negative rate, TNR = TN/(TN+FP)**
- **Positive predictive power = proportion of individuals designated a risk who in fact are a risk**
- **Negative predictive power = proportion of individuals identified as low risk and who in fact are low risk**
- **False positive rate, FPR = (1—specificity) = FP/(FP+TN)**
- **Risk ratio = TPR/FPR**
- **Odds ratio = (TP+TN)/(FP+FN)**: odds that person predicted to fail will do so/odds a person not predicted to fail will do so
- **Relative improvement over chance, ROC = CF — ((BR)(SR)+(1—BR)(1—SR))**

REFERENCES

CLINICAL IMPLICATIONS

- Structured/systematic approaches to violence risk prediction provide a more accurate and transparent record of the risk factors considered and the rationale behind decisions taken.
- Risk assessment batteries need to be streamlined and adapted to suit the population under study and the key questions asked.
- The Psychopathy Checklist and its derivatives appear to be significant predictors of violence in forensic and non-forensic settings.

LIMITATIONS

- There are a limited number of studies on the reliability and validity of published risk assessment tools outside the centres in which they were developed.
- The literature on accuracy of violence risk is predominately postdictive rather than predictive, and much needs to be done to improve current violence prediction accuracy, using prospective study designs.
- The lack of uniformity in the statistical procedures used to assess predictive accuracy, and the variation in choice of cut-off scores on risk prediction tools, make comparisons between studies difficult. The reporting of receiver operator characteristic data should improve this situation.

---


Correspondence: Dr Mairead Dolan, Edendfield Centre, Mental Health Services of Salford

--- (First received 26 October 1999, final revision 12 April 2000, accepted 13 April 2000)


