European Psychiatry 52 (2018) 47–53

Contents lists available at ScienceDirect

European Psychiatry

journal homepage: http://www.europsy-journal.com

Use of risk assessment instruments to predict violence in forensic psychiatric hospitals: a systematic review and meta-analysis

Taanvi Ramesh[a], Artemis Igoumenou[b], Maria Vazquez Montes[c], Seena Fazel[a,*]

[a] Department of Psychiatry, University of Oxford, Oxford, UK
[b] Consultant Forensic Psychiatrist, Barnet Enfield and Haringey Mental Health NHS Trust, UK
[c] Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

ARTICLE INFO

Article history:
Received 14 November 2017
Received in revised form 24 February 2018
Accepted 28 February 2018
Available online 4 April 2018

ABSTRACT

Background and Aims: Violent behaviour by forensic psychiatric inpatients is common. We aimed to systematically review the performance of structured risk assessment tools for violence in these settings.
Methods: The nine most commonly used violence risk assessment instruments used in psychiatric hospitals were examined. A systematic search of five databases (CINAHL, Embase, Global Health, PsycINFO and PubMed) was conducted to identify studies examining the predictive accuracy of these tools in forensic psychiatric inpatient settings. Risk assessment instruments were separated into those designed for imminent (within 24 hours) violence prediction and those designed for longer-term prediction. A range of accuracy measures and descriptive variables were extracted. A quality assessment was performed for each eligible study using the QUADAS-2. Summary performance measures (sensitivity, specificity, positive and negative predictive values, diagnostic odds ratio, and area under the curve value) and HSROC curves were produced. In addition, meta-regression analyses investigated study and sample effects on tool performance.
Results: Fifty-two eligible publications were identified, of which 43 provided information on tool accuracy in the form of AUC statistics. These provided data on 78 individual samples, with information on 6,840 patients. Of these, 35 samples (3,306 patients from 19 publications) provided data on all performance measures. The median AUC value for the wider group of 78 samples was higher for imminent tools (AUC 0.83; IQR: 0.71–0.85) compared with longer-term tools (AUC 0.68; IQR: 0.62-0.75). Other performance measures indicated variable accuracy for imminent and longer-term tools. Meta-regression indicated that no study or sample-related characteristics were associated with between-study differences in AUCs.
Interpretation: The performance of current tools in predicting risk of violence beyond the first few days is variable, and the selection of which tool to use in clinical practice should consider accuracy estimates. For more imminent violence, however, there is evidence in support of brief scalable assessment tools.
© 2018 The Authors. Published by Elsevier Masson SAS. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. INTRODUCTION

Violence in inpatient psychiatric wards is a major problem for health services, with effects on patient and staff psychiatric morbidity [1], wider implications on stigma for patients and recruitment in psychiatric hospitals, alongside costs associated with injury, staff sickness, and potential litigation by victims. There are higher reported rates of violence on forensic psychiatric wards compared to general psychiatry; a review of nearly 70,000 psychiatric patients from 122 studies in high income countries found that 48% of patients on forensic wards were violent over a mean follow-up of 31 months, which was almost double that for acute psychiatric wards (26%, mean time period: 19 months) and over two-fold that for other less acute psychiatric inpatient settings (22%, mean time period: 16 months) [2].

Despite its importance, few instruments have been designed for the prediction of violence specifically for inpatient populations. Current guidelines from the National Institute for Health and Care Excellence (NICE) [3] in England recommend the use of the Brøset Violence Checklist (BVC) [4,5] or the Dynamic Appraisal of Situational Aggression (DASA) [6] for the prediction of inpatient violence, although US and Australasian guidelines do not appear to recommend any such tools for acute management of schizophrenia inpatients [7,8].

* Corresponding author at: Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK.
E-mail address: seena.fazel@psych.ox.ac.uk (S. Fazel).

http://dx.doi.org/10.1016/j.eurpsy.2018.02.007
0924-9338/© 2018 The Authors. Published by Elsevier Masson SAS. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.1016/j.eurpsy.2018.02.007 Published online by Cambridge University Press

Previous work has typically combined forensic psychiatric patients with other psychiatric populations and prisoners when assessing the predictive accuracy of risk assessment instruments [9–12]. A meta-review of violence risk assessment systematic reviews and meta-analyses found that 90% of reviews published before 2010 included mixed samples of different populations, and thus the overall findings may not be informative to specific patient groups [13]. In addition, inpatient or institutional violence is often grouped together with community or offending outcomes in reviews [10,11,12]. As violence base rates and possible interventions, and also the strength of risk factors, are different between inpatients and community-dwelling individuals, there is a need for a review specifically on inpatient violence.

Thus, we have aimed to systematically review and meta-analyse the performance of structured risk assessment instruments used to predict inpatient violence in forensic psychiatric samples. In addition, we have investigated sources of variation between individual studies using meta-regression analyses.

2. METHODS

2.1. Review protocol

This review followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement [14]. A review protocol was published on PROSPERO on 23/11/16: (https://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016049789).

2.2. Risk assessment tools

Based on recent reviews and questionnaire surveys [15–17], the 11 most commonly used instruments for forensic inpatient violence risk prediction were identified. Actuarial instruments included the Brøset Violence Checklist (BVC) [4,5], the Classification of Violence Risk (COVR) [18,19], the Dynamic Appraisal of Situational Aggression (DASA) [6], the Level of Service Inventory-Revised (LSI-R) [20], the Psychopathy Checklist Revised (PCL-R) [21], the Psychopathy Checklist Screening Version (PCL:SV) [22], the Violence Risk Appraisal Guide (VRAG) [23,24] and the Violence Risk Scale (VRS) [25]. Structured professional judgement (SPJ) tools included the Historical Clinical Risk Management-20 (HCR-20) [26,27], the Short-Term Assessment of Risk and Treatability (START) [28,29] and the Violence Risk Screening-10 (V-RISK-10) [30,31]. Tools developed specifically for sexual violence were not included in this review as they are very rarely used in inpatients. Our systematic search returned no eligible studies focusing on the LSI-R or the V-RISK-10. Further information on each of the 9 included instruments can be found in Table 1.

2.3. Systematic search

A systematic search was conducted to identify studies that measured the predictive validity of the nine instruments in forensic psychiatric settings for the outcome of inpatient violence. We searched five databases (CINAHL, Embase, Global Health, PsycINFO and PubMed) from the earliest available start date up to January 2017, using a keyword search of titles and abstracts with the following search terms: *(PCL-R OR Psychopathy Checklist Revised OR HCR-20 OR Historical Clinical Risk Management OR PCL:SV OR Psychopathy Checklist Screening OR VRAG OR Violence Risk Appraisal Guide OR COVR OR Classification of Violence Risk OR LSI-R OR Level Service Inventory OR VRS OR Violence Risk Scale OR START OR Short Term Assessment Risk Treatability OR BVC OR Br?set Violence Checklist OR DASA OR Dynamic Appraisal of Situational Aggression OR V-RISK-10 OR Violence Risk Screening 10 OR risk assess*) AND inpatient* AND violen* AND risk AND (predict* OR valid*).*

Additional studies were identified through hand-searching references of the identified studies, using the Google Scholar "cited by" function, scanning the annotated bibliographies for each instrument, and corresponding with researchers in the field. Studies in all languages and those that were unpublished were considered for inclusion. Studies were excluded if: (1) they measured the predictive validity of selected scales of a tool, as

Table 1
Characteristics of the nine included violence risk assessment instruments.

Instrument type and name	No. of items	Static or Dynamic Items	Cut-off scores[a]	
Actuarial				
BVC	6	All dynamic	*High*	≥3
			Low	<3
COVR	-[b]	Mainly static	*High*	≥26
			Moderate	8–26
			Low	<8
DASA	7	All dynamic	*High*	≥4
			Low	<4
PCL-R	20	Mainly static	*High*	≥25
			Moderate	15–24
			Low	<15
PCL:SV	12	Mainly static	*High*	≥15
			Low	<15
VRAG	12	All static	*High*	≥14
			Moderate	−7–13
			Low	<−8
VRS	26	Both	*High*	≥42
			Low	<42
Structured professional judgement				
HCR-20	20	Both	*High*	≥30
			Moderate	20-29
			Low	<20
START	44	Both	-[c]	

[a] Information on cut-off scores relates only to those samples who reported a cut-off score; in some cases cut-off scores were unknown or a clinical risk judgement may have been used instead.
[b] COVR has a varying number of items depending on answers given to previous items.
[c] No cut-off score was used for START classifications, as the low, moderate and high risk categorisation was given from the violence risk estimate section.

https://doi.org/10.1016/j.eurpsy.2018.02.007 Published online by Cambridge University Press

the aim was to test the accuracy of the tool as a whole; (2) they focused on a specific subgroup of the forensic population (e.g., those with a diagnosis of learning disability), as our aim was to focus on the most common forensic psychiatric populations; (3) instruments were coded retrospectively without blinding to outcomes, to avoid any possible observer biases in evaluating outcomes; (4) they were calibration studies for the actuarial tools, as such development samples will provide inflated accuracy. Where studies used overlapping samples, the sample with the larger number of participants was used in order to avoid double-counting. Using this search strategy, we identified 52 studies eligible for inclusion.

To be included in the full meta-analysis, studies were required to report numbers of true positives, false positives, true negatives, and false negatives at a given tool-specific cut-off score for the outcome of inpatient violence over a defined time period. We contacted study authors if this information was unavailable in the manuscript and they were asked to fill in a standardised form. The desired full range of outcome data were available in the manuscripts of 11 eligible studies (13 samples). Further data was requested from the authors of the other 41 manuscripts and data was obtained for an additional 8 studies (22 samples). Of the 52 eligible studies, 43 (78 samples) gave an overall performance measure (the area under the curve value; AUC) and thus were included for calculating the median summary AUC value for a wider sample. The final number of studies included in the meta-analysis of other performance measures (i.e. true and false positives/negatives with AUCs) was 19 (amounting to 35 samples).

2.4. Quality assessment

The QUADAS-2 tool, designed to assess methodological quality for systematic reviews of studies investigating diagnostic or prognostic accuracy, provided a risk of bias for each study, with low or high risk of bias categorisations. All included studies showed a low risk of bias.

2.5. Data analysis

Risk assessment instruments were divided into two groups: those designed for the prediction of imminent violence over a 24-hour period following the assessment (BVC and DASA) and those designed for the prediction of violence over a longer period (COVR, HCR-20, PCL-R, PCL:SV, START, VRAG and VRS). Given that instruments used for violence risk assessment in a clinical setting are primarily used to identify higher risk individuals that may need monitoring, we combined subjects who were classified as moderate risk with those classified as high risk, and compared these two categories to low risk patients.

2.5.1. Meta-analytic model

We followed guidelines in the Cochrane collaboration for systematic reviews of diagnostic and prognostic test accuracy [32]. We examined two central measures of accuracy: sensitivity (the proportion of violent patients that a risk assessment tool predicted to be higher risk) and specificity (the proportion of non-violent patients that an instrument predicted to be low risk). We then developed a bivariate random-effects model that jointly analyzed pairs of sensitivities and specificities, taking into account their correlation with one another [33]. Without covariates, this model is a different parameterisation of the hierarchical summary receiver operating characteristic (HSROC) model [34]. We then used summary receiver operating characteristic (SROC) plots to present the results of each study in receiver operating characteristic (ROC) space, with each study plotted as a single sensitivity-specificity point. This produced a SROC curve, with a summary operating point (showing summary sensitivity and specificity values), a summary AUC value, 95% confidence region and 95% prediction region. We obtained summary accuracy estimates for the sensitivity, specificity, positive predictive value (PPV; the proportion of patients classified as higher risk who went on to be violent), negative predictive value (NPV; the proportion of patients classified as low risk who went on to not be violent), diagnostic odds ratio (DOR; the ratio of the odds of violent patients having been classified as higher risk relative to the odds of non-violent patients having been classified as low risk) and the area under the curve (AUC) value.

2.5.2. Heterogeneity

Heterogeneity is expected in meta-analyses of diagnostic or prognostic test accuracy due to the bivariate nature of the analysis and variation in cut-off scores; therefore, the standard Q and I^2 statistics are not recommended [35–39], but with no consensus on what to use [40]. Thus it is recommended that visual evaluation of the scatter of points from the SROC curve and the size of the ellipse of the prediction regions be used to assess heterogeneity. A greater scatter of points from the SROC curve and a larger prediction region are indicative of greater levels of heterogeneity [32].

2.5.3. Meta-regression and subgroup analyses

Meta-regression analyses were conducted to investigate the relationship between an overall accuracy estimate (the AUC value) and pre-specified study and sample characteristics, to test whether any had a moderating effect on the AUC. Sample-related variables included sample size, gender, mean age of participants, and proportion of patients with psychotic disorder, personality disorder, or violent index offence. Study-related variables included temporal design of the study (prospective vs. retrospective), type of instrument (actuarial vs. structured professional judgement), follow-up period post-assessment, and definition of violent outcome used (interpersonal violence vs. interpersonal violence and verbal aggression). Meta-regression analysis was performed for studies included in the meta-analysis. We planned to investigate any significant findings on meta-regression using subgroup analyses. We also performed an additional analysis of the alternative binning strategy (low/medium vs. high) for the longer-term tools.

All analysis was conducted on Stata [41], using the *midas* command to generate summary statistics and a SROC curve and the *metareg* command for meta-regression analyses. Summary PPVs and NPVs were not produced by the *midas* command and were therefore calculated as medians. Summary AUC values for the wider group of eligible samples were also calculated as medians.

3. RESULTS

3.1. Descriptive characteristics

For the wider sample of studies that reported on AUC values, information was collected for 6,840 participants in 78 samples from 43 independent publications. There were 5,680 (83%) male patients and 1,150 female patients. In the meta-analysis of all performance measures (with additional information on sensitivity and specificity), information was collected for 3,306 participants in 35 samples from 19 independent publications (Table 2). Standardised outcome information on numbers of true and false positives and negatives for 24 samples was obtained directly from study authors. When investigating all performance measures, there were 2,645 (80%) male patients and 661 female patients and the overall mean age of patients was 36.6 years (standard deviation [SD] = 3.5). There was some variation in both sample size (mean = 94.5; SD = 120.4) and rate of violence over the study

https://doi.org/10.1016/j.eurpsy.2018.02.007 Published online by Cambridge University Press

Table 2
Descriptive and demographic characteristics of samples for imminent and longer-term instruments included in the full meta-analysis (k = 35).

Category and group	Imminent (k = 6)	Longer-Term (k = 29)
Tool Information		
Type of tool		
Actuarial	6 (100)	13 (45)
Structured professional judgement	0 (0)	16 (55)
Tool used		
BVC	3 (50)	-
COVR	-	3 (10)
DASA	3 (50)	-
HCR-20	-	13 (45)
PCL-R	-	4 (14)
PCL:SV	-	1 (3)
START	-	3 (10)
VRAG	-	4 (14)
VRS	-	1 (3)
Sample characteristics		
Male participants (n (%))	1115 (80)	1549 (81)
Age (years; mean (SD))	37.0 (2.5)	36.4 (3.8)
Psychotic disorder (n (%))	508 (37)	931 (81)
Personality disorder (n (%))	122 (9)	449 (36)
Violent index offence (n (%))	715 (51)	1089 (73)
Study design		
Sample size (mean (SD))	232 (233)	66 (54)
Temporal design		
Retrospective	0 (0)	12 (41)
Prospective	6 (100)	15 (52)
Pseudo-prospective	0 (0)	2 (7)
Length of follow-up (days; mean (SD))	1.0 (0.0))	692.2 (978.6)
Outcome		
Violent outcome measured		
Only interpersonal physical violence	2 (33)	19 (66)
Including verbal aggression	4 (67)	10 (34)
Rate of violence during study (mean (SD))	23.8 (15.3)	32.6 (16.2)

Note: Data are number (%) of samples, unless stated otherwise. Percentages are reported in relation to only those samples where information was available for the variable in question. SD = standard deviation.

period (mean = 31% of the sample being violent; SD = 16.1). Each risk assessment instrument had between one and four studies assessing predictive validity, with the exception of the HCR-20, which was investigated in 13 studies. Studies were conducted in 12 different countries: Australia, Belgium, Canada, Denmark, Hong Kong, Ireland, Japan, the Netherlands, Norway, Spain, the UK and the USA.

3.1.1. Comparison between groups

In the meta-analysis of all performance measures, there were 1,394 patients in the 6 imminent tool samples (reported in 4 publications), compared to 1,912 patients in the 29 longer-term tool samples (15 publications). Both sample groups had approximately 80% male patients (Table 2) and there was little difference in mean age (37.0 and 36.4 years, respectively). Sample sizes for imminent tool studies ranged between 38 and 530 patients, while for longer-term tool studies, they spanned from 29 to 185. Follow-up length for all imminent tool samples had a 24-hr follow-up, while for longer-term tool samples, it was a mean of 692 days (SD = 979). The mean rate of violence over the defined follow-up period was 23.8% in the imminent tool sample compared with 32.6% for longer-term tools.

3.2. Predictive accuracy

3.2.1. Summary statistics

The studies included for the production of these summary statistics were those for which information on true and false positives and negatives was available (k = 35).

Predictive accuracy was different for the two groups of instruments (Table 3). In studies of imminent instruments, sensitivity was 0.59 (95% confidence interval [95% CI]: 0.29–0.83), while for longer-term instruments, it was 0.75 (95% CI: 0.65–0.83). The summary specificity for imminent tools was 0.99 (95% CI: 0.80–1.00) and for longer-term tools was 0.56 (95% CI: 0.46–0.66). A summary DOR for imminent tools could not be accurately calculated due to the number of zero-value categories (2 of the 6 samples included had one or more cells with zero values). The summary diagnostic odds ratio (DOR) for longer-term tools was 4.0 (95% CI: 3.0-6.0). The median PPV for imminent instruments was 0.36 (Interquartile range [IQR]: 0.10–0.93) and the median NPV was 0.99 (IQR: 0.85-1.00). The median PPV for longer-term instruments was 0.55 (IQR: 0.30-0.75) and the median NPV was 0.75 (IQR: 0.58-0.95).

Two different summary estimates of AUC values are reported based on different sample sizes. The first were calculated as median AUCs from all eligible studies that reported AUC values; this amounted to 78 samples and a total of 6,840 patients from 43 publications, based on 10 imminent tool samples (1,666 patients) and 68 longer-term tool samples (5,174 patients). The median AUC for imminent instruments was 0.83 (IQR: 0.71-0.85), while for longer-term instruments it was 0.68 (IQR: 0.62-0.75) (Table 3).

The second summary AUC value reported is that from the samples included in the meta-analysis (k = 35), as for the other reported performance measures. The summary AUC value for imminent tools in the meta-analysis sample was 0.90 (95% CI: 0.87-0.92) and for longer-term tools it was 0.71 (95% CI: 0.67-0.75).

3.2.2. HSROC curves

Figs. 1 and 2 show the hierarchical summary receiver operating characteristic (HSROC) curve formed from the meta-analysis of imminent and longer-term instruments, respectively. On both curves, the summary sensitivity, specificity point is plotted, along with a 95% confidence contour and a 95% prediction contour. The

https://doi.org/10.1016/j.eurpsy.2018.02.007 Published online by Cambridge University Press

Table 3
Summary accuracy estimates produced by two categories of violence risk assessment instruments.

	Imminent Instruments (k = 6)	Longer-Term Instruments (k = 29)
Summary estimates (95% confidence interval)		
Sensitivity	0.59 (0.29 – 0.83)	0.75 (0.65 – 0.83)
Specificity	0.99 (0.80 – 1.00)	0.56 (0.46 – 0.66)
PPV[a]	0.36 (0.10 – 0.93)	0.55 (0.30 – 0.75)
NPV[a]	0.99 (0.85 – 1.00)	0.75 (0.58 – 0.95)
DOR	-	4.00 (3.00 – 6.00)
AUC[a]	0.83 (0.71 – 0.85)	0.68 (0.62 – 0.75)

Note: Median AUC values calculated from wider samples (k = 78): 10 samples for imminent tools and 68 samples for longer-term tools.
[a] Median (interquartile range).

HSROC curve for imminent tools is approaching the top left-hand corner of the graph, indicating high accuracy, but the prediction contour is large, indicating high levels of between-study heterogeneity (Fig. 1). For longer-term tools, the HSROC curve is closer to the y = x diagonal that would indicate an uninformative test than it is to the top left-hand corner of space (Fig. 2). The prediction contour is also large, again indicating high levels of between-study heterogeneity.

3.2.3. Individual tool performance

Within the wider group of 78 samples, the majority of samples assessed the performance of the HCR-20 (k = 27) and the PCL-R (k = 10). These tools performed moderately for the prediction of inpatient violence with median AUCs of 0.70 (IQR: 0.62-0.80) and 0.64 (IQR: 0.61-0.69), respectively. Imminent instruments had higher AUC values; the BVC (k = 5) had a median AUC of 0.83 (IQR: 0.75–0.87) and the DASA (k = 5) also had a median AUC of 0.83 (IQR: 0.65-0.90). See Appendix Table 2 in Supplementary material for all accuracy measures for each instrument.

3.3. Investigation of heterogeneity and subgroup analyses

Meta-regression analyses were only performed for longer-term instrument samples, as there were too few imminent instrument samples (k = 6). No study- or sample-related variables were

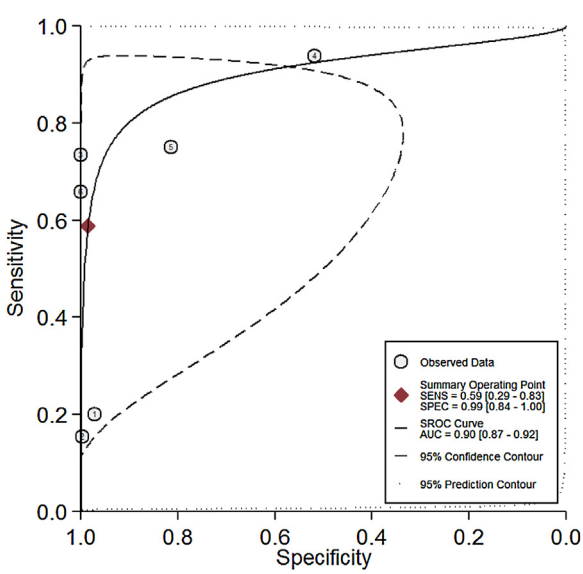

Fig. 1. Summary receiver operating characteristics (SROC) curve from bivariate analysis of imminent violence risk assessment instruments for forensic inpatient violence.
Note: Summary operating point = best fit for sensitivity and specificity. 95% confidence contour represents within-study heterogeneity. 95% prediction contour represents between-study heterogeneity.

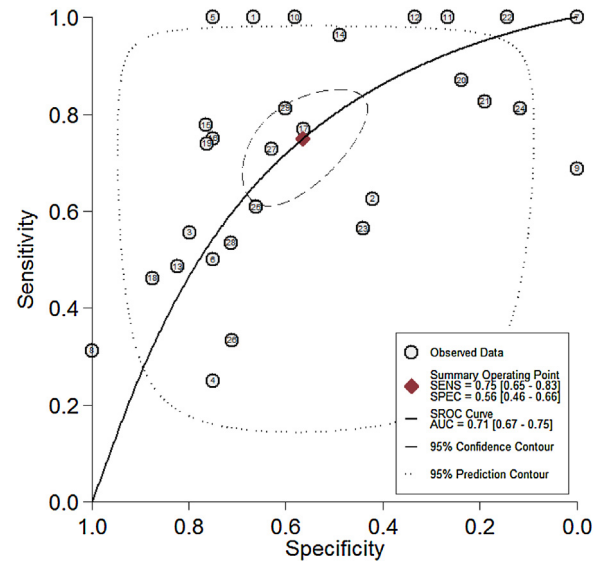

Fig. 2. Summary receiver operating characteristics (SROC) curve from bivariate analysis of longer-term violence risk assessment instruments for forensic inpatient violence.
Note: Summary operating point = best fit for sensitivity and specificity. 95% confidence contour represents within-study heterogeneity. 95% prediction contour represents between-study heterogeneity.

associated with between-study difference in AUCs (Appendix Table 3 in Supplementary material). When we used an alternative binning strategy (low/medium vs. high), the performance of the longer-term tools was marginally improved with regards to PPV and AUC (Appendix Table 4 in Supplementary material).

4. DISCUSSION

This systematic review and meta-analysis examined the predictive accuracy of 9 violence risk assessment instruments for inpatient violence in forensic psychiatric hospitals from 78 samples involving 7,705 patients from 14 different countries. The main finding was that instruments designed for the prediction of imminent violence performed better at predicting inpatient violence than instruments designed for longer-term follow-up periods, based on a range of performance measures. As a measure of overall accuracy, the median AUC for imminent tool studies was 0.83, compared to a median AUC of 0.68 for longer-term tools. Generally, AUC values greater than 0.8 indicate a highly accurate test and those below 0.7 indicate poor to moderate accuracy [42]. Imminent instruments performed particularly well for screening out low risk individuals: 99% of those who went on to not be violent were correctly predicted to be low risk (specificity) and 99% of those who were predicted to be low risk went on to not be violent (NPV).

4.1. Individual tool performance

The HCR-20 is the most widely-used violence risk assessment instrument internationally, yet our findings from this review show that it has at best moderate accuracy across a range of performance measures, with regard to the prediction of inpatient violence. These lower levels of accuracy are likely a consequence of how the HCR-20 has been developed, as it is a general violence risk assessment instrument with applications and recommendations for use in a broad range of contexts, populations and follow-up periods. Similarly, the PCL-R and VRAG performed poorly for the

https://doi.org/10.1016/j.eurpsy.2018.02.007 Published online by Cambridge University Press

prediction of inpatient violence. Although their performance may be acceptable for some populations in the community, the current evidence does not support their use for the prediction of inpatient violence in forensic psychiatry.

The two instruments designed specifically for imminent inpatient violence prediction (the BVC and the DASA) performed with higher accuracy for a number of measures. However, there were few studies (k = 10) despite being recommended by NICE. There were more studies focused on the poorer performing tools, such as the HCR-20, suggesting a need to move towards research examining short-term tools, and possibly optimizing them by considering novel risk factors [43].

4.2. Clinical implications

Our findings indicate that the use of instruments designed for the imminent prediction of violence over the 24-hour period post-assessment yielded higher accuracy for multiple measures of performance. In clinical practice, consideration should be given to the use of the BVC and the DASA, both of which are recommended tools in one clinical guideline for short-term management of violence and aggression in inpatient mental health settings [3]. Furthermore, the narrow 24-hour window within which violence is predicted allows for prevention and management strategies to be implemented when they may be most needed. Both the BVC and DASA are brief checklists (6 and 7 items, respectively), have the advantage of scalability and can easily be integrated into routine practice.

However, other clinical contexts will exist where longer-term instruments may be more relevant or appropriate; the high sensitivity (0.75) and moderate PPV (0.55) suggest these instruments may have a role for some patients. Considering the brevity of the BVC and DASA, they could act as a screen before a longer term tool is used considering the expense involved in administering time-consuming and resource-intensive instruments [44].

However, for both imminent and longer-term tools, it is important for there to be a link with clinical interventions and outcomes to link the risk prediction element with subsequent management of risk. One randomised controlled trial (RCT) has been conducted finding a positive effect (reduction in inpatient violent incidents) when the BVC was used in a forensic psychiatric sample combined with implementation of a violence management strategy and training [47].

4.3. Strengths and limitations

To our knowledge, this is the first comprehensive review and meta-analysis of violence risk assessment instruments in the context of their predictive accuracy for inpatient violence in forensic psychiatric populations. There has been one previous review of risk assessment for inpatient violence in forensic psychiatric patients [45]. However, it used mean correlation coefficients between violence risk assessment scores and inpatient violence, which is limited to examine predictive accuracy. Further, only three violence risk assessment instruments (the HCR-20, PCL-R and PCL:SV) were included in that review.

Recent criticism of risk assessment literature has stated that there is an insufficient focus on subpopulations in a specific context [46]. Unlike previous reviews of risk assessment tools, the current one investigates a particular patient group in one setting. In addition, the literature on predictive accuracy of violence risk assessment has been limited by relying on one or two measures of accuracy [46]. The AUC value, for example, is often reported in isolation; however, it does not indicate whether this discrimination is clinically useful, nor does it provide any information on the calibration of the instrument's predictions with actual future violence [48]. To address this, we investigated a range of accuracy measures although none of the included studies reported calibration measures.

One limitation is that only studies reporting true and false positives and negatives could be included in the full meta-analysis. However, median AUCs were reported for the wider sample of eligible studies. Further, we corresponded with authors requesting unpublished data and increased the number of possible samples from 11 to 35 samples that report a range of performance measures. Another limitation is the large amount of between-study heterogeneity, perhaps due to variations in cut-off scores used for risk classifications. A number of other possible explanations were investigated in meta-regression and no associations were found to explain the variation between tools. This heterogeneity is expected, especially in prognostic (as opposed to diagnostic) studies, and the use of a random-effects model accounted for this variation. Further, where possible, the same cut-off scores were applied for each sample of the same instrument.

There were differences between the imminent and longer-term groups of studies with regard to the type of primary outcome used (interpersonal violence only vs. interpersonal violence and verbal aggression), which could explain their relative performance. Although this was investigated in meta-regression analyses and found to have no effect on the AUC accuracy estimate for longer-term tools, this analysis could not be performed for imminent instruments due to lack of available data. It is possible, therefore, that the better performance of the imminent tools (based on AUCs) is based on higher rates of softer outcomes (i.e. aggression), which will inflate base rates.

We also found marginally improved performance in some performance measures when we used a different binning strategy (low/medium vs high). Whether this merits a change in how these tools are used in practice and for which inpatient settings requires further work.

4.4. Future directions

Future research on violence risk assessment in forensic inpatient settings should focus more on imminent instruments as this meta-analysis found a smaller proportion of the research literature based on these instruments. Another useful direction for research would be further exploration of whether there should be a screen before longer-term instruments are used [44]. As the two imminent tools in this study rely predominantly on dynamic variables, research could investigate the role of novel dynamic variables to improve risk prediction, and whether adding static variables can add incremental performance. Further to this, new technologies that have been developed for the use of risk prediction and monitoring should be examined [49]. From a methodological perspective, future work in this area should report multiple estimates of predictive accuracy in order to provide a more complete picture of an instrument's performance, including measures of calibration. Overall, this meta-analysis supports previous recommendations that future work in violence risk assessment requires the development and validation of tools designed for specific populations [46,50,51].

Acknowledgements

We thank the following study authors for providing tabular data for the analyses: Dr. Kaoru Arai, Dr. Oliver Chan, Professor Geoff Dickens, Dr. Óscar Herrero, Dr. Helen Miles, Professor Robert Snowden, Professor Lindsay Thomson and Dr. Vivienne de Vogel.

https://doi.org/10.1016/j.eurpsy.2018.02.007 Published online by Cambridge University Press

SF is a Wellcome Trust Senior Research Fellow in Clinical Science (202836/Z/16/Z). We would like to disclose no conflicts of interest of funding sources for this review.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.eurpsy.2018.02.007.

References

[1] Wildgoose J., Briscoe M, Lloyd K. Psychological and emotional problems in staff following assaults by patients. The Psychiatrist 2003;27(8):295–7.
[2] Bowers L, Stewart D, Papadopoulos C, et al. Inpatient violence and aggression: A literature review. Report from the Conflict and Containment Reduction Research Programme. Kings College London: Institute of Psychiatry; 2011.
[3] National Institute for Health and Care Excellence, Violence and aggression: short-term management in mental health, health and community settings. NICE Guideline (NG10) ed, In: 2015.
[4] Almvik R, Woods P, Rasmussen K. The Brøset Violence Checklist sensitivity, specificity, and interrater reliability. Journal of interpersonal violence 2000;15 (12):1284–96.
[5] Linaker OM, Busch-Iversen H. Predictors of imminent violence in psychiatric inpatients. Acta Psychiatrica Scandinavica 1995;92(4):250–4.
[6] Ogloff J, Daffern M. Dynamic appraisal of situational aggression: Inpatient version. Melbourne, Victoria, Australia: Monash University and Forensicare; 2002.
[7] Galletly C, Castle D, Dark F, et al. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the treatment of schizophrenia and related disorders. Australian & New Zealand Journal of Psychiatry 2016;50 (5):410–72.
[8] Lehman AF, Lieberman JA, Dixon LB, et al. Practice guideline for the treatment of partients with schizophrenia. American Journal of Psychiatry 2004161(2 Suppl).
[9] Campbell MA, French S, Gendreau P. The prediction of violence in adult offenders a meta-analytic comparison of instruments and methods of assessment. Criminal Justice and Behavior 2009;36(6):567–90.
[10] Fazel S, Singh JP, Doll H, et al. Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: systematic review and meta-analysis. BMJ 2012;345:e4692.
[11] Singh JP, Grann M, Fazel S. A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. Clinical psychology review 2011;31(3):499–513.
[12] Whittington R, Hockenhull J, McGuire J, et al. A systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour: update 2002–8. Health Technology Assessment 2013;17(50):1–128.
[13] Singh JP, Fazel S. Forensic risk assessment: A metareview. Criminal Justice and Behavior 2010;37(9):965–88.
[14] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS med 2009;6(7): e1000097.
[15] Hurducas CC, Singh JP, de Ruiter C, et al. Violence risk assessment tools: A systematic review of surveys. International Journal of Forensic Mental Health 2014;13(3):181–92.
[16] Singh JP, Desmarais SL, Hurducas C, et al. International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. International Journal of Forensic Mental Health 2014;13(3):193–206.
[17] Singh JP, Desmarais SL, Otto RK, et al. The International Risk Survey: Use and perceived utility of structured violence risk assessment tools in 44 countries. In: International Perspectives on Violence Risk Assessment. Oxford University Press; 2016.
[18] Monahan J, Steadman HJ, Silver E, et al. Rethinking risk assessment: The MacArthur study of mental disorder and violence. Oxford University Press; 2001.
[19] Monahan J, Steadman HJ, Robbins PC, et al. An actuarial model of violence risk assessment for persons with mental disorders. Psychiatric services 2005;56 (7):810–5.
[20] Andrews D, Bonta J. LSI-R: The Level of Service Inventory-Revised. Toronto, Ontario, Canada: Multi-Health Systems: Inc; 1995.
[21] Hare RD. The Hare psychopathy checklist-revised: Manual. Multi-Health Systems, Incorporated; 1991.
[22] Hart SD, Cox DN, Hare RD. The Hare psychopathy checklist: Screening version (PCL: SV). MHS-Multi-Health Systems, Incorporated; 1995.
[23] Quinsey VL, Harris GT, Rice ME, et al. Actuarial prediction of violence. In: Quinsey V, Harris G, Rice M, Cormier C, editors. The Law and Public Policy. Violent offenders: Appraising and Managing Risk. Washington, DC, US: American Psychological Association; 2006. p. 155–96.
[24] Quinsey V, Harris G, Rice M, et al. Violent offenders: appraising and managing risk. Washington, DC: American Psychological Association; 2006.
[25] Wong S, Gordon A. Violence Risk Scale (VRS). Saskatoon, Saskatchewan. 2000.
[26] Douglas K, Hart S, Webster C, et al. HCR-20 version 3: assessing risk for violence. Burnaby, BC, Canada: Mental Health, Law and Policy Institute, Simon Fraser University; 2013.
[27] Webster C, Douglas K, Eaves D, et al. HCR-20: Assessing Risk for Violence (Version 2). Burnaby, British Columbia, Canada: Simon Fraser University. Mental Health, Law, and Policy Institute; 1997.
[28] Webster CD, Martin M, Brink J, et al. START: The Short-term assessment of risk and treatability. Hamilton: St Joseph's Healthcare; 2004.
[29] Webster C, Martin M, Brink J, et al. Manual for the Short-Term Assessment of Risk and Treatability (START) (Version 1.1). Coquitlam, Canada: British Columbia Mental Health & Addiction Services; 2009.
[30] Bjorkly S, Hartvig P, Heggen FA, et al. Development of a brief screen for violence risk (V-RISK-10) in acute and general psychiatry: An introduction with emphasis on findings from a naturalistic test of interrater reliability. European Psychiatry 2009;24(6):388–94.
[31] Hartvig P, Østberg B, Alfarnes S, et al. Violence Risk Screening-10 (V-RISK-10). Oslo, Norway: Centre for Research and Education in Forensic Psychiatry; 2007.
[32] Macaskill P, Gatsonis C, Deeks J, et al. Cochrane handbook for systematic reviews of diagnostic test accuracy. Version 09 0. London: The Cochrane Collaboration; 2010.
[33] Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of clinical epidemiology 2005;58(10):982–90.
[34] Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Statistics in medicine 2001;20(19):2865–84.
[35] Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. Statistics in medicine 2010;29(12):1282–97.
[36] Jackson D, White IR, Riley RD. Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. Statistics in medicine 2012;31 (29):3805–20.
[37] Verde PE. Meta-analysis of diagnostic test data: modern statistical approaches. Deutsche Nationalbibliothek; 2008.
[38] White IR. Multivariate random-effects meta-regression: updates to mvmeta. Stata Journal 2011;11(2):255.
[39] Zhou Y, Dendukuri N. Statistics for quantifying heterogeneity in univariate and bivariate meta-analyses of binary data: The case of meta-analyses of diagnostic accuracy. Statistics in medicine 2014;33(16):2701–17.
[40] Naaktgeboren CA, Ochodo EA, Van Enst WA, et al. Assessing variability in results in systematic reviews of diagnostic studies. BMC Medical Research Methodology 2016;16:6.
[41] StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.; 2015.
[42] Tape TG., The area under an ROC curve. Interpreting diagnostic tests, 2006. Available from: http://gim. unmc. edu/dxtests/roc3.htm.
[43] Eriksen BMS, Bjørkly S, Lockertsen Ø, et al. Low cholesterol level as a risk marker of inpatient and post-discharge violence in acute psychiatry—A prospective study with a focus on gender differences. Psychiatry research 2017;255:1–7.
[44] Rosenfeld B, Foellmi M, Khadivi A, et al. Determining when to conduct a violence risk assessment: Development and initial validation of the Fordham Risk Screening Tool (FRST). Law and human behavior 2017;41(4):325.
[45] Hogan N, Ennis L, Assessment F. Assessing risk for forensic psychiatric inpatient violence: A meta-analysis. Open Access Journal of Forensic Psychology 2010;2:137–47.
[46] Douglas T, Pugh J, Singh I, et al. Risk assessment tools in criminal justice and forensic psychiatry: the need for better data. European Psychiatry 2017;42:134–7.
[47] Abderhalden C, Needham I, Dassen T, et al. Structured risk assessment and violence in acute psychiatric wards: randomised controlled trial. The British Journal of Psychiatry 2008;193(1):44–50.
[48] Singh JP. Predictive validity performance indicators in violence risk assessment: A methodological primer. Behavioral Sciences & the Law 2013;31(1):8–22.
[49] Gulati G, Cornish R, Al-Taiar H, et al. Web-based violence risk monitoring tool in psychoses: pilot study in community forensic patients. Journal of forensic psychology practice 2016;16(1):49–59.
[50] Fazel S, Wolf A, Larsson H, et al. Identification of low risk of violent crime in severe mental illness with a clinical prediction tool (Oxford Mental Illness and Violence tool [OxMIV]): a derivation and validation study. Lancet Psychiatry 2017;4(6):461–8.
[51] Wolf A, Fanshawe T, Sariaslan A, et al. Prediction of violent crime on discharge from secure psychiatric hospitals: A clinical prediction rule (FoVOx). European Psychiatry 2017;47:88–93.

https://doi.org/10.1016/j.eurpsy.2018.02.007 Published online by Cambridge University Press