

The Short- to Medium-Term Predictive Validity of Static and Dynamic Risk-of-Violence Measures in Medium- to Low-Secure Forensic and Civil Inpatients

Brayden Finch^a, Derek G. Gilligan^a, Sean A. Halpin^a and Megan E. Valentine^b

^a*School of Psychology, Faculty of Science and Information Technology, University of Newcastle, Newcastle, NSW, Australia;* ^b*School of Mathematical and Physical Sciences, Faculty of Science and Information Technology, University of Newcastle, Newcastle, NSW, Australia*

The prediction and subsequent management of aggression by psychiatric inpatients is a crucial role of the mental health professional. This retrospective cohort study examines the predictive validity of 10 static and dynamic risk-of-violence measures and subscales in 37 forensic and 37 civil inpatients residing in a medium- to-low security psychiatric facility for a period of up to 6 months. Retrospective file records were sourced to conduct an AUC analysis of the ROC curve for short- and medium-term follow-up periods. The hypothesis that dynamic measures would be better predictors than static measures over the short term was supported. Albeit to a lesser extent, dynamic measures were still better predictors than static measures over the medium term. This result was seen in both civil and forensic groups. Three previously untested measures were found to predict aggression within the sample. It is recommended that mental health services employ the use of dynamic measures when making short-term risk-of-violence predictions for civil and/or forensic inpatients.

Key words: civil; dynamic; forensic; risk; short term; static; violence.

The prediction and management of aggression and violence displayed by individuals with mental health disorders is a prominent issue in modern psychiatric settings (Belfrage & Douglas, 2002). Individuals who display violence, the most extreme form of aggression (Anderson & Bushman, 2002), are viewed by society as being the most dangerous to others. Consequently, predicting and managing violence is a priority of treating organisations, professionals, and researchers (Monahan et al., 2001). There exists a public and legal expectation that mental health professionals can fully predict violence displayed by individuals with mental health

disorders (Monahan et al., 2001). Despite these expectations, the ability of mental health professionals to predict violence using clinical judgement alone is modest at best (Monahan et al., 2001). Consequently, many actuarial risk-assessment measures have been developed to predict more accurately when a patient may become violent, with more than 120 currently available across general and psychiatric settings (Singh, Grann, & Fazel, 2011). Validation of these measures has become a priority to assist mental health professionals and organisations in making informed decisions about the best available methods of violence risk assessment in a

Correspondence: Brayden Finch, Lake Macquarie Mental Health Service, 72 Watt St, Newcastle, NSW 2300, Australia. Email: Brayden.finch@uon.edu.au Phone: +612 4904 9000

particular setting. Differences in the types of risk factors used in such measures may help to determine which type of measure most accurately predicts violence in a given setting and over a given time frame.

Forensic and Civil Inpatients

There is uncertainty regarding differences in risk assessment applied to forensic versus civil populations. Civil inpatients typically have much shorter observation periods than do forensic inpatients, which influences the type of research that can be performed on such samples (Steinert, 2002). Preliminary research shows that civil inpatients typically present with more acute psychopathological states (Steinert, 2002), which suggests they may experience more rapid changes in risk state than do forensic inpatients (Belfrage & Douglas, 2002). Further research is required to determine what types of measures are best at predicting aggression in each of these populations.

Static vs. Dynamic Risk Factors

Static risk factors have been the primary focus of research on risk measures in the mental health context (Douglas & Skeem, 2005). Static risk factors are those that do not fluctuate over time, such as gender or a history of violence (Andrews & Bonta, 1998). Conversely, dynamic risk factors are those that do fluctuate with time and circumstance, such as anti-social attitudes or active psychosis (Andrews & Bonta, 1998). Importantly, static risk factors are seen as unmalleable, whereas dynamic risk factors are able to be reduced following targeted intervention, resulting in an overall reduction of risk (Webster, Douglas, Belfrage, & Link, 2000). It is therefore worth while to further investigate the utility of dynamic risk measures, both for the prediction and for the management of violence.

Static risk factors may be better predictors of violence over the long term (Quinsey, Harris, Rice, & Cormier, 2006), and dynamic

risk factors better over the short term (Chu, Thomas, Daffern, & Ogloff, 2013; Douglas, Ogloff, Nicholls, & Grant, 1999; McNeil, Gregory, Lam, Binder, & Sullivan, 2003). Inconsistent definitions of ‘long-term’ and ‘short-term’ are common in the literature, with short-term ranging from anything to a few hours to a few months, and long-term ranging from a few months to years (Chu et al., 2013). While many studies have compared static and dynamic risk-assessment measures (see Chu et al., 2013), few have examined predictive validity at one month or less. This has implications for the generalizability of the findings, since the average length of psychiatric inpatient admissions in Australia is less than a month (Australian Institute of Health and Welfare, 2011; Chu et al., 2013). Given that the majority of the risk-assessment literature is of studies in a North American context, filling this gap in the research is important in determining which tools are best for use over what duration in an Australian context.

In 2013, Chu et al compared static and dynamic risk-of-violence assessment measures in a high-security Australian forensic hospital over the short term (defined as up to 1 month after time of rating) and compared this to the medium term (defined as between 1 and 6 months after time of rating). Dynamic risk measures demonstrated better predictive validity than static risk measures over the short term, but static measures did not demonstrate better predictive ability than dynamic measures over the medium term, as predicted. Chu et al. (2013) reasoned that the highly structured environment of the high-security setting and subsequent low levels of violence may have been among the factors that influenced this result.

Statistical Comparisons of Risk Measures

Risk-of-violence measures are often developed to be used in one of two ways. Purely actuarial assessments give risk factors numerical values

to produce a probabilistic estimate of the individual's likelihood of violence (Singh et al., 2011). Conversely, Structured Professional Judgement (SPJ) assessments use a combination of actuarial methods and structured assessment approaches to inform clinical decision-making (Douglas & Ogloff, 2003). While it has been argued that SPJ measures should be examined only in the way they were designed to be used (i.e., final risk judgements made by the clinician based on actuarial and other information), this makes it very difficult to separate differences that are due to broader methodology (actuarial vs SPJ) from differences due to particular tools (Heilbrun, Yasuhara, & Shah, 2010). Therefore, when comparing SPJ tools to actuarial ones, a standardised method of evaluating only the actuarial properties of the measures is used.

The most regularly employed statistical technique to compare risk measures involves analysing the area under the curve (AUC) of the receiver operating characteristic (ROC; Singh, Desmarais, & Van Dorn, 2013). ROC curves are created by plotting each risk-assessment measure's false positive rate against its true positive rate across score thresholds (Hajian-Tilaki, 2013). The AUC is then calculated, resulting in a predictive validity score between 0 and 1 for each measure, which can then be compared. In a risk-of-violence context, the AUC is thus an index of how well a risk-assessment tool discriminates between offenders and non-offenders across all possible cut-off scores. In general, measures producing AUCs of .5 to .6 are classified as a poor predictors, .6 to .7 are classified as modest predictors, .7 to .8 are classified as acceptable, .8 to .9 are classified as excellent, and measures producing AUCs greater than .9 are considered outstanding positive predictors (Hosmer & Lemeshow, 2000).

Other Measures

While there are many independently developed, widely tested risk-of-violence measures available to mental health professionals, this is

not the standard method of risk assessment in all practice settings. Many health departments and organisations have developed their own measures in an attempt to standardise the way in which data from mental health professionals is collected. In public health settings in New South Wales, Australia, the Mental Health Outcomes and Assessment Tools risk-assessment module (MHOAT-Risk) is routinely used to assess risk; however, the predictive validity of this measure has not been examined. The Health of the Nation Outcome Scale-Secure (HoNOS-secure) is used to assist in making clinical decisions based on level of risk in secure psychiatric facilities. Even though it was developed for risk management rather than risk assessment, limited research on its predecessor, the HoNOS-MDO, reveals it may have some predictive ability (van den Brink, 2010). As there is little information regarding how MHOAT-Risk was developed or how it is to be scored, and as the HoNOS-secure was developed for risk management rather than prediction, it is likely these measures are not as accurate at predicting aggression as more established risk-assessment measures.

Aims

The present study aimed to examine the predictive validity of several static and dynamic measures in a sample of forensic and civil inpatients residing in a low- to medium-secure psychiatric hospital at follow-ups of 1, 3, and 6 months. Specifically, we aimed to replicate Chu et al.'s (2013) study that found that dynamic measures were better predictors of aggression than were static measures over the *short term* (1-month follow-up). We also aimed to show that static measures are better predictors of aggression than dynamic measures over the *medium term* (3-month and 6-month follow-ups). We reasoned that the lower security setting used in the present study may allow for this finding, which Chu et al. (2013) failed to demonstrate. Additionally, we aimed to extend upon Chu et al. (2013) and examine the ability of these

measures to predict aggression for both forensic and civil groups. Finally, we aimed to examine the predictive ability of three additional measures: two that are routinely used within the public health service (MHOAT-Risk and HoNOS-secure) and a third that was developed for the present study from a review by Douglas and Skeem (2005), which we have named the Dynamic Risk Scale (DRS; see Appendix B).

In order to achieve our research aims, we tested the following hypotheses:

Hypothesis 1: Dynamic measures would be better at predicting aggression than static measures over the short term (1-month follow-up).

Hypothesis 2: Static measures would be better at predicting aggression than dynamic measures over the medium term (3-month and 6-month follow-ups).

Hypothesis 3: The patterns of predictive ability demonstrated by dynamic and static measures in each of the forensic and civil groups would be similar to those demonstrated in the total sample.

Hypothesis 4: The MHOAT-Risk and HoNOS-secure would have less predictive ability than more established dynamic risk measures.

Hypothesis 5: Due to its sound theoretical basis, the DRS would have similar predictive ability to more established dynamic risk measures.

Method

Study Design

This retrospective cohort study examined the ability of various risk-of-aggression assessment measures to predict an incident of physical or verbal aggression during follow-up periods of 1, 3, or 6 months from the time of assessment. Risk-assessment and follow-up aggression outcome data was used from a convenience sample of forensic and civil inpatients in a psychiatric hospital.

Power Calculation

A prospective power analysis was completed to determine the minimum number of participants that would be required to meaningfully interpret the ROC curves used in the statistical analysis. The power analysis revealed that a sample size of 70 participants would yield a 95% confidence interval around each point used to plot the ROC curve of approximately $\pm 12\%$. This amount of error was considered adequate.

Participants

The sample comprised 74 male inpatients residing in either a medium-secure unit or a low-secure unit at a psychiatric hospital in New South Wales, Australia. Inpatients in both units are classified as either forensic or civil. Forensic inpatients had committed a serious offence and had subsequently been found not guilty by reason of mental illness. Civil inpatients had displayed levels of risk in their previous inpatient setting that warranted referral and transfer to the current facility.

Inpatients who were residents for at least 6 months in either the medium-secure or low-secure units were eligible for inclusion in the study. Data from every forensic inpatient were included. Data from civil inpatients were collected until there were at least 70 participants in total, and until there was an equivalent number of inpatients in both groups.

Measures

Historical Clinical Risk Management 20 Factors—Version 2 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997)

The HCR-20 is a 20-item violence risk-assessment measure designed to be used with the SPJ method. Items are divided into 10 Historical (H) items comprising largely static factors, plus 5 Clinical (C) items comprising dynamic factors that reflect current mental and clinical status, and 5 Risk Management (R) items also comprising dynamic factors

that reflect future situational risk. Each item is coded as 0, 1, or 2, for a possible total score of 40, with a higher score indicating a higher level of risk. The HCR-20 is one of the most widely used tools to assess the risk of violence in forensic and psychiatric populations (Douglas, Hart, Webster, & Belfrage, 2013). The HCR-20 has previously demonstrated superior predictive ability to other measures combining dynamic and static factors (Chu et al., 2013).

Psychopathy Checklist–Revised (PCL-R; Hare, 2003)

The PCL-R is a risk-assessment measure which uses the actuarial method and comprises 20 mostly static items. Each item is coded as 0, 1, or 2, for a possible total score of 40, with a score of 30 and above being an indicator of psychopathy. The items are coded based on semi-structured interview and file reviews. The PCL-R is widely used as a predictor of violence, and has a large body of supporting evidence across different contexts (Leistico, Salekin, DeCoster, & Rogers, 2008; Walters, 2003). The PCL-R has previously demonstrated the best predictive validity of all the static measures over the short to medium term (Chu et al., 2013).

Short-Term Assessment of Risk and Treatability, Version 1.1 (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009)

The START is a risk-assessment measure shown to have promising predictive validity with forensic populations (Chu et al., 2013; Chu, Thomas, Ogloff, & Daffern, 2011; Nicholls, Brink, Desmarais, Webster, & Martin, 2006). It uses the SPJ method and is composed of 20 purely dynamic items. Each item is rated as either a *strength* or a *vulnerability*, and coded as a 0, 1, or 2. However, one study has found it difficult to code particular items as a strength or as a vulnerability, and such items have appeared to be highly collinear (e. g., Braithewaite, Charette, Crocker, & Reyes,

2010). Therefore, only the vulnerabilities scale of the START was used for a total possible score of 40, with higher scores indicating higher risk.

Dynamic Risk Scale (DRS; based on Douglas & Skeem, 2005)

The DRS is a risk-assessment measure developed for the present study based on Douglas and Skeem’s (2005) review, which listed and operationalised 9 promising dynamic risk factors (see Table 1). It was reasoned that these dynamic factors should be predictive when combined as items in an actuarial way. The items are explicated in Douglas and Skeem (2005), with clear operational definitions provided. Each item was coded as 0 (= ‘no/absent’), 1 (= ‘partially/possibly present’), or 2 (= ‘yes/definitely present’), for a possible total score of 18. Higher scores indicated greater risk.

Mental Health Outcomes and Assessment Tools–Risk-Assessment Module (MHOAT-Risk; NSW Health, 2001)

The MHOAT is a collection of standardised clinical measures implemented by New South Wales Health in 2001 to standardise statewide data collection by mental health professionals (NSW Health, 2001). The MHOAT risk-assessment module (MHOAT-Risk) is a collection of variables and is routinely used by staff across NSW Health, including the

Table 1. DRS Items.

Impulsiveness
Negative affectivity: anger
Negative affectivity: negative mood
Psychosis
Antisocial attitudes
Substance use and related problems
Interpersonal relationships
Treatment alliance and adherence: treatment and medication compliance
Treatment alliance and adherence: treatment-provider alliance

recruitment site. The measure requires the clinician to rate the patient as ‘yes’, ‘no’, or ‘unknown’ on *background* and *current* items within 4 categories of risk, including *general risk factors*, *suicide*, *violence/aggression*, and *other vulnerabilities*. Only the MHOAT-Risk categories *violence/aggression* (6 background factors, 8 current factors) and *general risk factors* (6 background factors, 4 current factors) were examined. While the background factors appear to be largely static and the current factors dynamic, the items are difficult to dichotomise in this manner as they are vague and not operationalised. There is no scoring manual to accompany the measure, and it is thus scored based on the item names only. Furthermore, the present authors were unable to locate any information about how the items were developed. An official NSW report on the implementation of the MHOAT states: ‘it is acknowledged that the utility of the standard measures will require review’ (Chippis, Raphael, & Coombs, 2002; p. 238). The present paper aims contribute information for possible future review of the MHOAT-Risk measure. In order to examine the predictive ability of the measure empirically, a decision was made to allocate a score of 1 to all ‘yes’ responses, and a score of 0 for all ‘no’ or ‘unknown’ responses, to produce a total possible score of 24, with a higher score indicating a higher level of risk.

Health of the Nation Outcome Scale–Secure, version 2b (HoNOS-secure; Dickens, Sugarman, & Walker, 2007)

The HoNOS is used within NSW Health as a standardised measure of the health and social functioning of people with severe mental illness (NSW Health, 2001). The HoNOS for users of secure and forensic services (HoNOS-secure) was developed as some items in the original HoNOS proved not applicable to secure settings (Royal College of Psychiatrists, 2015). The HoNOS and HoNOS-secure are not risk-assessment measures but, rather, allow patients to be rated in

terms of need for clinical risk-management protocols and care following risk assessment (Royal College of Psychiatrists, 2015; Dickens, Sugarman, & Walker, 2007). However, there is considerable evidence to suggest that the original HoNOS has good predictive validity in generalist mental health settings (Pirkis, Burgess, Kirk, Dodson, & Coombs, 2005; Shrinkfield & Ogloff, 2015; Webster, Bretherton, Goulter, & Fawcett, 2013). Furthermore, there is evidence to suggest the HoNOS-secure’s predecessor, the HoNOS-MDO, has some predictive ability for risk of violence (van den Brink, 2010). The HoNOS-secure was thus included to test whether it has similar predictive capacity.

The HoNOS-secure includes an amended version of the 12 original HoNOS items, as well as 7 additional *secure* items, which involve rating risk in the near future, defined as within weeks or months (Dickens, Sugarman, & Walker, 2007). Each item scored from 0 to 4, for a total possible score of 76. Higher scores indicate a greater need for secure care. While there is evidence of the HoNOS-secure’s reliability (Dickens et al., 2007), its predictive validity for aggression has not been examined over a short or a medium time frame.

Measures containing a clear majority of static items were classified as *static measures* in this study. Measures that contained at least an equal number of dynamic and static items were classified as *dynamic measures* (see Table 2).

Table 2. Static vs. dynamic measures.

Static	Dynamic
H-scale of HCR-20	HCR-20 Total C-scale of HCR-20 ^a R-scale of HCR-20 ^a C+R scales of HCR-20 ^a
PCL-R	START ^a MHOAT-Risk HoNOS-secure ^a DRS ^a

^aPurely dynamic measure.

Measure of Aggressive Incidents

Acts of aggression were coded from daily case notes in each inpatient's clinical file for the time period in question. Aggressive incidents were separated into *interpersonal violence* (including biting, hitting, kicking, punching, and throwing objects), *verbal threat* (including threats to kill or harm others), and *any aggression* (either of the above; Chu et al., 2013; Steadman et al., 1998).

For each follow-up period, each inpatient was allocated a score of 1 or 0, with a score of 1 indicating that they had committed at least one incident of interpersonal violence, and a score of 0 indicating that no incidents had occurred. Incidents of verbal threat were scored in the same way. If interpersonal violence and/or verbal threat had occurred at least once within a given period, a score of 1 was allocated to any aggression for that period. The short-term follow-up period spanned 0–1 month, and the medium-term follow-up periods spanned 0–3 and 0–6 months, respectively, based on Chu et al. (2013).

Procedure

Data were retrieved from clinical records for the risk-assessment measures. The risk measures had been scored by individual case managers (nurses), psychologists, or via team-based scoring at unique 13-weekly intervals based on the inpatient's date of admission. The 6-month period with the most completed risk-assessment measures was chosen. Risk-assessment data were collected before outcome data for each inpatient. To obtain outcome data, daily case notes written by hospital staff were reviewed for acts of aggression. Demographic information was also collected from patient files. All data were collected and de-identified prior to analysis in accordance with the institutional ethics approval using a Microsoft Excel spread sheet.

Statistical Analysis

Statistical analysis of the data was carried out using the Statistics Package for Social Sciences Version 21 (SPSS). Receiver Operating Characteristics (ROCs) were constructed for each of the measures, the three individual subscales of the HCR-20, and the C+R scales of the HCR-20, based on the sensitivity and specificity of the measure or subscale across all possible score thresholds (Hajian-Tilaki, 2013). Areas under each of the ROC curves (AUCs) were produced to give an indication of the ability of the measures and subscales to predict interpersonal violence, verbal threat, and any aggression over follow-up periods of 1, 3, and 6 months. Analysis was repeated for the forensic group and the civil group. Pearson's chi-square tests of association were used to calculate differences in the proportion of forensic and civil inpatients who were aggressive at least once within each of the follow-up periods for the different types of aggression. A priori, a Type I error of $\alpha = .05$ was assumed. However, consistent with Chu et al. (2013), False Discovery Rate (FDR) corrections were made within each time period for each type of aggression. These corrections control for Type I errors, which may arise when performing multiple comparisons of AUCs (Benjamini & Hochberg, 1995).

Results

Sample Characteristics

Inpatients' demographic information is shown in Table 3. Of the 74 inpatients included in the study, 37 (50%) were legally classified as forensic, 17 of whom resided in the medium-secure unit. The remaining 37 inpatients (50%) were legally classed as civil, 13 of whom resided in the medium-secure unit. All inpatients had received a primary mental health diagnosis. Over one-third had a secondary diagnosis of personality disorder. Most inpatients had previous psychiatric admissions, and over one-third had an offence

Table 3. Inpatient demographics.

Demographic	Total		Forensic		Civil	
	n = 74	%	n = 37	%	n = 37	%
<i>Security</i>						
Medium	30	40	17	46	13	35
Low	44	60	20	54	24	65
<i>Primary mental health diagnosis</i>						
Schizophrenia	60	81	29	78	31	84
Schizoaffective disorder	7	10	4	11	3	8
Bipolar disorder	5	7	2	5	3	8
Delusional disorder	1	1	1	3	0	0
Substance-induced psychosis	1	1	1	3	0	0
<i>Secondary mental health diagnosis</i>						
Psychopathy	8	11	8	22	0	0
Anti-social personality disorder	5	7	5	14	0	0
Anti-social PD and psychopathy	1	1	1	3	0	0
Traits of anti-social PD and/or psychopathy	10	14	3	8	7	19
Borderline personality disorder	3	4	1	3	2	5
Intellectual disability	18	24	9	24	9	24
Prior head injury	11	15	7	19	4	11
Previous psychiatric admissions	63	85	28	76	35	95
Previous forensic admissions	6	8	4	11	2	5
Offence history	29	39	14	38	15	41
Substance use history	51	69	29	78	22	60
Ever been in serious relationship	25	34	16	43	9	24

history. A majority of inpatients had a substance use history.

The number of inpatients who were aggressive (interpersonal violence, verbal threat, any aggression) at least once within each period is shown in Table 4. One-third of the total sample committed aggressive acts over the 6-month follow-up period, including 11 inpatients who engaged in physical violence towards patients or staff and 18 inpatients who verbally threatened harm.

More civil inpatients committed at least one act of interpersonal violence than their forensic counterparts over the 6-month follow-up period, $\chi^2(1) = 5.23, p < .05$. There was no difference between the proportion of civil or forensic inpatients who committed at least one act of verbal threat, or who

committed at least one act of any aggression, within any of the follow-up periods.

Predictive Validity of the Risk-Assessment Measures for the Total Sample

Predictive Accuracy for Interpersonal Violence

The AUCs produced by static and dynamic risk measures for predicting different types of aggression in the total sample across the follow-up periods can be seen in Table 5. All dynamic measures were excellent-to-outstanding predictors of interpersonal violence at 1-month follow-up (AUCs .85 to .95). In contrast, the static measures of H-scale and PLC-R were inadequate predictors. Similarly, only the dynamic measures were significant

Table 4. Number of total, forensic and civil inpatients who were aggressive at least once within each of the follow-up periods.

Aggressive behaviour	Up to 1 month			Up to 3 months			Up to 6 months		
	total (n = 74)	forensic (n = 37)	civil (n = 37)	total (n = 74)	forensic (n = 37)	civil (n = 37)	total (n = 74)	forensic (n = 37)	civil (n = 37)
Interpersonal violence	6 (8%)	2	4	10 (14%)	2	8	11 (15%)	2*	9*
Verbal threat	11 (15%)	5	6	16 (22%)	9	7	18 (24%)	11	7
Any aggression	14 (19%)	6	8	23 (31%)	10	13	26 (35%)	12	14

Note: Some participants committed both interpersonal violence and verbal threat within the same period.

*Significant difference based on Pearson's chi-square.

predictors of interpersonal violence in the medium-term follow-up periods, although the AUCs were classified as acceptable to excellent for 3-month (.73 to .85) and 6-month (.73 to .84) follow-up periods.

Predictive Accuracy for Verbal Threat

All dynamic measures were acceptable-to-outstanding predictors of verbal threat at 1-month follow-up (AUC's .77 to .93). In contrast, the static measures were inadequate predictors. At 3-month follow-up, the dynamic measures were acceptable-to-outstanding predictors of verbal threat (AUCs .77 to .92), and the static PCL-R was a modest predictor (AUC .67). At 6-month follow-up, all dynamic measures were modest-to-excellent predictors of verbal threat (AUCs .68 to .86), and the static PCL-R was a modest predictor (AUC .68) and thus comparable to some dynamic measures. The other static measure, the H-scale, remained an inadequate predictor over all three follow-up periods.

Predictive Accuracy for Any Aggression

At 1-month follow-up, dynamic measures were excellent-to-outstanding predictors of any aggression (AUCs .83 to .95), and static measures were inadequate predictors. At 3 months, dynamic measures were excellent-to-outstanding predictors of any aggression (AUCs .80 to .94), and the static PCL-R was a modest predictor (AUC .65). At 6-month follow-up, the dynamic measures were acceptable-to-outstanding predictors of any aggression (.73 to .90), and the static measures were modest predictors (AUCs .64 to .66), meaning all measures were significant predictors of any aggression at 6-month follow-up.

Predictive Validity of the Risk-Assessment Measures for Forensic Inpatients

The AUCs produced by the static and dynamic measures in predicting aggression at each follow-up period for the group of forensic inpatients can be seen in Table 6.

Table 5. Predictive accuracy of static and dynamic risk measures for inpatient aggression in the total sample.

Measure	Up to 1 month		Up to 3 months		Up to 6 months	
	AUC (SE)	95% CI	AUC (SE)	95% CI	AUC (SE)	95% CI
<i>Interpersonal violence</i>						
(s) H Scale	.63 (.11)	[.42, .84]	.68 (.08)	[.52, .83]	.64 (.08)	[.48, .80]
(s) PCL-R	.59 (.08)	[.43, .75]	.58 (.07)	[.43, .72]	.56 (.07)	[.42, .70]
(d) HCR-20 Total	.89* (.05)	[.80, .98]	.82* (.07)	[.69, .94]	.80* (.06)	[.67, .92]
(d) C Scale	.93* (.04)	[.85, 1.00]	.86* (.07)	[.72, 1.00]	.84* (.07)	[.71, .97]
(d) R Scale	.85* (.06)	[.74, .97]	.73* (.09)	[.56, .90]	.73* (.08)	[.57, .88]
(d) C+R Scale	.92* (.04)	[.85, .99]	.82* (.07)	[.69, .94]	.81* (.06)	[.68, .94]
(d) START	.94* (.03)	[.87, 1.00]	.85* (.06)	[.72, .96]	.84* (.06)	[.72, .95]
(d) MHOAT-Risk	.90* (.05)	[.79, 1.00]	.84* (.06)	[.72, .96]	.83* (.06)	[.72, .94]
(d) HoNOS-secure	.95* (.03)	[.91, 1.00]	.83* (.08)	[.68, .99]	.83* (.07)	[.68, .97]
(d) DRS	.89* (.04)	[.81, .98]	.78* (.08)	[.63, .93]	.78* (.07)	[.64, .92]
<i>Verbal threat</i>						
(s) H Scale	.54 (.10)	[.35, .73]	.55 (.09)	[.37, .73]	.58 (.08)	[.42, .74]
(s) PCL-R	.61 (.10)	[.42, .80]	.67* (.08)	[.51, .83]	.68* (.07)	[.54, .83]
(d) HCR-20 Total	.80* (.06)	[.68, .92]	.80* (.05)	[.70, .90]	.74* (.07)	[.61, .86]
(d) C Scale	.86* (.06)	[.67, .91]	.80* (.06)	[.69, .90]	.70* (.08)	[.56, .86]
(d) R Scale	.77* (.08)	[.62, .93]	.79* (.06)	[.67, .90]	.72* (.07)	[.59, .86]
(d) C+R Scale	.84* (.07)	[.71, .92]	.82* (.05)	[.72, .93]	.73* (.08)	[.58, .88]
(d) START	.93* (.03)	[.86, .99]	.92* (.03)	[.86, .98]	.86* (.06)	[.74, .97]
(d) MHOAT-Risk	.87* (.05)	[.76, .97]	.86* (.05)	[.77, .95]	.81* (.05)	[.70, .91]
(d) HoNOS-secure	.92* (.04)	[.84, .99]	.92* (.03)	[.86, .99]	.86* (.06)	[.74, .99]
(d) DRS	.79* (.06)	[.67, .91]	.77* (.05)	[.66, .87]	.68* (.07)	[.53, .82]
<i>Any aggression</i>						
(s) H Scale	.59 (.08)	[.43, .75]	.63 (.07)	[.49, .77]	.64* (.07)	[.51, .77]
(s) PCL-R	.60 (.08)	[.44, .77]	.65* (.07)	[.52, .78]	.66* (.07)	[.53, .79]
(d) HCR-20 Total	.85* (.05)	[.75, .95]	.85* (.04)	[.77, .94]	.80* (.05)	[.70, .90]
(d) C Scale	.89* (.05)	[.80, .99]	.85* (.05)	[.76, .94]	.78* (.06)	[.66, .90]
(d) R Scale	.83* (.07)	[.70, .95]	.80* (.06)	[.69, .91]	.76* (.06)	[.64, .87]
(d) C+R Scale	.89* (.05)	[.78, .99]	.86* (.04)	[.77, .95]	.79* (.06)	[.68, .91]
(d) START	.95* (.03)	[.91, 1.00]	.94* (.03)	[.89, .99]	.90* (.04)	[.82, .98]
(d) MHOAT-Risk	.89* (.04)	[.80, .98]	.89* (.04)	[.82, .96]	.86* (.04)	[.78, .94]
(d) HoNOS-secure	.95* (.03)	[.90, 1.00]	.94* (.04)	[.87, 1.00]	.90* (.05)	[.80, .99]
(d) DRS	.84* (.05)	[.73, .94]	.80* (.05)	[.70, .90]	.73* (.06)	[.61, .86]

Note: (s) = static measure. (d) = dynamic measure. Boldface font denotes significance after FDR corrections. **p* < .05.

Predictive Accuracy for Interpersonal Violence

While there was a clear distinction between static and dynamic measures’ ability to predict interpersonal violence in the forensic group across all three time periods, these

results were not significant following FDR corrections (see Table 6).

Predictive Accuracy for Verbal Threat

At 1-month follow-up, dynamic measures were excellent-to-outstanding predictors of

Table 6. Predictive accuracy of static and dynamic risk measures for inpatient aggression in the forensic group.

Measure	Up to 1 month		Up to 3 months		Up to 6 months	
	AUC (SE)	95% CI	AUC (SE)	95% CI	AUC (SE)	95% CI
<i>Interpersonal violence</i>						
(s) H Scale	.45 (.19)	[.08, .82]	.45 (.19)	[.08, .82]	.45 (.19)	[.08, .82]
(s) PCL-R	.32 (.08)	[.17, .48]	.32 (.08)	[.17, .48]	.32 (.08)	[.17, .48]
(d) HCR-20 Total	.88 (.08)	[.72, 1.00]	.88 (.08)	[.72, 1.00]	.88 (.08)	[.72, 1.00]
(d) C Scale	.97* (.03)	[.91, 1.00]	.97* (.03)	[.91, 1.00]	.97* (.03)	[.91, 1.00]
(d) R Scale	.94* (.04)	[.86, 1.00]	.94* (.04)	[.86, 1.00]	.94* (.04)	[.86, 1.00]
(d) C+R Scale	.96* (.04)	[.89, 1.00]	.96* (.04)	[.89, 1.00]	.96* (.04)	[.89, 1.00]
(d) START	.98* (.02)	[.93, 1.00]	.98* (.02)	[.93, 1.00]	.98* (.02)	[.93, 1.00]
(d) MHOAT-Risk	.97* (.03)	[.92, 1.00]	.97* (.03)	[.92, 1.00]	.97* (.03)	[.92, 1.00]
(d) HoNOS-secure	.97* (.03)	[.92, 1.00]	.97* (.03)	[.92, 1.00]	.97* (.03)	[.92, 1.00]
(d) DRS	.93* (.05)	[.84, 1.00]	.93* (.05)	[.84, 1.00]	.93* (.05)	[.84, 1.00]
<i>Verbal threat</i>						
(s) H Scale	.66 (.13)	[.40, .91]	.67 (.11)	[.45, .89]	.69 (.10)	[.50, .88]
(s) PCL-R	.77 (.10)	[.57, .97]	.78* (.08)	[.62, .95]	.76* (.08)	[.60, .91]
(d) HCR-20 Total	.87* (.08)	[.71, 1.00]	.90* (.05)	[.80, 1.00]	.82* (.08)	[.67, .96]
(d) C Scale	.84* (.09)	[.66, 1.00]	.83* (.07)	[.69, .97]	.68 (.11)	[.47, .89]
(d) R Scale	.88* (.10)	[.67, 1.00]	.89* (.06)	[.77, 1.00]	.82* (.07)	[.67, .97]
(d) C+R Scale	.88* (.09)	[.72, 1.00]	.90* (.05)	[.79, 1.00]	.75* (.11)	[.54, .95]
(d) START	.97* (.03)	[.91, 1.00]	.93* (.04)	[.85, 1.00]	.83* (.08)	[.67, 1.00]
(d) MHOAT-Risk	.97* (.03)	[.91, 1.00]	.92* (.05)	[.83, 1.00]	.86* (.06)	[.75, .98]
(d) HoNOS-secure	.96* (.03)	[.90, 1.00]	.97* (.03)	[.91, 1.00]	.87* (.08)	[.71, 1.00]
(d) DRS	.81* (.08)	[.65, .97]	.83* (.07)	[.69, .96]	.67 (.11)	[.46, .88]
<i>Any aggression</i>						
(s) H Scale	.67 (.11)	[.45, .90]	.69 (.10)	[.49, .89]	.71* (.09)	[.52, .89]
(s) PCL-R	.69 (.11)	[.47, .91]	.74* (.09)	[.56, .91]	.72* (.08)	[.56, .89]
(d) HCR-20 Total	.91* (.07)	[.78, 1.00]	.94* (.04)	[.86, 1.00]	.86* (.07)	[.73, .99]
(d) C Scale	.89* (.08)	[.73, 1.00]	.87* (.06)	[.76, .99]	.73* (.10)	[.53, .93]
(d) R Scale	.91* (.09)	[.74, 1.00]	.93* (.05)	[.83, 1.00]	.86* (.07)	[.73, .99]
(d) C+R Scale	.93* (.07)	[.79, 1.00]	.94* (.04)	[.86, 1.00]	.80* (.10)	[.60, .99]
(d) START	1.00* (.01)	[.99, 1.00]	.97* (.02)	[.92, 1.00]	.88* (.07)	[.73, 1.00]
(d) MHOAT-Risk	.99* (.01)	[.98, 1.00]	.95* (.03)	[.89, 1.00]	.90* (.05)	[.81, 1.00]
(d) HoNOS-secure	.99* (.01)	[.96, 1.00]	1.00* (.00)	[1.00, 1.00]	.91* (.07)	[.77, 1.00]
(d) DRS	.85* (.07)	[.72, .99]	.86* (.06)	[.75, .98]	.72* (.10)	[.52, .92]

Note: (s) = static measure. (d) = dynamic measure. Boldface font denotes significance after FDR corrections. **p* < .05.

verbal threat in the forensic group (AUCs .81 to .97), while the static measures were inadequate predictors. At 3-month follow-up, the dynamic measures were again excellent-to-outstanding predictors of verbal threat in the forensic group (AUCs .83 to .97), and the

static PCL-R was an acceptable predictor (AUC .78). At 6-month follow-up, only 6 of the 8 dynamic measures were acceptable-to-excellent predictors of verbal threat in the forensic group (.75 to .87), and the remaining dynamic DRS and C-scale were inadequate

predictors. Also at 6-month follow-up, the static PCL-R was an acceptable predictor of verbal threat in the forensic group (AUC .76), and the static H-scale was an inadequate predictor.

Predictive Accuracy for Any Aggression

At 1-month follow-up, dynamic measures were excellent-to-outstanding predictors of any aggression in the forensic group (AUCs .85 to 1.00), and the static measures were inadequate predictors. At 3-month follow-up, dynamic measures were again excellent-to-outstanding predictors of any aggression in the forensic group (AUCs .86 to 1.00), and the static PCL-R was an acceptable predictor (AUC .74). At 6-month follow-up, all measures were significant predictors of any aggression in the forensic group. Dynamic measures were acceptable-to-outstanding predictors (AUCs .72 to .91), and both static measures were acceptable predictors (AUCs .71 to .72).

Predictive Validity of the Risk-Assessment Measures for Civil Inpatients

The AUCs produced by the static and dynamic measures in predicting aggression at each follow-up period for the group of civil inpatients can be seen in [Table 7](#).

Predictive Accuracy for Interpersonal Violence

As shown in [Table 7](#), at 1-month follow-up, 7 of the 8 dynamic measures were excellent-to-outstanding predictors of interpersonal violence in the civil group (AUCs .86 to .95), while the dynamic R-scale and the static measures were inadequate. At 3-month follow-up, 5 of the 8 dynamic measures were acceptable-to-excellent predictors of interpersonal violence in the civil group (AUCs .78 to .84), while the dynamic R-scale, C+R scale, and DRS were inadequate predictors. Also at 3 months, both static measures were acceptable-to-excellent predictors of interpersonal

violence (AUCs .77 to .85). At 6-month follow-up, 5 of the 8 dynamic measures were acceptable-to-excellent predictors of interpersonal violence in the civil group (AUCs .75 to .78), while the dynamic R-scale, C+R scale, and DRS were inadequate predictors. Also at 6 months, both static measures were acceptable predictors (AUCs .75 to .78).

Predictive Accuracy for Verbal Threat

At the 1-, 3-, and 6-month follow-ups, only four dynamic measures (C-scale, C+R scale, HoNOS-secure and START) were significant predictors of verbal threat in the civil group after FDR corrections. They produced similar results across time periods, with excellent-to-outstanding AUCs at 1 month (.81 to .90) and acceptable-to-outstanding AUCs over 3 months (.79 to .90) and 6 months (.79 to .90). All other dynamic and static measures were inadequate predictors of verbal threat in the civil group in all follow-up periods.

Predictive Accuracy for Any Aggression

At 1, 3, and 6-month follow-ups, all dynamic measures were significant predictors of any aggression in the civil group, except the dynamic R-scale, which was inadequate over all follow-up periods. The significant dynamic measures produced similar results across time periods, with acceptable-to-outstanding AUCs over 1-month (.79 to .94), 3-month (.74 to .92) and 6-month (.73 to .92) follow-ups. Neither of the static measures were adequate predictors of any aggression in the civil group over any of the follow-up periods.

Discussion

Findings, Comparisons, and Implications

The current study has clearly demonstrated that dynamic risk measures are better than static risk measures at predicting aggression

Table 7. Predictive accuracy of static and dynamic risk measures for inpatient aggression in the civil group.

Measure	Up to 1 month		Up to 3 months		Up to 6 months	
	AUC (SE)	95% CI	AUC (SE)	95% CI	AUC (SE)	95% CI
<i>Interpersonal violence</i>						
(s) H Scale	.81* (.08)	[.66, .97]	.85* (.06)	[.72, .97]	.78* (.09)	[.61, .95]
(s) PCL-R	.80 (.10)	[.59, 1.00]	.77* (.09)	[.59, .94]	.75* (.09)	[.58, .92]
(d) HCR-20 Total	.91* (.05)	[.80, 1.00]	.78* (.10)	[.59, .96]	.75* (.09)	[.56, .93]
(d) C Scale	.90* (.07)	[.76, 1.00]	.79* (.11)	[.58, 1.00]	.76* (.10)	[.56, .96]
(d) R Scale	.77 (.10)	[.57, .97]	.58 (.13)	[.33, .84]	.59 (.12)	[.36, .82]
(d) C+R Scale	.88* (.07)	[.74, 1.00]	.71 (.11)	[.49, .93]	.70 (.10)	[.50, .90]
(d) START	.95* (.04)	[.87, 1.00]	.83* (.08)	[.68, .99]	.84* (.07)	[.69, .98]
(d) MHOAT-Risk	.89* (.07)	[.75, 1.00]	.84* (.08)	[.69, 1.00]	.84* (.07)	[.69, .98]
(d) HoNOS-secure	.95* (.03)	[.89, 1.00]	.80* (.11)	[.58, 1.00]	.79* (.10)	[.59, .99]
(d) DRS	.86* (.07)	[.71, 1.00]	.69 (.11)	[.47, .92]	.68 (.10)	[.48, .89]
<i>Verbal threat</i>						
(s) H Scale	.39 (.13)	[.14, .64]	.32 (.12)	[.09, .56]	.32 (.12)	[.09, .56]
(s) PCL-R	.48 (.13)	[.22, .73]	.45 (.12)	[.21, .68]	.45 (.12)	[.21, .68]
(d) HCR-20 Total	.73 (.10)	[.53, .92]	.69 (.10)	[.50, .88]	.69 (.10)	[.50, .88]
(d) C Scale	.90* (.05)	[.80, 1.00]	.85* (.07)	[.71, .99]	.85* (.07)	[.71, .99]
(d) R Scale	.64 (.11)	[.42, .87]	.66 (.10)	[.45, .86]	.66 (.10)	[.45, .86]
(d) C+R Scale	.81* (.09)	[.64, .98]	.79* (.08)	[.63, .95]	.79* (.08)	[.63, .95]
(d) START	.90* (.06)	[.78, 1.00]	.90* (.06)	[.79, 1.00]	.90* (.06)	[.79, 1.00]
(d) MHOAT-Risk	.76* (.10)	[.57, .95]	.75* (.09)	[.58, .93]	.75* (.09)	[.58, .93]
(d) HoNOS-secure	.87* (.06)	[.75, 1.00]	.86* (.06)	[.74, .98]	.86* (.06)	[.74, .98]
(d) DRS	.78* (.09)	[.60, .95]	.76* (.09)	[.59, .93]	.76* (.09)	[.59, .93]
<i>Any aggression</i>						
(s) H Scale	.51 (.12)	[.27, .74]	.58 (.10)	[.37, .78]	.54 (.10)	[.34, .75]
(s) PCL-R	.55 (.12)	[.32, .78]	.59 (.10)	[.39, .78]	.59 (.10)	[.40, .78]
(d) HCR-20 Total	.79* (.08)	[.63, .95]	.74* (.08)	[.58, .90]	.73* (.08)	[.57, .89]
(d) C Scale	.90* (.05)	[.80, 1.00]	.84* (.07)	[.69, .98]	.82* (.07)	[.68, .97]
(d) R Scale	.72 (.10)	[.52, .91]	.63 (.10)	[.43, .83]	.64 (.10)	[.44, .83]
(d) C+R Scale	.84* (.07)	[.70, .98]	.76* (.08)	[.60, .92]	.76* (.08)	[.60, .92]
(d) START	.94* (.04)	[.85, 1.00]	.92* (.05)	[.82, 1.00]	.92* (.04)	[.84, 1.00]
(d) MHOAT-Risk	.80* (.08)	[.64, .96]	.83* (.07)	[.70, .96]	.83* (.07)	[.70, .96]
(d) HoNOS-secure	.93* (.05)	[.83, 1.00]	.87* (.07)	[.73, 1.00]	.87* (.07)	[.74, 1.00]
(d) DRS	.81* (.08)	[.66, .97]	.73* (.09)	[.57, .90]	.73* (.08)	[.57, .90]

Note: (s) = static measure. (d) = dynamic measure. Boldface font denotes significance after FDR corrections. **p* < .05.

over the short term. This robust finding was seen in both civil and forensic inpatient groups and is consistent with Chu et al. (2013).

Dynamic measures were superior to static measures at predicting interpersonal violence, verbal threat, and any aggression over the 1-

month follow-up for the total sample. The hypothesis that dynamic measures would be better predictors of aggression than static measures over the short term was therefore supported. The dynamic measures were also superior to static measures over the medium

term (3-month and 6-month follow-ups). Therefore, the hypothesis that static measures would be better predictors of aggression than dynamic measures over the medium term was not supported.

While both Chu et al. (2013) and the present study demonstrated that dynamic measures are better predictors of aggression than static measures over the short term, neither study was able to demonstrate that static measures are superior to dynamic measures over the medium term. This was despite the rationale that the medium- to low-secure setting used in the present study may display a different pattern of aggression from that of the high-secure setting used by Chu et al. (2013). Importantly, however, the present results show that the static PCL-R was a better predictor of verbal threat and any aggression in the medium term than it was in the short term, and that the static H-scale was a better predictor of any aggression at 6 months than it was at 1 month or 3 months. It therefore appears that the static measures in the present study were better predictors of any aggression and verbal threat over the medium term than the short term. This is in contrast to Chu et al. (2013), who found that static measures were inadequate predictors over all time periods. It is however consistent with studies that have found static measures to be good predictors of aggression over the longer term (Quinsey, Harris, Rice, & Cormier, 1998). As we did no testing to compare the AUCs of the measures across time-frames, we were unable to determine whether dynamic measures were better predictors over the short term than the medium term in the total sample. As the 3-month and 6-month follow-ups are inclusive of the previous periods (i.e., 3-month includes 1 month, 6-month includes both 1 and 3 months), the data are highly correlated, and so this would warrant further research and more sophisticated analysis.

In comparing to Chu et al. (2013), it is apparent that the AUCs produced by most measures in the present study were very high, to the point of producing AUCs of 1.00 in

some instances. One possible explanation may be that 'SPSS software allows to depict [sic] ROC curve [sic] in unit square space by trapezoidal rule, (i.e., nonparametric method) and nonparametric estimate of AUC and its SE and 95% CI' (Hajian-Tilaki, 2013; pp. 5–632). This method of estimation of AUC may not be the best fit for the data. Another possible explanation may involve potential work-up bias (Hajian-Tilaki, 2013). As the measures examined in this study are routinely scored for each inpatient every 13 weeks, and thus most inpatients would have been scored on each measure several times, this may have resulted in better prediction of violence and therefore higher AUCs. This is a point of difference of the present study compared with Chu et al. (2013), who scored the inpatients at admission. A third possible explanation is that experienced mental health clinicians who had extensive knowledge and training in the measures completed the risk measures they administered and with the inpatients they were coding. While previous studies have found no difference between blind researchers and non-blind treating clinicians in the rating of risk-assessment measures (de Vogel & de Ruiters, 2004; de Vogel & de Ruiters, 2006), none of the assessors in such studies had prior experience with the chosen risk-assessment measure. It is therefore unknown whether a combination of rating clinicians being non-blind and having expertise in the use of the measures may have impacted on the predictability for aggression and therefore the AUC.

The high predictive ability of dynamic measures in the short term was demonstrated in both the present medium- to-low-secure setting and the high-secure setting of Chu et al. (2013). This suggests that the measures perform similarly at predicting aggression regardless of imposed restrictions and levels of supervision. It is therefore recommended that dynamic measures should be used when predicting risk of aggression in psychiatric inpatients in the short term, regardless of security level. Furthermore, the finding that static measures appear to be better predictors

over the medium than the short term suggests that the advantage of dynamic over static measures decreases as time from prediction increases. It therefore remains possible that there is a longer follow-up period in which static measures are better predictors of aggression than dynamic measures. However, the results of the present study and those of Chu et al. (2013) suggest that it is unlikely this time point occurs within 6 months from prediction.

A second aim of the present study was to examine whether dynamic and static measures performed similarly for forensic and civil groups. The results showed that dynamic measures outperformed static measures over both the short and the medium term in both the forensic and civil groups, as they had done in the total sample. Therefore the hypothesis that the patterns of predictive ability demonstrated by dynamic and static measures in each of the forensic and civil groups would be similar to those demonstrated in the total sample was supported. However, there were some inconsistencies with regards to precisely which and how many measures were predictive of the different types of aggression within each group. Most notably, a number of dynamic scales were inadequate predictors for various aggression types and follow-up periods in the civil group. This is inconsistent with the results of the total sample in which all dynamic measures were significant predictors across all conditions. Specifically, the R-scale was not a significant predictor of any type of aggression at any period in the civil group. Furthermore, half of the dynamic measures (R-scale, HCR-20, MHOAT-Risk, and DRS) were inadequate predictors of verbal threat across all three follow-up periods for the civil group. In addition, the two static measures were acceptable-to-excellent predictors of interpersonal violence at the 6-month follow-up, which is inconsistent with the total sample where neither measure was predictive.

In general, the results suggest dynamic measures should be used for short-term

predictions of aggression in both forensic and civil groups. There is some suggestion that the R-scale may not perform as well as other dynamic measures when predicting aggression in civil inpatients. This holds implications for the use of the HCR-20 with such groups. These and other dynamic measures, such as the MHOAT-Risk and DRS, may also be inadequate predictors of verbal threat in civil inpatients. However, it must be noted that the analysis of forensic and civil groups in this study is based on small samples. Therefore these claims should be investigated further before any clinical recommendations can be made.

A third aim of the present study was to compare the risk-assessment performance of measures routinely administered within NSW Health, namely the MHOAT-Risk, HoNOS-secure, and DRS. These measures performed consistently with the more established dynamic risk-assessment measures across aggression types and follow-up periods. Therefore, the hypothesis that the MHOAT-Risk and HoNOS-secure would have less predictive ability than more established dynamic risk measures was not supported, and the hypothesis that the DRS would have similar predictive ability to more established dynamic risk measures was supported.

While the HoNOS-secure was not designed as a risk-assessment measure, it seems it would be a useful clinical measure to inform and direct risk-assessment processes in the company of the MHOAT-Risk. Both of these measures produced predictive ability consistent with more established risk-assessment measures and have the advantage of being efficient to score. A further benefit of the HoNOS-secure is that it provides information regarding need for more comprehensive risk assessment. Therefore, the routine use of the MHOAT-Risk and the HoNOS-secure as screening measures in the present setting appears justified. A wider battery of the more established measures could then be utilised when a more thorough risk assessment is indicated by these measures. While a

case could be made to use the DRS instead of the full HCR-20 for short-term risk-of-violence prediction due to it containing more than half the number of items, the 5-item C-scale of the HCR-20 performed just as well across aggression types and follow-up periods. Thus, there is support for the use of the C-scale when making time-efficient short-term risk predictions in similar settings.

Limitations

While attempts were made to address some limitations of Chu et al. (2013), certain shortcomings were unable to be rectified. As seen in many studies (Owen, Tarantello, Jones, & Tennant, 1998), staff may have under-reported incidents of aggression within the client records. It is likely that staff working in inpatient settings have a higher threshold for reporting incidents of aggression, particularly verbal threat, which may be interpreted as being less serious (Owen, Tarantello, Jones, & Tennant, 1998). Furthermore, while the present sample had lower security and fewer restrictions than the sample in Chu et al. (2013), it is still likely that immediate and responsive interventions to aggression would have limited future incidents of aggression. Coupled with the necessarily small sample size due to the nature of the units, the present study was based on low levels of aggression. This is a common issue in the risk-of-violence literature and limits the findings of the present study, particularly when interpreting the results of the forensic and civil groups.

An additional limitation of the present study is that various clinicians performing both individual and team-based ratings originally scored the risk measures. While the results demonstrate the effectiveness of measures as used in real-world clinical practice, differences in accuracy between these raters and methods were not accounted for. Furthermore, as the data retrieval point was selected based on availability of data, it is unknown whether differences exist between these time points and others with less complete data.

Finally, the measures were not re-administered at each time point in either study, which may have provided information about their ability to detect dynamic fluctuations in risk state.

In terms of statistical limitations, *p*-values as opposed to confidence intervals (CIs) were utilised to determine whether a measure had statistically significant predictive validity. However, the *p*-value and the confidence interval are both estimates (Hajian-Tilaki, 2013), and our results show some inconsistencies between the two, particularly when the sample was split into forensic and civil groups. It would be expected that where an AUC has a high *p*-value, the 95% CI would be inclusive of .5, meaning that the AUC was not significantly different from chance. However, this was not always the case in the present results (see Table 7 results for R-scale at predicting interpersonal violence at 1 month for an example). Such inconsistencies warrant further investigation.

Future Research

The current study focused on examining the predictive ability of static and dynamic measures in a psychiatric inpatient setting over the short to medium term. Future research should test the predictive accuracy of static and dynamic measures over long-term follow-up periods of at least one to two years in order to examine whether static measures are better predictors of aggression than dynamic measures over a longer time period, and how long after prediction this may occur. More advanced statistical analysis of comparisons of AUCs between time-periods would be useful in conducting such research. Ideally a comparison would also be made between forensic and civil samples to replicate the finding that dynamic and static risk assessments performed similarly with both groups, and between aggression types to determine whether this occurs in a similar fashion for predicting all types of aggression.

The prediction and management of aggression displayed by psychiatric inpatients

remains a critical component of their rehabilitation and care. Measures that focus on dynamic risk factors have been shown to successfully predict aggression in civil and forensic inpatients over the short term. The continued development and implementation of such measures is likely to result in improved risk-management strategies within these settings, resulting in an increase in the safety of inpatients, staff members, and visitors alike.

Declaration of Interest

The authors report no declarations of interest

Geolocation Information

Newcastle and Lake Macquarie, NSW, Australia

References

- Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*, 27–51.
- Andrews, D. A., & Bonta, J. (1998). *The level of service inventory-revised: Screening version*. Toronto, ON: Multi-Health Systems.
- Australian Institute of Health and Welfare. (2011). *Mental health services in Australia*. Retrieved from <http://www.aihw.gov.au/publication-detail/?id=10737420191>
- Belfrage, H., & Douglas, K. S. (2002). Treatment effects on forensic psychiatric patients measured with the HCR-20 violence risk assessment scheme. *International Journal of Forensic Mental Health*, *1*, 25–36.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, *57*, 289–300.
- Braithwaite, E., Charette, Y., Crocker, A., & Reyes, A. (2010). The predictive validity of clinical ratings of the Short-Term Assessment of Risk and Treatability (START). *International Journal of Forensic Mental Health*, *9*, 271–281.
- Chippis, J., Raphael, B., & Coombs, T. (2002). The mental health outcomes and assessment tools training project: Creating the foundation for improved quality of care. *NSW Public Health Bulletin*, *13*, 237–238.
- Chu, C. M., Thomas, S. D. M., Ogloff, J. R. P., & Daffern, M. (2011). The predictive validity of the short-term assessment of risk and treatability (START) in a secure forensic hospital: Risk factors and strengths. *International Journal of Forensic Mental Health*, *10*, 337–345.
- Chu, C. M., Thomas, S. D., Daffern, M., & Ogloff, J. R. P. (2013). The short- to medium-term predictive accuracy of static and dynamic risk assessment measures in a secure forensic hospital. *Assessment*, *20*, 230–241.
- de Vogel, V., & de Ruiter, C. (2004). Differences between clinicians and researchers in assessing risk of violence in forensic psychiatric patients. *The Journal of Forensic Psychiatry & Psychology*, *15*, 145–164.
- de Vogel, V., & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime and Law*, *12*, 321–336.
- Dickens, G., Sugarman, P., & Walker, L. (2007). HoNOS-secure: A reliable outcome measure for users of secure and forensic mental health services. *Journal of Forensic Psychiatry and Psychology*, *18*, 507–514.
- Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20V3: Assessing risk of violence – User guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Douglas, K. S., & Ogloff, J. R. P. (2003). Multiple facets of risk for violence: The impact of judgmental specificity on structured decisions about violence risk. *International Journal of Forensic Mental Health*, *2*, 19–34.
- 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.
- Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law*, *11*, 347–383.
- Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine*, *4*(2), 627–635.
- Hare, R. D. (2003). *The hare psychopathy checklist-revised manual* (2nd ed.). North Tonawanda, NY: Multi-Health Systems.
- Heilbrun, K., Yasuhara, K., & Shah, S. (2010). Violence risk assessment tools: Overview and critical analysis. In R. K. Otto & K. S. Douglas

- (Eds.), *Handbook of violence risk assessment* (pp. 1–17). New York, NY: Routledge.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley.
- Leistico, A. R.; Salekin, R. T.; DeCoster, J.; Rogers, R. (2008). *Law and Human Behavior*, 32 (1), 28–45.
- McNiel, D. E., Gregory, A. L., Lam, J. N., Binder, R. L., & Sullivan, G. R. (2003). Utility of decision support tools for assessing acute risk of violence. *Journal of Consulting and Clinical Psychology*, 71, 945–953.
- Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. New York, NY: Oxford University Press.
- New South Wales Health. (2001). Mental Health Outcome Assessment Training Resources. Retrieved from <http://www.health.nsw.gov.au/mhdao/DM/Pages/professionals.aspx>
- Nicholls, T. L., Brink, J., Desmarais, S., Webster, C. D., & Martin, M. L. (2006). The short-term assessment of risk and treatability (START). *Assessment*, 13, 313–327.
- Owen, C., Tarantello, C., Jones, M., & Tennant, C. (1998). Repetitively violent patients in psychiatric units. *Psychiatric Services*, 49, 1458–1461.
- Pirkis, J., Burgess, P., Kirk, P., Dodson, S., & Coombs, T. (2005). *Review of standardised measures used in the national outcomes and casemix collection (NOCC)*. Sydney: NSW Institute of Psychiatry
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.
- Royal College of Psychiatrists. (2015). HoNOS-secure. Retrieved from <http://www.rcpsych.ac.uk/quality/honos/secure.aspx>
- Shrinkfield, G., & Ogloff, J. (2015). Use and interpretation of routine outcome measures in forensic mental health. *International Journal of Mental Health Nursing*, 24, 11–18.
- Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression of studies involving 25,980 participants. *Clinical Psychology Review*, 31, 499–513.
- Singh, J. P., Desmarais, S. L., & Van Dorn, R. A. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioural Sciences and Law*, 31, 55–73.
- Steadman, H. J., Mulvey, E. P., Monahan, J., Robbins, P. C., Appelbaum, P. S., Grisso, T., & Silver, E. (1998). Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods. *Archives of General Psychiatry*, 55, 393–401.
- Steinert, T. (2002). Prediction of inpatient violence. *Acta Psychiatrica Scandinavica*, 106, 133–141.
- van den Brink, R. D. (2010). Routine violence risk assessment in community forensic mental healthcare. *Behavioral Sciences & The Law*, 28, 396–410.
- Walters, G. D. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factor scores: A meta-analysis. *Law and Human Behavior*, 27, 541–558.
- Webster, J., Bretherton, F., Goulter, N., & Fawcett, L. (2013). Does an educational intervention improve the usefulness of the Health of the Nation Outcome Scales in an acute mental health setting? *International Journal of Mental Health Nursing*, 22, 322–328.
- Webster, C. D., Douglas, K. S., Belfrage, H., & Link, B. (2000). Capturing change: An approach to managing violence and improving mental health. In S. Hodgins & R. Müller-Isberner (Eds.), *Violence among the mentally ill* (pp. 119–144). Dordrecht, Netherlands: Kluwer/Academic.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk of violence (Version 2)*. Burnaby, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Manual for the short-term assessment of risk and treatability (START: Version 1.1)*. Port Coquitlam, British Columbia, Canada: British Columbia Mental Health and Addiction Services.