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
BPsych (Hons)

This thesis is submitted in partial fulfilment of the requirements for the degree of Master of Clinical Psychology, School of Psychology, University of Newcastle, Australia

March 2015

Statement of Originality

The thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968.

Signed ______________________

Brayden Finch
Master of Clinical Psychology candidate

Acknowledgment of Authorship

I hereby certify that the work embodied in this thesis contains a manuscript of which I am a joint author. I have included as part of the thesis a written statement, endorsed by my supervisor, attesting to my contribution to the joint manuscript:

I, Brayden Finch, attest that I was primarily responsible for the review of literature and the writing of the critical review, data analysis, and writing of the manuscript contained within this thesis. I received support from Megan Valentine, statistics consultant, for the data analysis and reporting of method and results. My work was forwarded to supervisors Dr Derek Gilligan and Dr Sean Halpin for review, and amendments were made based on their feedback. I am the primary author of the work contained herein.

Signed _______________________

Brayden Finch
Master of Clinical Psychology candidate

Signed________________________

Sean A. Halpin, PhD
Academic Supervisor

Acknowledgement of Financial Support

This thesis has been supported by an award from the NAHSSS funded by the Commonwealth Department of Health and Ageing. The views expressed in this thesis do not necessarily represent those of the NAHSSS, its Administrator, Services for Australian Rural and Remote Allied Health (SARRAH) and/or the Government Department of Health and Ageing.

Acknowledgement of Professional and Personal Support

To my colleagues, my family, my friends, my partner Alice, my clinical and research supervisors Derek and Sean, my clinical supervisors Virginia and Louise, and my statistics consultant Megan – I sincerely thank you for all your encouragement and support. You ROC.

Table of Contents

List of tables:

Thesis Abstract

The prediction and subsequent management of aggression in psychiatric inpatients is a crucial role of the mental health professional. This retrospective cohort study combines a critical literature review and a research paper on the ability of static and dynamic risk assessment measures to predict aggression in psychiatric inpatients in the short- to medium-term. The critical literature review presents an overview of the development of static and dynamic risk assessment measures before critically discussing the research on violence prediction in psychiatric inpatient populations over the short- to medium-term. The research paper 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 over 1-, 3-, and 6-month follow-up periods. Retrospective file records were sourced to conduct an AUC analysis of the ROC curve for short and medium follow-up periods to test the ability of each measure to predict interpersonal violence, verbal threat, and any aggression. 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.

Critical Literature Review

Overview

The prediction and management of aggression displayed by individuals with mental health disorders is a prominent issue in modern corrective and psychiatric settings (Belfrage & Douglas, 2002; Owen, Tarantello, Jones, & Tennant, 1998). Aggression can be defined as any behaviour directed towards another individual with the intent of causing immediate harm (Anderson & Bushman, 2002). Violence is one type of aggression that has extreme harm as its goal (Anderson & Bushman, 2002). There exists a public and legal expectation that mental health professionals are competent in the prediction and management of 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 risk assessment measures have been developed in an attempt to more accurately predict when a patient might become violent. Recently, validation of these measures has become the focus of a body of risk assessment literature. Such research aims to assist mental health professionals and organisations to make informed decisions about the best available methods of violence risk assessment in a particular context. Differences in the types of risk factors used in such measures (i.e. static and unchanging versus dynamic and changing) have been studied in various psychiatric samples to determine which type of measure most accurately predicts violence in a given setting and over a given time frame. The current critical review examines this literature, providing an overview of the development of such literature and a discussion of studies that have tested the predictive ability of static and dynamic risk measures in inpatient settings over prediction periods of up to 6 months.

Importance

A decision about level of dangerousness to others is a common standard for commitment of an individual with mental illness to an inpatient, outpatient or forensic unit (Monahan et al., 2001). Individuals who display violence, the most extreme form of aggression (Anderson & Bushman, 2002), are viewed as most dangerous to others. Therefore, predicting and managing violence is of high interest to treating organisations, professionals, and researchers (Monahan et al., 2001). While violence can cause significant harm in the community, violence in institutional and inpatient settings are a frequent and serious problem (Owen et al., 1998). Staff, patients, and visitors can be physically or psychologically injured, property can be destroyed, and treatment programs disrupted (Gadon, Johnstone, & Cooke, 2006). For the individual who enacts violence, the result can include lengthier and more restrictive admissions at significant financial cost to service providers (Gadon, Johnstone & Cooke, 2006). In Australia, the average cost of a forensic bed per day is $942 (Australian Institute of Health and Welfare, 2011). Predicting violence within a mental health setting is therefore highly important, not only in ensuring the safety of inpatients, staff, and visitors, but for the service-level implications of decisions regarding admission and discharge.

Development of Risk Assessment

Despite the importance of violence prediction, the risk assessment field is relatively young (Ogloff & Davis, 2005). Three generations of risk assessment have been identified. Professor John Monahan (1984), a leading researcher in the field, coined the phrase “first generation” risk research in reference to the limited number of studies available pre-1980s. This early research was criticised as relying heavily on the unstructured, idiosyncratic opinions of the predicting clinicians (Monahan, 1981).

More than half a century of risk assessment research has continually revealed that this type of unstructured clinical judgement alone is inaccurate, even for experienced clinicians (Ægisdóttir et al., 2006; Meehl, 1954). Monahan (1981) illustrated this point when he reported that psychiatrists and psychologists were correct in no more than one in three predictions of violence over follow-up periods of several years.

Monahan (1984) subsequently called for a second generation of risk research employing enhanced strategies to overcome the shortcomings of the unstructured clinical judgement approach. This came to fruition in the early 1980s, with a body of research that focused on short prediction time-frames, situational variables, and specified populations (Ogloff & Davis, 2005; Otto, 1992). The use of clinical impression as a reliable form of violence prediction was largely abandoned during this period, as the low accuracy resulted in a common view that this was an unethical practice (Ogloff & Davis, 2005). However, some commenters argued that violence prediction could be ethical if it was performed using actuarial assessment (Grisso & Appelbaum, 1992). This new method began with researchers reviewing the files of prisoners or psychiatric inpatients and coding the available information into potential risk factors (Ogloff & Davis, 2005). The follow-up records of such prisoners or inpatients were then obtained to determine which individuals had gone on to become violent (Ogloff & Davis, 2005). Using statistical analysis, those factors were identified which, when combined, were most closely related to those individuals who had gone on to behave violently (Ogloff & Davis, 2005). This resulted in the development of actuarial risk assessment measures, which gave risk factors numerical values to produce a probabilistic estimate of the individual's likelihood of violence (Singh, Grann, & Fazel, 2011). Each individual was appraised using the same criteria to arrive at an ordinal risk score which could be compared (Singh, Grann, & Fazel, 2011). The use of actuarial assessments in the prediction of violence is well-supported by research evidence (Quinsey, Harris, Rice, & Cormier, 1998). Mossman's (1994) meta-analysis on the actuarial prediction of violence found the combined accuracy of actuarial assessments was significant higher than that of unstructured clinical judgement. Similarly, Bonta, Law, and Hanson (1998) found that objective risk assessment was the best predictor of violence compared with a host of other variables, including historical factors, personality factors and substance use factors.

Despite the strong support for actuarial tools, there is some controversy associated with their use. It has been noted that using large samples is important when creating and validating actuarial tools to ensure confidence intervals remain narrow and risk categories do not overlap (Heilbrun, Yasahura, & Shah, 2010). However, the resulting confidence intervals are necessarily wide when applied to violence prediction of a single case, meaning there is a large degree of overlap between categories which limits risk prediction accuracy for that single case (Hart, Michie, & Cook, 2007). This point has been extensively debated in the literature (Harris & Rice, 2007; Heilbrun, Douglas & Yasahura, 2009; Mossman, 2007). Further, there has been contention that the weighting and scoring of variables in actuarial tools can mean that idiosyncratic factors important to a single case are overlooked (Heilbrun, Yasahura, & Shah, 2010). However, other evidence suggests weighting does not significantly alter results (Grann & Langstrom, 2007). Finally, while actuarial tools improve the predictive abilities of a clinician, there may be a limit to the extent of this improvement, and so it is argued newer improved theory and methods should be developed (Kroner, Mills, & Reddon, 2005).

A third generation of risk assessment has been identified more recently (Heilbrun, Ogloff, & Picarello, 1999). This generation is characterised by its aims –

rather than attempting to determine whether an individual will be violent or not, third-generation measures attempt to assess an individual's level of risk for violence (Ogloff & Davis, 2005). This generation of risk assessment considers a broader range of risk factors than the second generation, which used mostly unchanging historical factors, termed "static" factors (Douglas & Skeem, 2005). Examples of static factors include a history of violence, relationship instability or a diagnosis of a major mental illness. While some third generation measures utilise these static factors, they also consider factors that are more changeable and modifiable, termed "dynamic" factors (Douglas & Skeem, 2005). Examples of dynamic factors include negative attitudes, lack of support or active symptoms of mental illness. This approach has seen a shift from the mere prediction of violence towards what has been said to be the more important goal of risk assessment – the management of that risk (Daffern & Howells, 2002; Heilbrun, 1997; Ogloff & Davis, 2005).

Alongside continued development of actuarial methods, a new method of risk assessment was developed with this third generation. Structured Professional Judgement (SPJ) uses a combination of actuarial methods and structured assessment approaches to inform clinical decision making (Douglas & Ogloff, 2003). In addition to the information gathered for the actuarial component, assessors also collect wider collateral information via interview, records, and other sources (Heilbrun, Yasahura, & Shah, 2010). The assessor then utilises the score of the actuarial assessment in combination with this collateral information to make their own clinical decision regarding the level of risk, usually as high, medium or low (Heilbrun, Yasahura, & Shah, 2010).

Analysis of Risk Measures

A number of studies have attempted to compare actuarial and SPJ risk assessments to determine which is the better method of violence prediction (Catchpole & Gretton, 2003; Dahle, 2006; de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004; Douglas, Yeomans, & Boer, 2005). While it has been argued that SPJ measures should only be tested 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 due to broader methodology (actuarial vs SPJ) from differences due to particular tools (Heilbrun, Yasahura, & Shah, 2010). Therefore, when comparing SPJ tools to actuarial ones, a standardised method of evaluating the actuarial properties of all 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). In a risk of violence context, the AUC is an index of how well a risk assessment tool discriminates between offenders and non-offenders across all possible cut-off scores. The evidence thus far suggests the actuarial and SPJ methods are at least comparable predictors when examined via this method (Heilbrun, Yasahura, & Shah, 2010).

While the AUC of the ROC curve is the preferred statistical method for many experts in the violence prediction field, there are issues associated with its use. There is little consistency between papers in terms of cut-off scores for reporting the strength of an AUC (Singh, Desmarais, & Van Dorn, 2013). There is also concern that the AUC technique is subject to misuse, including being interpreted with far too much confidence (Sjöstedt & Grann, 2002). Problems also exist when interpreting AUCs in a meta-regression or systematic way, as the method does not allow for exploration of sources of heterogeneity (Thompson & Higgins, 2002).

Static vs. Dynamic Factors

Risk factors are often classified according to their potential responsiveness to treatment interventions (Belfrage & Douglas, 2002). Risk factors can be classified as being either static or dynamic in nature. Static factors are those that do not fluctuate over time, such as gender or a history of violence (Andrews & Bonta, 1998). Static factors therefore relate closely to the idea of risk status, which is an enduring categorisation of a person’s general level of risk (Douglas & Skeem, 2005). Conversely, dynamic risk factors fluctuate with time and circumstances, such as anti-social attitudes or active psychosis (Andrews & Bonta, 1998). Dynamic factors are more closely related to a person’s risk state, or their current level of risk (Douglas & Skeem, 2005). The majority of research on risk measures in a mental health context has been performed on instruments focussed on static risk factors (Douglas & Skeem, 2005). Importantly, static risk factors are typically unmalleable to change, whereas dynamic risk factors might be reduced following targeted intervention, resulting in an overall reduction of risk (Webster, Douglas, Belfrage, & Link, 2000). It is therefore worthwhile to further investigate the utility of measures incorporating dynamic risk factors, for both prediction of violence, and ongoing management of risk (Heilbrun, 1997).

Civil vs. Forensic Inpatients

It is critical that risk measures are validated using samples composed of those for whom the measure is intended (Webster, Douglas, Eaves and Hart, 1997). Further research on static and dynamic risk factors has been said to be key in the future prediction and management of violence in forensic and civil inpatients (Belfrage &

Douglas, 2002). This is based on the success of applying such techniques to the correctional field, in which validation of risk assessment measures is more advanced (Belfrage & Douglas, 2002).

There is little information regarding differences in risk assessment when applied to forensic versus civil populations. Civil inpatients generally have shorter observation periods than forensic inpatients which influences the type of research that can be performed on such samples (Steinert, 2002). Preliminary research suggests that psychopathological states are generally more acute in the case of civil inpatients (Steinert, 2002), which suggests they are more likely to experience rapid changes in risk state than forensic inpatients (Belfrage & Douglas, 2002). Further research is required to guide the selection of the most appropriate risk assessment measure for each type of inpatient.

Risk Assessment Prediction Periods

It has been suggested that 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). However, what constitutes "long-term" and "short-term" has been defined inconsistently within the literature, with short-term ranging from 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 Tables 1 through 4), few have examined predictive validity at 1 month or less (Chu et al., 2013). This holds implications for the generalizability of the research findings, since the average length of psychiatric inpatient admissions in Australia is less than a month (Australian Institute of Health and Welfare, 2011). There is little data comparing static and dynamic risk measures over what seems a reasonable operationalization of the “short-term” duration of a typical hospital admission (Chu et al., 2013). Filling this gap in the research is an important step in the successful routine practical application of risk assessment research findings.

While such research is in its infancy, several studies have tested the predictive accuracy of risk assessment measures in forensic psychiatric inpatient settings with follow-up periods of up to 6 months. In the following review of these studies, outcomes will be discussed in terms of physical violence against persons (*violence*), verbal threat or property damage (*non-violence*), or one or both of these types of aggression (*any aggression*).

Short-Term Risk Assessment

Static measures.

Until recently, there had been no study which tested the predictive accuracy of static measures in follow-up periods of up to 1 month. However, Chu et al. (2013) examined a sample of 66 high-security forensic inpatients to compare the predictive validity of five risk assessment measures at 1-, 3-, and 6-month follow-ups. This included three mostly static measures or subscales. The Historical, Clinical, Risk Management – 20 Factors (HCR-20; Webster et al., 1997) is a 20-item risk assessment tool which for clinical purposes is primarily designed to be adapted to the SPJ method and consists of both static and dynamic factors. Items are divided into ten Historical (H) items comprising static factors, plus five Clinical (C) and five Risk Management (R) items both comprising dynamic factors. It is one of the most widely-used tools to assess the risk of violence in psychiatric populations (Douglas, Hart, Webster, & Belfrage, 2013). The Psychopathy Check List – Revised (PCL-R; Hare, 2003) comprises 20 mostly static risk items. Despite being designed as a diagnostic tool, the PCL-R is frequently used as a predictor of violence, either in isolation or more commonly as a part of an actuarial or SPJ tool (e.g. as an item in the HCR-20; Heilbrun, Yasahura, & Shah, 2010). The Violence Risk Appraisal Guide (Quinsey et al., 1998) is a prominent actuarial risk assessment tool which uses only static factors. The results of the Chu et al. (2013) study for the H-scale of the HCR-20, the PCL-R and the VRAG can be seen in Table 1.

Table 1.

Static Risk Assessment Over the Short Term

			AUC		
Follow-up	Study	Measure	Violent	Non-violent	Any
1 month	Chu et al. (2013)	HCR-20 H-scale	.67	.60[v]	.62
	Chu et al. (2013)	PCL-R Total	.72	.61[v]	.67
	Chu et al. (2013)	VRAG	.61	.50[v]	.53

Note. AUC = area under the curve. "Violent," "Nonviolent," and "Any" refers to physical violence, verbal threat/property damage, and any inpatient aggression, respectively. Superscript v denotes verbal threat.

As shown in Table 1, all three measures displayed mostly poor to modest predictive ability across aggression types, with the highest result being an acceptable AUC of .72 for the PCL-R predicting violence. These results suggest that static measures are inadequate at predicting inpatient aggression over the short-term (Chu et al., 2013). However, this was a single study so further research is required to support this contention.

Dynamic measures.

The AUCs from studies testing dynamic measures over follow-up periods of up to 1 month can be seen in Table 2. It is evident that there have been several more studies examining the predictive ability of dynamic risk measures rather than static risk measures over the short-term. These studies have tended to either predict aggression over the very short-term (up to 24 hours) or at 1-month follow-up.

Table 2.

Dynamic Risk Assessment Over the Short Term

			AUC		
Follow-up	Study	Measure	Violent	Non-violent	Any
Same shift	Almvik et al. (2007)	BVC*	.94		
12 hours	Abderhalden et al. (2004)	BVC*	.88		
24 hours	Almvik et al. (2007)	BVC*	.69		
	Almvik et al. (2000)	BVC*	.82		
	Ogloff and Daffern (2006)	BVC*	.83		
	Woods and Almvik (2002)	BVC*	.82		
	Barry-Walsh et al. (2009)	DASA:IV*	.80[st], .65[pt]	.73[v], .67[p]	.69
	Daffern and Howells (2007)	DASA:IV*	.65		
	Ogloff and Daffern (2006)	DASA:IV*	.82		
	Daffern and Howells (2007)	HCR-20 C-scale*	.63		
	Ogloff and Daffern (2006)	HCR-20 C-scale*	.73		
	Daffern et al. (2005)	LSI:R:SV	.60		
1 month	Chu et al. (2013)	HCR-20 Total	.78	.68[v]	.72
	Chu et al. (2013)	HCR-20 C-scale*	.73	.73[v]	.73
	Chu et al. (2013)	HCR-20 R-scale*	.75	.73[v]	.73
	Chu et al. (2013)	LSI:R:SV	.66	.53[v]	.57
	Chu et al. (2013)	START*	.75	.79[v]	.74
~1 month	Braithwaite et al. (2010)	START*			.65-.66

Note. AUC = area under the curve. * denotes a purely dynamic measure. ~ denotes an approximation. “Violent,” “Nonviolent,” and “Any” refers to physical violence, verbal threat/property damage, and any inpatient aggression, respectively. Superscripts v, p denote verbal threat, and property damage, respectively. Superscripts st, pt denote violence toward staff and patients respectively. Table adapted from Chu et al. (2013).

As seen in Table 2, the Brøset Violent Checklist (BVC; Woods & Almvik, 2002) displayed excellent predictive ability for violence over a 12- to 24-hour period (Abderhalden et al., 2004; Almvik, Woods, & Rasmussen, 2000; Ogloff & Daffern, 2006; Woods & Almvik, 2002). It also displayed outstanding predictive ability for violence when follow-up was within the same shift it was administered (Almvik, Woods, & Rasmussen, 2007); however, the BVC only displayed modest predictive ability over a 24-hour period in this study. Another purely dynamic actuarial measure, the Dynamic Appraisal of Situational Aggression IV (DASA-IV; Ogloff & Daffern, 2006), displayed modest to excellent predictive ability for violence over 24 hours (Barry-Walsh, Daffern, Douglas, & Ogloff, 2006; Daffern & Howells, 2007; Ogloff & Daffern, 2006). In addition, the DASA-IV displayed modest to acceptable predictive ability for non-violence and any aggression over the same period (Barry-Walsh, Daffern, Douglas, & Ogloff, 2006). While the BVC and the DASA-IV have the advantages of being simple to use, time-efficient, and appear to be strong predictors of aggression over the very short-term, they were designed to predict violence over periods of 24 hours only (Ogloff & Daffern, 2006).. The Level of Service Inventory – Revised: Short Version (LSI-R:SV; Andrews & Bonta, 1998), which was designed for use over longer periods and comprises mainly dynamic but some static factors, only displayed modest predictive ability for violence when tested over 24 hours (Daffern, Ogloff, Ferguson, & Thompson, 2005).

The majority of data regarding the ability of dynamic risk assessment measures to predict aggression at 1-month follow-up comes from the study by Chu et al. (2013). In this study the LSI-R:SV displayed poor to modest predictive ability across aggression conditions over 1 month, consistent with Daffern et al. (2005). Chu et al. (2013) also tested the HCR-20 Total, C-scale, and R-scales, which displayed similar modest to acceptable predictive ability over all aggression conditions at 1-month follow-up. The authors noted the fact that the dynamic C- and R-scales of the HCR-20, and not the static H-scale, produced strong results over this time period supports the notion that dynamic factors are better predictors over the short-term. Chu et al. (2013) also tested the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009), which is a newer purely dynamic measure shown to have promising predictive validity with forensic populations (Nicholls, Brink, Desmarais, Webster, & Martin, 2006). The START displayed acceptable predictive ability at 1-month follow-up across aggression types. Chu et al. (2013) concluded that the START was a strong predictor of aggression over the short- to medium-term, and argued that the dynamic nature of its items played a significant role in its higher performance than measures with more static items. However, in the only other study testing the START in an inpatient facility at 1-month follow-up, it demonstrated only modest predictive ability for any aggression (Braithwaite, Charette, Crocker, and Reyes, 2010).

The literature on predicting aggression in psychiatric inpatients over the short-term is very limited. It appears that risk measures which utilise dynamic factors have improved clinical utility over those which include mostly static factors, especially over the very short-term. Further research is required to substantiate this claim.

Medium-Term Risk Assessment

Static measures.

There have been many more studies testing the predictive ability of risk measures over the medium-term. The AUCs produced by such studies examining static measures can be seen in Table 3.

Table 3.

Static Risk Assessment Over the Medium Term

Follow-up	Study	Measure	AUC Violent	AUC Non-violent	AUC Any
3 months	Chu et al. (2013)	HCR-20 H-scale	.64	.60^{v}	.67
	Gray et al. (2003)	HCR-20 H-scale	.77	.73^{v}, .82^{p}	
	Chu et al. (2013)	PCL-R Total	.67	.66^{v}	.72
	Gray et al. (2003)	PCL-R Total	.70	.60^{v}, .76^{p}	
	Chu et al. (2013)	VRAG	.53	.56^{v}	.58
~3.5 months	Nicholls et al. (2004)	HCR-20 H-scale	.56^{m}, .57^{f}		.58^{m}, .69^{f}
	Nicholls et al. (2004)	PCL-R:SV Total	.59^{m}, .63^{f}		.60^{m}, .72^{f}
~4 months	Doyle et al. (2002)	HCR-20 H-scale	.70	.66^{v}	
	Dolan and Davies (2006)	PCL-R:SV Total	.65		.65
	Doyle et al. (2002)	PCL-R:SV Total	.76		
	Doyle et al. (2002)	VRAG	.71	.64	
Up to 6 months	McDermott et al. (2008)	HCR-20 H-scale	.60^{ia}, .84^{pa}		
	McDermott et al. (2008)	PCL-R Total	.68^{ia}, .84^{pa}		
6 months	Chu et al. (2013)	HCR-20 H-scale	.59	.56^{v}	.55
	Grevatt et al. (2004)	HCR-20 H-scale	.54	.28v, .32p	.40
	Chu et al. (2013)	PCL-R Total	.65	.63^{v}	.61
	Walters and Heilbrun (2010)	PCL-R Total			.57-.63
	Vitacco et al. (2009)	PCL-R:SV Total	.54		
	Chu et al. (2013)	VRAG	.57	.45^{v}	.54
	Snowden et al. (2009)	VRAG	.54		
~6 months	Fujii et al. (2005)	HCR-20 H-scale			.58 - .73

Note. AUC = area under the curve. * denotes a purely dynamic measure. ~ denotes an approximation. “Violent,” “Nonviolent,” and “Any” refers to physical violence, verbal threat/property damage, and any inpatient aggression, respectively. Superscripts v, p denote verbal threat, and property damage, respectively. Superscripts ia, pa denote impulsive and predatory aggression, respectively. Superscripts st, pt denote violence toward staff and patients respectively. Superscripts m, f denote male and female, respectively. Table adapted from Chu et al. (2013).

As shown in Table 3, such studies typically span follow-up periods of between 3 and 6 months. Beginning with studies using a 3-month follow-up, Chu et al. (2013) found the H-scale of the HCR-20 and the PCL-R displayed modest to acceptable predictive ability across aggression types. The VRAG also produced poor predictive ability across aggression types in this study, and it was concluded that all three measures were inadequate over the 3-month follow-up. Gray et al. (2003) compared four violence risk assessment tools in a sample of 34 institutionalised mentally disordered patients with a follow-up period of 3 months. The H-scale and PCL-R displayed similar modest to acceptable predictive ability; however, a clinical measure of mental state was equally effective in predicting violence. It should be noted that the small sample size in this study limited the strength of these findings.

Several studies have tested static measures over 3- to 4-month follow-ups. Doyle, Dolan and McGovern (2002) compared the H-scale, the short version of the PCL-R (PCL-R:SV) and the VRAG over a 12-week period using a sample of 87 mentally disordered inpatients in the United Kingdom. All three measures produced similar modest to acceptable predictive ability across aggression types, demonstrating that assessment tools developed within North America performed similarly in the United Kingdom. Dolan and Davies (2006) tested the PCL-R:SV over the 12-weeks post-admission in 134 male inpatients with a diagnosis of schizophrenia, finding it displayed modest predictive ability for violence and any aggression. Nicholls, Ogloff, and Douglas (2004) tested the predictive validity of risk assessment tools over a period of 108 days in a large sample of 268 involuntarily hospitalised male and female psychiatric patients, whether remaining in hospital or in follow-up in the community. The H-scale and the PCL-R:SV displayed mostly poor to modest predictive ability across conditions, with the highest AUC being for the female sample when predicting any aggression. The authors offered several reasons for the generally poor results, including small samples of subgroups tested, methodological flaws preventing accurate follow-up, and reliance on staff reports and records. In sum it appears studies testing static measures in inpatient facilities over 3 to 4 months have found poor to acceptable AUCs.

Several studies have also tested the predictive accuracy of risk measures over 6 months. Chu et al. (2013) tested the H-scale of the HCR-20, the PCL-R, and the VRAG and found all measures displayed poor to modest predictive ability for all aggression types at 6-month follow-up. It was concluded that none of the static measures were adequate predictors at 6-month follow-up despite their hypothesis that these measures would outperform dynamic measures over this period.

McDermott, Quanbeck, Busse, Yastro, and Scott (2008) tested the predictive accuracy of the PCL-R and the HCR-20 in predicting violence in forensic inpatients with a follow-up period of up to 6 months. Violence was separated into three subtypes: impulsive, predatory, and psychotic. The PCL-R and H-scales were excellent predictors of predatory violence, but modest predictors of impulsive violence. Incidents of impulsive violence were most frequent. The authors concluded that despite the strengths of the PCL-R and H-scale in identifying predatory violence, these might not be the most effective measures to use in this type of setting.

Grevatt, Thomas-Peter and Hughes (2004) tested the H and C scales of the HCR-20 to predict violence in the 6 months following a retrospective file review of 44 male forensic inpatients. The static H-scale displayed poor predictive ability of violence; however, the dynamic C-scale was also a poor predictor. Despite this, it appeared the static H-scale was better at predicting isolated incidents of violence than the C-scale. The small sample size and retrospective approach may have contributed to the poor outcomes presented in this study.

Walters and Heilbrun (2010) used the PCL-R to test two samples of offenders who were either forensic inpatients (n = 216) or inmates referred for psychiatric review (n = 230). The PCL-R total score demonstrated poor to modest predictive validity of any aggression, although Facet 4 of the PCL-R, which comprises items on anti-sociality, was found to be a better predictor. The low base rates of violence and reliance on often under-reported incident reports limited these findings.

Vitacco et al. (2009) tested the predictive validity of the PCL-R:SV in a sample of forensic inpatients over a 6-month follow-up period. The measure displayed poor predictive ability for violence. Similarly, Snowden, Gray, Taylor, & Fitzgerald, (2009) tested the predictive validity of the VRAG in a sample of forensic inpatients residing in a medium-secure unit in the United Kingdom over 6 months. The VRAG displayed poor predictive ability of violence. Low base rates of violence and a small sample sizes may have limited the power and generalizability of these studies.

Fujii, Tokioka, Lichton and Hishinuma (2005) tested the predictive validity of the HCR-20 and subscales on a sample of 169 psychiatric inpatients in a Hawaiian institution with a follow-up period of 193 days. They compared the predictive ability of the measure and subscales with three different ethnic groups within the sample. The H-scale displayed poor to acceptable predictive ability for any aggression depending on the ethnic group of the inpatient. This study highlights the importance of using risk measures validated for use with the intended sample.

In sum, it appears that static measures range from being poor to acceptable predictors of violence over medium-term follow-up periods. Only one study investigated static measures over the short-term, so it is difficult to compare static measures over the short and medium periods. Further research is required to determine whether static measures behave differently over the short- and medium-term.

Dynamic measures.

As was the case with static measures, there are more studies testing dynamic measures over medium-term follow-up periods than short-term, with most studies either using 3- or 6-month follow-ups. The results of such studies can be seen in Table 4.

Table 4.

Dynamic Risk Assessment Over the Medium Term

Follow-up	Study	Measure	AUC		
			Violent	Non-violent	Any
3 months	Chu et al. (2013)	HCR-20 Total	.75	.69^{v}	.78
	Gray et al. (2003)	HCR-20 Total	.81	.79^{v}, .83^{p}	
	Chu et al. (2013)	HCR-20 C-scale*	.75	.77^{v}	.78
	Gray et al. (2003)	HCR-20 C-scale*	.79	.74^{v}, .77^{p}	
	Chu et al. (2013)	HCR-20 R-scale*	.75	.74^{v}	.76
	Chu et al. (2013)	LSI:R:SV	.54	.55^{v}	.60
	Chu et al. (2013)	START*	.79	.82^{v}	.83
~3 months	Nonstad et al. (2010)	START*	.77		
~3.5 months	Nicholls et al. (2004)	HCR-20 H+C scales	.56^{m}, .62^{f}		.59^{m}, .74^{f}
	Nicholls et al. (2004)	HCR-20 C-scale*	.55^{m}, .62^{f}		.58^{m}, .70^{f}
	Nicholls et al. (2004)	VSC*	.53		.58^{m}, .59^{f}
Up to 6 months	McDermott et al. (2008)	HCR-20 Total	.69^{ia}, .89^{pa}		
	McDermott et al. (2008)	HCR-20 C-scale*	.71^{ia}, .86^{pa}		
	McDermott et al. (2008)	HCR-20 R-scale*	.70^{ia}, .69^{pa}		
6 months	Snowden et al. (2009)	COVR	.73		
	Chu et al. (2013)	HCR-20 Total	.62	.62^{v}	.59
	Grevatt et al. (2004)	HCR-20 H+C scales	.56	.45^{v}, .41^{p}	.48
	Chu et al. (2013)	HCR-20 C-scale*	.61	.72^{v}	.60
	Grevatt et al. (2004)	HCR-20 C-scale*	.60	.81^{v}, .65^{p}	.72
	Chu et al. (2013)	HCR-20 R-scale*	.67	.62^{v}	.62
	Chu et al. (2013)	LSI:R:SV	.48	.44^{v}	.43
	Chu et al. (2013)	START*	.74	.79^{v}	.74
~6 months	Fujii et al. (2005)	HCR-20 Total			.58-.73
	Fujii et al. (2005)	HCR-20 C-scale*			.58-.74
	Fujii et al. (2005)	HCR-20 R-scale*			.55-.73

Note. AUC = area under the curve. * denotes a purely dynamic measure. ~ denotes an approximation. “Violent,” “Nonviolent,” and “Any” refers to physical violence, verbal threat/property damage, and any inpatient aggression, respectively. Superscripts v, p denote verbal threat, and property damage, respectively. Superscripts ia, pa denote impulsive and predatory aggression, respectively. Superscripts st, pt denote violence toward staff and patients respectively. Superscripts m, f denote male and female, respectively. Table adapted from Chu et al. (2013).

Beginning with studies using 3-month follow-ups, Chu et al. (2013) tested the predictive validity of dynamic measures the HCR-20 and subscales, the LSI-R:SV and the START over this period. The HCR-20 Total score, which contains an equal number of dynamic and static factors, displayed modest to acceptable predictive ability across aggression conditions. On the subscale level, both the C and R scales displayed acceptable predictive ability, and both these dynamic scales were superior to the static H-scale over this period. This finding was consistent with the previous work of Gray et al. (2003), strengthening the claim that the dynamic scales of the HCR-20 are superior to the static scale at 3-month follow-up. Chu et al. (2013) also found that the purely dynamic START measure displayed acceptable to excellent predictive ability over 3-month follow-up for all aggression types. This was consistent with Nonstad et al. (2010), who tested the predictive validity of the START in 47 high-security inpatients in a Norwegian mental health facility with a 90-day follow-up period. The START displayed acceptable predictive ability, and the authors reasonably concluded that the START showed promise in violence prediction over this follow-up period. The LSI-R:SV did not perform as well as the other measures in the study by Chu et al. (2013) at 3-month follow-up, displayed poor predictive ability across aggression types.

As discussed previously, Nicholls, Ogloff, and Douglas (2004) tested the predictive validity of dynamic risk measures over follow-up period of up to 108 days in a large sample of 268 involuntarily hospitalised male and female psychiatric patients. The combination of the static H-scale and dynamic C-scale, as well as the C-scale in isolation, displayed poor to modest predictive ability for males and females committing violence and males committing any aggression, but acceptable predictive ability for females committing any aggression. A third, purely dynamic measure, the Violence Screening Checklist (McNeil & Binder, 1994), displayed poor predictive ability across aggression types.

In studies testing dynamic measures over 6-months, Snowden et al. (2009) tested the predictive validity of the Classification Of Violence Risk (COVR; Monohan et al., 2006) in a sample of forensic inpatients residing in a medium-secure unit in the United Kingdom. The COVR was an acceptable predictor of violence over this time frame, outperforming the static VRAG.

Chu et al. (2013) found the HCR-20 Total, C-scale and R-scale displayed mostly modest predictive ability, which was worse than the mostly acceptable predictive validity displayed by the same measures over 3 months. Similarly, the START displayed mostly acceptable predictive ability at 6 months compared to mostly excellent predictive ability at 3 months. The LSI-R:SV displayed equally poor predictive ability at both time periods indicating it is an inadequate predictor over the medium-term. Compared to results at the short-term, this trend suggests the dynamic measures become weaker as time from prediction increases. Grevatt et al. (2004) found the H plus C scales of the HCR-20 displayed poor predictive ability across aggression types, but the C-scale in isolation displayed more variable predictive ability with results ranging from modest to excellent. McDermott et al. (2008) found that the HCR-20 Total, C-scale and R-scale all displayed excellent predictive ability of predatory aggression, and modest to acceptable predictive ability of impulsive aggression. The authors concluded that the HCR-20 was superior to the static PCL-R for use in their forensic inpatient setting over 6-month follow-up. Finally, Fujii et al. (2005) compared three different ethnic groups in a psychiatric inpatient facility in Hawaii. The HCR-20 Total, C-scale and R-scale displayed poor to acceptable predictive ability depending on ethnic group. This was similar to results produced by the static H-scale, and again lent support for the validation of measures for use in the intended sample.

In sum, dynamic measures appeared to produce similar predictive ability within the span of the medium-term period, with suggestions these measures are weaker at 6-month follow-ups than at 1 month, and possibly weaker than at 3 months. Furthermore, there is little difference between static and dynamic measures in the medium-term. It should be highlighted that the studies reviewed here utilised samples from forensic psychiatric inpatient settings. However, these samples differed on many dimensions including the country and health system within which the facilities operated. It is unknown how much these and other dimensions account for the variability seen in the data between studies. It is therefore important to continue to test all static and dynamic risk measures in a wide range of psychiatric contexts to determine which measures are best within which context.

Conclusion

Violence risk assessment has evolved from an unstructured and arguably unethical one-in-three clinical decision into an informed, structured and much more accurate clinical decision based on actuarial processes and comparison samples (Ogloff & Davis, 2005). The research has been applied primarily within the correctional field, and there is a growing need for similar research in psychiatric settings (Ogloff & Davis, 2005; Owen et al., 1998). Violence in such settings comes at great cost to the individual, other inpatients, treating staff, and the facilities and bodies responsible for holding the inpatient (Gadon, Johnstone & Cooke, 2006; Owen et al., 1998). The comparison of static and dynamic factors is a promising area in psychiatric inpatient settings. While research suggests that dynamic measures may be better predictors of risk over the short-term, and that static measures may be better predictors of risk over the longer-term (Douglas & Skeem, 2005), the base of studies testing such claims is thus far limited. In addition, the majority of the research to date occurs in a North American context, and so examining the applicability of such research to an Australian context is warranted. Furthermore, while it is important to continue to validate risk measures in the settings they are going to be used (Webster et al., 1997), little is known about violence risk prediction in forensic versus civil samples. Future studies should continue to compare static and dynamic risk measures over various time frames and in a range of settings to assist mental health clinicians and bodies to make decisions about which measures are best within which context.

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[1], Derek G. Gilligan[1], Sean A. Halpin[1], Megan E. Valentine[2]

[1] School of Psychology, Faculty of Science and Information Technology, University of Newcastle, Australia

[2]School of Mathematical and Physical Sciences, Faculty of Science and Information Technology, University of Newcastle, Australia

Corresponding author:

Derek Gilligan

Forensic Mental Health Services, Hunter New England Mental Health.
Morisset Hospital,
P.O Box 833
Newcastle NSW 2300

Derek.Gilligan@hnehealth.nsw.gov.au

Abstract

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 are better predictors of aggression over the short-term was supported. 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.

Keywords: risk, violence, static, dynamic, forensic, civil, short-term

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 competently 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 more accurately predict 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 make informed decisions about the best available methods of violence risk assessment in a 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 forensic inpatients, which influences the type of research that can be performed on such samples (Steinert, 2002). Preliminary research suggests that civil inpatients typically present with more acute psychopathological states (Steinert, 2002), which suggests they may experience more rapid changes in risk state than 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 worthwhile to further investigate the utility of dynamic risk measures, for both the prediction and 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 1 month or less. This holds 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 the majority of the risk assessment literature is performed 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 one of 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, Grann, & Fazel, 2013). 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 only be examined 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 due to broader methodology (actuarial vs SPJ) from differences due to particular tools (Heilbrun, Yasahura, & 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 a modest predictors, .7 to .8 are classified as acceptable, .8 to .9 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 Health, 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. Despite being 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) finding that dynamic measures were better predictors of aggression than 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 which are routinely used within the public health service (MHOAT-Risk and HoNOS-secure) and a third which 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 in follow-up periods of 1, 3, or 6 months from 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 +-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 was included. Data from civil inpatients was 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 five Clinical (C) items comprising dynamic factors that reflect current mental and clinical status, and five Risk Management (R) items also comprising dynamic factors that reflect future situational risk Each item is coded as either a 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 (Lestico, 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, Thomas, Ogloff, & Daffern, 2011; Chu et al., 2013; 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 either 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 Appendix B). 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 either 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 state-wide 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 recruitment site. The measure requires the clinician to rate the patient as *Yes*, *No,* or *Unknown* on *Background* and *Current* items within four categories of risk including *General Risk Factors*, *Suicide*, *Violence/Aggression*, and *Other Vulnerabilities*. Only the MHOAT-Risk categories Violence/Aggression (six background factors, eight current factors) and General Risk Factors (six background factors, four 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” (Chipps, Raphael & Coombes, 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, 2013). The HoNOS for users of secure and forensic services (HoNOS-secure) was developed as some items in the original HoNOS proved inapplicable 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; Sugarman & Walker, 2007). However, there is considerable evidence to suggest that the original HoNOS has good predictive validity in generalist mental health settings (Pirkis et al. 2005; Shrinkfield & Ogloff, 2014; Webster et al. 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 seven additional *secure* items which involve rating risk in the near future, defined as within weeks or months (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, Sugarman, & Walker, 2007), its predictive validity for aggression has not been examined over a short-to-medium time frame.

Measures containing a clear majority of static items were classified as *static measures* in this study. Measures which contained at least an equal number of dynamic and static items were classified as *dynamic measures* (see Table 5).

Table 5.

Static vs. Dynamic Measures

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

Note. * denotes purely 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), v*erbal 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 they had committed at least one incident of interpersonal violence, and a score of 0 indicating 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 in a given period, a score of 1 was allocated to any aggression for that period. The short-term follow-up period spanned 0 to 1 month, and the medium-term follow-up periods spanned 0 to 3 and 0 to 6 months respectively, based on Chu et al. (2013).

Procedure

Data was retrieved from clinical records for the risk assessment measures. The risk measures had been scored by individual case mangers (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 was 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 was collected and de-identified prior to analysis in accordance with the institutional ethics approval using a Microsoft Excel spreadsheet.

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

Table 6.

Inpatient Demographics

Demographic	Total n=74	%	Forensic n=37	%	Civil n=37	%
Security						
Medium	30	*40*	17	*46*	13	*35*
Low	44	*60*	20	*54*	24	*65*
Primary mental health diagnosis						
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*
Secondary mental health diagnosis						
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*

The number of inpatients who were aggressive (interpersonal violence, verbal threat, any aggression) at least once within each period is shown in Table 7. 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.

Table 7.

Number of Total, Forensic and Civil Inpatients Who Were Aggressive at Least Once Within Each of the Follow-Up Periods

	Up to 1 Month			Up to 3 Months			Up to 6 Months		
Aggressive Behaviour	n (=74)	f (=37)	c (=37)	n (=74)	f (=37)	c (=37)	n (=74)	f (=37)	c (=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.
f , c denote forensic and civil inpatients respectively
* denotes a significant difference based on Pearson's Chi Square

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 8. 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 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.

Table 8.

Predictive Accuracy of Static and Dynamic Risk Measures for Inpatient Aggression in the Total Sample

	Up to 1 month		Up to 3 months		Up to 6 months	
Measure	AUC (SE)	95% CI	AUC (SE)	95% CI	AUC (SE)	95% CI
			Interpersonal Violence			
(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
			Verbal Threat			
(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
			Any Aggression			
(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. * denotes significant at .05
boldface font denotes significance after FDR corrections
(s), (d) denotes static and dynamic measure respectively

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 9.

Table 9.

Predictive Accuracy of Static and Dynamic Risk Measures for Inpatient Aggression in the Forensic Group

	Up to 1 month		Up to 3 months		Up to 6 months	
Measure	AUC (SE)	95% CI	AUC (SE)	95% CI	AUC (SE)	95% CI
			Interpersonal Violence			
(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
			Verbal Threat			
(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
			Any Aggression			
(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. * denotes significant at .05
boldface font denotes significance after FDR corrections
(s), (d) denotes static and dynamic measure respectively

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 9).

Predictive accuracy for verbal threat

At 1-month follow-up, dynamic measures were excellent to outstanding predictors of 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 six of the eight 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-.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 10.

Table 10.

Predictive Accuracy of Static and Dynamic Risk Measures for Inpatient Aggression in the Civil Group

	Up to 1 month		Up to 3 months		Up to 6 months	
Measure	AUC (SE)	95% CI	AUC (SE)	95% CI	AUC (SE)	95% CI
			Interpersonal Violence			
(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
			Verbal Threat			
(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
			Any Aggression			
(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. * denotes significant at .05
boldface font denotes significance after FDR corrections
(s), (d) denotes static and dynamic measure respectively

Predictive accuracy for interpersonal violence

As shown in Table 10, at 1-month follow-up, seven of the eight 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, five of the eight 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, five of the eight 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 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 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 to that of the high-secure setting in 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 et al., 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; p. 5-6). 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 as 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 the risk measures were completed by experienced mental health clinicians who had extensive knowledge and training in the 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 Ruiter, 2004; de Vogel & de Ruiter, 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 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 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 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 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 five-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 (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 the risk measures were originally scored by various clinicians performing both individual and team-based ratings. 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 readministered 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 the AUC was not significantly different from chance. However, this was not always the case in the present results (see Table 10 results for R-scale at predicting interpersonal violence at 1 month for an example). Such inconsistencies warrant further investigation.

Future Research

The current study focussed 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 1 to 2 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.

REFERENCES

Abderhalden, C., Needham, I., Miserez, B., Almvik, R., Dassen, T., Haug, H. J., & Fischer, J. E. (2004). Predicting inpatient violence in acute psychiatric wards using the Brøset-Violent-Checklist: A multicentre prospective cohort study. *Journal of Psychiatric and Mental Health Nursing, 11*, 422-427.

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341-382.

Almvik, R., Woods, P., & Rasmussen, K. (2000). The Brøset Violent Checklist: Sensitivity, specificity, and interrater reliability. *Journal of Interpersonal Violence, 15*, 1284-1296.

Almvik, R., Woods, P., & Rasmussen, K. (2007). Assessing risk for imminent violence in the elderly: The Brøset Violent Checklist. *International Journal of Geriatric Psychiatry, 22*, 862-867.

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

Barry-Walsh, J., Daffern, M., Duncan, S., & Ogloff, J. R. P. (2009). A test of the dynamic appraisal of situational aggression in a NZ forensic psychiatric service. *Australasian Psychiatry, 17*, 493-496.

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.

Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders. *Psychological Bulletin, 123*, 123-142.

Braithwaite, E., Charette, Y., Crocker, A. G., & 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.

Catchpole, R. E. H., & Gretton, H. M. (2003). The predictive validity of risk assessment with violent young offenders: A one-year examination of criminal outcome. *Criminal Justice and Behavior, 30*, 688-708.

Chipps, J., Raphael, B., and 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*(11-12), 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.

Daffern, M., & Howells, K. (2002). Psychiatric inpatient aggression: A review of structural and functional assessment approaches. *Aggression and Violent Behavior, 7,* 477-497.

Daffern, M., & Howells, K. (2007). The prediction of imminent aggression and self-harm in personality disordered patients of a high security hospital using the HCR-20 clinical scale and the Dynamic Appraisal of Situational Aggression. *International Journal of Forensic Mental Health, 6*, 137-143.

Daffern, M., Howells, K., & Ogloff, J. (2007). What's the point? Towards a methodology for assessing the function of psychiatric inpatient aggression. *Behaviour Research and Therapy, 45,* 101-111.

Daffern, M., Ogloff, J. R. P., Ferguson, M., & Thomson, L. (2005). Assessing risk for aggression in a forensic psychiatric hospital using the Level of Service Inventory-Revised: Screening Version. *International Journal of Forensic Mental Health, 4,* 201-206.

Dahle, K. P. (2006). Strengths and limitations of actuarial prediction of criminal re-offence in a German prison sample: A comparative study of the LSI-R, HCR-20, and PCL-R. *International Journal of Law and Psychiatry, 29*, 431-442.

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.

de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health, 3*, 149-165.

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.

Dolan, M., & Davies, G. (2006). Psychopathy and institutional outcome in patients with schizophrenia in forensic settings in the UK. *Schizophrenia Research*, *81*(2-3), 277-281.

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.

Douglas, K. S., Yeomans, M. & Boer, D. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32*, 479-510.

Doyle, M., Dolan, M., & McGovern, J. (2002). The validity of North American risk assessment tools in predicting in-patient violent behaviour in England. *Legal and Criminological Psychology, 7*, 141-154.

Fujii, D. E. M., Tokioka, A. B., Lichton, A. I., & Hishinuma, E. (2005). Ethnic differences in prediction of violence risk with the HCR-20 among psychiatric inpatients. *Psychiatric Services*, *56*, 711-716.

Gadon, L., Johnstone, L., & Cooke, D. (2006). Situational variables and institutional violence: A systematic review of the literature. *Clinical Psychology Review, 26*, 515-534.

Grann, M., & Langstrom, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh? *Criminal Justice and Behavior*, *34*(1) 22-36.

Gray, N. S., Hill, C., McGleish, A., Timmons, D., MacCulloch, M. J., & Snowden, R. J. (2003). Prediction of violence and self-harm in mentally disordered offenders: A prospective study of efficacy of HCR-20, PCL-R, and psychiatric symptomatology. *Journal of Consulting and Clinical Psychology, 71*, 443-451.

Grevatt, M., Thomas-Peter, B., & Hughes, G. (2004). Violence, mental disorder and risk assessment: Can structured clinical assessments predict the short-term risk of inpatient violence. *Journal of Forensic Psychiatry & Psychology, 15*, 278-292.

Grisso, T., & Appelbaum, P. S. (1992). Is it unethical to offer predictions of future violence? *Law and Human Behavior, 16,* 621-633.

Hajian-Tilaki, K. (2013). Reciever operating characteristic (ROC) curve anaylsis for medical diagnostic test evaluation. *Caspian Journal of International Medicine, 4*(2), 627-635.

Hare, R. D. (2003). *The Hare Psychopathy Checklist–Revised manual (2nd ed.).* North Tonawanda, NY: Multi-Health Systems.

Harris, G. T. & Rice, M. E. (2007). Adjusting actuarial violence risk assessments based on aging or the passage of time. *Criminal Justice and Behaviour, 34,* 267-313.

Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the "margins of error" of group v. individual predictions of violence. *British Journal of Psychiatry, 190,* 60-65.

Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: Importance of legal decision-making context. *Law and Human Behavior, 21*, 347-359.

Heilbrun, K., Douglas, K. S., & Yasuhara, K. (2009). Controversies in violence risk assessment. In J.L. Skeem, K.S. Douglas, & S.O. Lilienfeld (Eds.), *Psychological science in the courtroom: Controversies and consensus* (pp. 333-356). New York, NY: Guilford.

Heilbrun, K., Ogloff, J. R. P., & Picarello, K. (1999). Dangerous offender statutes in the United States and Canada: implications for risk assessment. *International Journal of Law and Psychiatry, 22*, 393-415.

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.

Kroner, D. G., Mills, J. F., & Reddon, J. R. (2005). A coffee can, factor analysis, and prediction of antisocial behavior: The structure of criminal risk. *International Journal of Law and Psychiatry 28*(4), 360-374.

McDermott, B. E., Quanbeck, C. D., Busse, D., Yastro, K., & Scott, C. L. (2008). The accuracy of risk assessment instruments in the prediction of impulsive versus predatory aggression. *Behavioral Sciences & the Law, 26*, 759-777.

McNeil, D. E., & Binder, R. L. (1994). Screening for risk of inpatient violence: Validation of an actuarial tool. *Law and Human Behavior, 18,* 579-586.

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.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press. Reprinted with new Preface, 1996, by Jason Aronson, Northvale, NJ.

Monahan, J. (1981). *The clinical prediction of violent behavior*. Washington, DC: Government Printing House.

Monahan, J. (1984) The prediction of violent behaviour: Toward a second generation of theory and policy. *American Journal of Psychiatry, 141,* 10-15.

Monahan, J., Steadman, H. J., Appelbaum, P. S., Grisso, T., Mulvey, E. P., Roth, L. H., Robbins, P. C., Banks, S. & Silver, E. (2006). The classification of violence risk. *Behavioral Science and Law, 24*, 721-730.

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.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783-792.

Mossman, D. (2007). Avoiding errors about “margins of error”. *British Journal of Psychiatry, 191*, 561.

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.

Nicholls, T. L., Ogloff, J. R. P., & Douglas, K. S. (2004). Assessing risk for violence among male and female civil psychiatric patients: The HCR-20, PCL: SV, and VSC. *Behavioral Sciences & the Law, 22*, 127-158.

Nonstad, K., Nesset, M. B., Kroppan, E., Pedersen, T. W., Nøttestad, J. A., Almvik, R., & Palmstierna, T. (2010). Predictive validity and other psychometric properties of the Short-Term Assessment of Risk and Treatability (START) in a Norwegian high secure hospital. *International Journal of Forensic Mental Health, 9*, 294-299.

Ogloff, J. R. P., & Daffern, M. (2006). The dynamic appraisal of situational aggression: An instrument to assess risk for imminent aggression in psychiatric inpatients. *Behavioral Sciences & the Law, 24*, 799-813.

Ogloff, J., & Davis, M. (2005). Assessing risk for violence in the Australian context. In D. Chappell & P. Wilson (Eds.), *Issues in Australian Crime and Criminal Justice* (pp. 301-338). Chatswood: Lexis Nexis Butterworths.

Otto, R. K. (1992). Predicting dangerous behavior: A review and analysis of second generation research. *Forensic Reports, 5*, 103-133.

Owen, C., Tarantello, C., Jones, M., & Tennant, C. (1998). Repetitively violent patients in psychiatric units. *Psychiatric Services, 49* (11), 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

Shinkfield, G. and 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*(3), 499-513.

Singh, J. P., Desmarais, S. L., & Van Dorn, R. A. (2013). Measurement of predictive vaility in violence risk assessment studies: a second-order systematic review. *Behavioural Sciences and Law, 31*(1), 55-73.

Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments? *International Journal of Forensic Mental Health, 1*, 179-183.

Snowden, R. J., Gray, N., Taylor, J., & Fitzgerald, S. (2009). Assessing risk of future violence among forensic psychiatric patients with the Classification of Violence Risk (COVR). *Psychiatric Services, 60*, 1522-1626.

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.

Thompson, S. G., and Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21*, 1559-1573.

van den Brink, R. D. (2010). Routine violence risk assessment in community forensic mental healthcare. *Behavioral Sciences & The Law, 28*(3), 396-410.

Vitacco, M. J., van Rybroek, G. J., Rogstad, J. E., Yahr, L. E., Tomony, J. D., & Saewert (2009). Predicting short-term institutional aggression in forensic patients: A multi-trait method for understanding subtypes of aggression. *Law and Human Behavior, 33*, 308-319.

Walters, G. D. (2003). Predicting institutional adjustment and recidivismwith the Psychopathy Checklist factor scores: A meta-analysis. *Law and Human Behavior, 27*, 541–558.

Walters, G. D., & Heilbrun, K. (2010). Violence risk assessment and facet 4 of the Psychopathy Checklist: Predicting institutional and community aggression in two forensic samples. *Assessment, 17*, 259-268.

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.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Middleton, C. (2004). *Manual for the Short-Term Assessment of Risk and Treatability (START; Version 1.0).* Port Coquitlam, British Columbia, Canada: British Forensic Psychiatric Services Commission.

Woods, P., & Almvik, R. (2002). The Brøset Violent Checklist (BVC). *Acta Psychiatrica Scandinavica, 106*(412), 103-105.

Appendix A – Ethics Clearance

Health
Hunter New England
Local Health District

26 August 2014

Mr Brayden Finch
School of Psychology
University of Newcastle

Dear Mr Finch,

Re: To Short-to-Medium Term Predictive Validity of Risk of Violence Measures in a Medium-Low Security Forensic Inpatient Setting (14/08/20/5.05)

HNEHREC Reference No: 14/08/20/5.05
NSW HREC Reference No: LNR/14/HNE/304
SSA Reference No: LNRSSA/14/HNE/305

Thank you for submitting an application for authorisation of this project. I am pleased to inform you that authorisation has been granted for this study to take place at the following sites:

- [redacted]

The following conditions apply to this research project. These are additional to those conditions imposed by the Human Research Ethics Committee that granted ethical approval:

1. Proposed amendments to the research protocol or conduct of the research which may affect the ethical acceptability of the project, and which are submitted to the lead HREC for review, are copied to the research governance officer;
2. Proposed amendments to the research protocol or conduct of the research which may affect the ongoing site acceptability of the project, are to be submitted to the research governance officer.

Yours faithfully

Dr Nicole Gerrand
Research Governance Officer
Hunter New England Local Health District

Hunter New England Research Ethics & Governance Unit
Locked Bag 1
New Lambton NSW 2305
Telephone: (02) 49214950 Facsimile: (02) 49214818
Email: HNELHD-HREC@hnehealth.nsw.gov.au
http://www.hnehealth.nsw.gov.au/research_ethics_and_governance_unit

Appendix B – Items Included in the DRS

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

Appendix C - Submission Guidelines for Intended Journal: *Assessment*

The editor invites high quality manuscripts covering a broad range of topics and techniques in the area of psychological assessment. These may include empirical studies of assessment of personality, psychopathology, cognitive functions or behavior, articles dealing with general methodological or psychometric topics relevant to assessment, or comprehensive literature reviews in any of these areas. This journal encourages submissions evaluating a) new assessment methodologies and techniques for both researchers and practitioners, b) how assessment methods and research informs understanding of major issues in clinical psychology such as the structure, classification, and mechanisms of psychopathology, and c) multi-method assessment research and the integration of assessment methods in research and practice. Additionally, the journal encourages submissions introducing useful, novel, and non-redundant instruments or demonstrating how existing instruments have applicability in new research or applied contexts. All submissions should provide strong rationales for their efforts and articulate important implications for assessment science and/or practice

Research participants may represent both clinical and nonclinical populations.

In general, regular articles should not exceed 30 pages of text, excluding Title Page, Abstract, Tables, Figures, Footnotes and Reference list.

Authors submitting manuscripts to the journal should not simultaneously submit them to another journal, nor should manuscripts have been published elsewhere, including the World Wide Web, in substantially similar form or with substantially similar content.

This journal is a member of the Committee on Publication Ethics (COPE)

Manuscript Submission:

Manuscripts must be submitted in Microsoft Word or Rich Text Format (rtf) electronically at http://mc.manuscriptcentral.com/asmnt. Figures may be submitted using any of the formats listed below. If requesting a masked blind review, please ensure that both a manuscript file with no identifying author information and a separate title page with author details are included in your submission. Questions should be directed to the ASSESSMENT Editorial Office by email: assessment.editorial@gmail.com.

Preparation of Manuscripts:

Authors should carefully prepare their manuscripts in accordance with the following instructions.

Authors should use the Publication Manual of the American Psychological Association (6th edition, 2009) as a guide for preparing manuscripts for submission.

All manuscript pages, including reference lists and tables, must be typed double-spaced.

The first page of the paper (the title page) should contain the article title, the names and affiliations of all authors, authors' notes or acknowledgments, and the names and complete mailing addresses of the corresponding author. If requesting a masked blind review, the first page should contain only the article title and the title page should be uploaded as a separate document.

The second page should contain an abstract of no more than 150 words and five to seven keywords that will be published following the abstract.

The following sections should be prepared as indicated:

Tables. Each table should be fully titled, double-spaced on a separate page, and placed at the end of the manuscript. Tables should be numbered consecutively with Arabic numerals. Footnotes to tables should be identified with superscript lowercase letters and placed at the bottom of the table. All tables should be referred to in the text.

Figures. Electronic copies of figures can be submitted in one of the following file formats: TIFF, EPS, JPEG, or PDF. All figures should be referred to in text. Each figure should appear on a separate page at the end of the manuscript but before the tables, and all titles should appear on a single, separate page.

Endnotes. Notes should appear on a separate page before the References section. Notes should be numbered consecutively and each endnote should be referred to in text with a corresponding superscript number.

References. Text citations and references should follow the style of the Publication Manual of the American Psychological Association (6th edition, 2009).

Authors who want to refine the use of English in their manuscripts might consider utilizing the services of SPi, a non-affiliated company that offers Professional Editing Services to authors of journal articles in the areas of science, technology, medicine or the social sciences. SPi specializes in editing and correcting English-language manuscripts written by authors with a primary language other than English. Visit http://www.prof-editing.com for more information about SPi's Professional Editing Services, pricing, and turn-around times, or to obtain a free quote or submit a manuscript for language polishing.

Please be aware that SAGE has no affiliation with SPi and makes no endorsement of the company. An author's use of SPi's services in no way guarantees that his or her submission will ultimately be accepted. Any arrangement an author enters into will be exclusively between the author and SPi, and any costs incurred are the sole responsibility of the author.