

Differential predictive validity of the Historical, Clinical and Risk Management Scales (HCR–20) for inpatient aggression

Laura E. O’Shea, Marco M. Picchioni, Fiona L. Mason, Philip A. Sugarman,
Geoffrey L. Dickens

This is the accepted manuscript © 2015, Elsevier
Licensed under the Creative Commons Attribution-NonCommercial-
NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>



The published article is available from doi:
<http://dx.doi.org/10.1016/j.psychres.2014.07.080>

Abstract

The HCR-20 may be a better predictor of inpatient aggression for selected demographic and clinical groups but homogeneity of study samples has prevented definitive conclusions. The aim of this study, therefore, was to test the predictive validity of the HCR-20 as a function of gender, diagnosis, age, and ethnicity while controlling for potential covariates. A pseudo-prospective cohort study ($n=505$) was conducted in a UK secure/forensic mental health setting using routinely collected data. The HCR-20 predicted aggression better for women than men, and for people with schizophrenia and/or personality disorder than for other diagnostic groups. In women, the presence of the risk management items (R5) was important while men's aggression was best predicted solely by current clinical features from the C5 scale. R5 items were better than C5 items for predicting aggression in people with organic and developmental diagnoses. Our data provide additional information on which HCR-20 raters can formulate overall summary judgements about risk for inpatient aggression based on important demographic and clinical characteristics.

Keywords: Violence, Risk Assessment, Diagnosis, Inpatients, Mental Health, Schizophrenia, Personality Disorders

1. Introduction

Aggression by patients in institutional forensic mental health settings is common (Bowers et al., 2011) and has a profound adverse impact on the therapeutic environment and the safety and wellbeing of patients and staff (McKenzie and Curr, 2005). Accurate prediction, prevention and management are therefore crucial objectives. Individual risk assessment and risk management planning are widely considered to be best practice, and structured professional judgement has become the gold standard technique to guide this (National Institute for Mental Health in England, 2004). Of the available tools, the HCR-20 (Webster et al., 1997) is the most commonly used (Khiroya et al., 2009). It is highly rated for utility (Khiroya et al., 2009), can be scored reliably (Douglas et al., 2002-2008), and has convergent validity with other tools that predict aggression including the Violence Risk Appraisal Guide (VRAG; Doyle et al., 2002) and the Psychopathy Checklist – Revised (PCL-R; Warren et al., 2005). A meta-analytic comparison of six risk assessment guides found that the HCR-20 produced the largest mean effect size for institutional violence (Campbell et al., 2009). Despite this, little is known about variation in performance of the HCR-20 between clinical and demographic groups. Meta-analysis has revealed significantly greater effect sizes in samples containing higher proportions of women, Caucasians and people with schizophrenia, and lower effect size in samples containing higher proportions of people with a personality disorder (O’Shea et al., 2013). However, much of the literature comprises potentially biased and relatively small, homogenous samples of young, male, Caucasian patients with schizophrenia and personality disorder diagnoses. This has prevented detailed examination of between-group performance and consequently has limited the generalisation of research findings to patients that differ from existing validation samples. This is

important since clinicians must interpret and apply group-derived data to the case at hand (Buchanan, 2013). Since risk assessment guides like the HCR-20 inform decisions about continued compulsory detention and deprivations of liberty it is crucial to test whether they are accurate in all the groups to which they are applied. This study aimed to extend current knowledge by sampling from the largest sample in the literature in order to establish whether the predictive efficacy of the HCR-20 differs by gender, age, ethnicity and diagnosis while controlling for potential covariates. Hypotheses were theory-driven in that the predictive efficacy of a risk assessment tool would be expected to be best among samples most like those whom they were developed and/or validated with (Buchanan, 2013). We therefore hypothesised that HCR-20 would perform best in groups most like validation samples: better in males than females, in younger than older patients, in Caucasians than non-Caucasians, and in patients with schizophrenia and personality disorder diagnoses compared with other patients. We also aimed to examine the relevance and relative importance of individual HCR-20 items to risk prediction across different groups.

2. Method

2.1 Setting and participants

St Andrew's is a provider of specialist secure and forensic mental health inpatient care located at four sites in England. Different sites have separate managerial structures but all are governed by a single clinical risk management policy. Patient accommodation is provided in gender specific wards at levels of medium security or low security/locked rehabilitation, and there are a small number of 'open' rehabilitation beds. Some wards provide accommodation for subgroups of patients defined by age (e.g., older adults or adolescents), or by diagnostic group

(e.g. autistic spectrum disorder, borderline personality disorder, and neuropsychiatric conditions), while others provide care for more heterogeneous groups. The sampling frame of eligible participants were inpatients aged over 18 years who were resident in any ward between 1st September 2010 and 31st March 2011; who did not have an intellectual disability; who had at least one HCR-20 risk assessment completed during that time period, and remained in the service for at least three months after assessment. Patients were excluded if their HCR-20 assessment was missing data in excess of prorating guidelines (Webster et al., 1997: p. 21).

2.2 Study design

This naturalistic study employed a pseudo-prospective cohort design. Participants were routinely assessed during admission for violence risk by their clinical team. Organisational risk management policy indicates that the HCR-20 should be completed in accordance with guidelines in the HCR-20 version 2 manual (Webster et al, 1997: p.15): namely that for clinical use multiple sources of information should be considered. In the current study setting HCR-20 assessment is conducted by registered psychologists or graduate psychology assistants under their supervision and ratified by the clinical team. A pre-study audit found that 92% of HCR-20 assessments reported information sources used; of these, 68% had used patient interviews, 97% medical records, and 75% consultation with other clinical staff (available upon request). Information about subsequent aggressive and violent behaviour was extracted from entries in clinical notes.

2.3 Procedure

The study was approved by the clinical audit and service evaluation committee. Data were collated from anonymised versions of patients' records linked by allocation of a unique code number.

2.4 Measures

2.4.1 Violence risk assessment

Risk assessment was conducted with the HCR-20 version 2 (Webster et al., 1997; Webster et al., 1995); since the completion of the current study the new HCR-20: version 3 (Douglas, et al., 2013) has been introduced. The version 2 tool comprises 20 items split into three scales; the Historical (H10), the Clinical (C5) and the Risk Management (R5) (see Table 1 for list of items). Each item is rated as: "not present" (No or 0), "possibly present or present to some degree" (Possible or 1), or "definitely present" (Yes or 2) (Webster and Douglas, 2001; p. 2). The "Psychopathy" item was omitted since the majority of patients had not been assessed using the PCL-R. Psychopathy is regularly omitted from HCR-20 assessments due to the additional time and expertise required to administer the PCL-R (Webster et al., 2002) with negligible effect on the predictive efficacy of the tool (Guy et al., 2010). In clinical practice, the R5 items, which forecast how individuals will adjust to future circumstances, are rated twice, once under the premise that the individual will be institutionalized ('In') and again under the premise that the individual will be in the community in the near future ('Out'). Since we were testing inpatient aggression we only used the 'In' scenario ratings. In addition, a separate overall summary judgement (SJ) of future aggression risk (low, moderate or high) is made on

completion of the risk assessment. Omitted items were prorated in accordance with guidance in the HCR-20 manual (Webster et al., 1997: p. 21).

[INSERT TABLE 1 ABOUT HERE]

2.4.2 Demographic and clinical data

- Data comprised information about age, gender, date of admission, legal status, self-reported ethnicity, and ICD-10 psychiatric diagnoses as recorded by the patients' consultant psychiatrist. For the purpose of analysis individuals were grouped according to the major ICD-10 categories relating to mental and behavioural disorders (World Health Organisation, 1992).

Because a large number of participants had both schizophrenia and personality disorder we treated these as a distinct comorbid group. Diagnosis of a substance use disorder was entered as a separate variable for all subsequent analyses. After assignment to groups a number of patients remained with multiple diagnoses. There was no theoretically coherent way of categorising these patients into groups of sufficient size to enable analysis and they were excluded from analyses of diagnostic difference.

2.4.3 Aggressive outcomes

Aggressive and violent behaviour occurring in the three months following risk assessment was identified for each patient from their clinical record. Each entry is electronically flagged by the author at the time of entry when any of a range of risk behaviours is identified. We obtained all entries flagged for: "Aggression – Physical", "Aggression – Verbal", "Fire Setting", "Hostage Taking", "Intimidation/Bullying", "Self Harm/Suicide" and "Sexual Offending". Data were coded using the Overt Aggression Scale (OAS; Yudofsky et al., 1986).

Incidents are assigned to one of four categories (verbal aggression, physical aggression against objects, physical aggression against self [not reported on in this paper] and physical aggression against people) and then rated on a criterion-referenced severity scale from 1 (least severe outcome) to 4 (most severe outcome). Severity of verbal aggression on the OAS ranges from angry shouting (1) to clear threats of violence (4); for physical aggression against objects from slamming a door (1) to setting fires or throwing objects dangerously (4); and for physical aggression against other people from making threatening gestures (1) to attacks resulting in severe physical injury (4). Rating was conducted by three trained researchers blind to the HCR-20 risk assessments. Inter-rater reliability was tested on a sample of $n=260$ incidents. Agreement was in the excellent range; categorical agreement on aggression category $K=0.91$, and weighted kappa for severity agreement range $K= 0.81-1.0$ (mean 0.90). To reduce the number of aggression types analysed we amalgamated data into two categories: any aggression (*'any aggression'*; including verbal aggression, physical aggression towards objects and physical aggression towards people) and physical aggression towards people (*'physical aggression'*). Within both categories aggression was dichotomised as absent/present within the follow-up period.

2.5 Data analysis

Descriptive statistics were calculated in order to examine HCR-20 risk ratings and the incidence of aggression within the follow up period. Pearson's Chi squared tests were conducted to examine whether prevalence of aggressive behaviour, risk levels and sample characteristics differed between clinical and demographic groups. Differences in mean HCR-20 scores between

groups were investigated using Independent *t* tests and one-way ANOVAs¹. Scores for the HCR-20 total, subscale totals, and the SJ were assessed for their predictive validity for inpatient aggression. Predictive validity was investigated using the area under the curve (AUC) values generated from Receiver Operating Characteristic (ROC) analysis which is the preferred method since it is unaffected by base rates (Mossman, 1994). The ROC-AUC parameter ranges from 0 to 1, with 0.5 reflecting prediction that is no better than chance: 0.75 is generally considered the lower boundary of a large effect size (Dolan and Doyle, 2000). Differences in predictive efficacy between groups were examined using the *rocreg* function in Stata version 12 for Windows. *Rocreg* performs regression using ROC principles and hence calculates sensitivity and specificity based on variables of interest, whilst controlling for covariates. This allowed us to examine whether AUC values differed as a function of gender, age, ethnicity and diagnosis whilst controlling for other relevant characteristics. We also used *rocreg* analyses to calculate AUC values for the individual HCR-20 items for each of the groups whilst controlling for covariates (item-outcome analysis). Significance from the *rocreg* analyses was inferred from absence of zero from 99% confidence intervals (equivalent to $p < 0.01$). Except where stated, analyses were conducted using PASW Statistics version 18 for Windows (SPSS Inc., Version 18).

3. Results

3.1 Participants – characteristics and aggression

Of 546 eligible patients 41 were excluded because of excessive missing data from their HCR-20 assessment, leaving a final total sample of $n=505$ (response rate 92.5%). The SJ for

¹ Full results for differences between groups are presented in the supplementary material, tables 1 to 4.

violence risk was available for $n=339$ (67.1%) patients. There were 349 (69.1%) males and 156 (30.9%) females. Just under half the sample were Caucasian (46.5%, $n=235$); 18.2% ($n=92$) were non-Caucasian and 35.2% ($n=178$) declined to indicate their ethnicity. Mean age was 39.8 years. The majority of patients were compulsorily detained under the civil (40.0%, $n=202$) or forensic (57.2%, $n=289$) parts of the Mental Health Act 1983 (amended in 2007) while $n=14$ (2.8%) were admitted informally; the mean length of time between admission and assessment was 849 days ($SD=1334$). A quarter of the sample (25%, $n=126$) resided in medium secure wards; the remainder were in low secure, locked or open wards (75%, $n=379$), which were grouped together for the purpose of analysis. In total, 14.7% ($n=74$) had a recorded substance use diagnosis [F10-19]. Final diagnostic groups were: Schizophrenia [F20-29] (43.8%, $n=221$), Personality Disorder [F60-69] (14.3%, $n=72$), Schizophrenia & Personality Disorder [F20-29 & F60-69] (8.3%, $n=42$), Developmental [F80-89] (6.5%, $n=33$) and Organic [F00-09] (6.3%, $n=32$). Of the 72 people with a personality disorder diagnosis only 11 (15.3%) were not diagnosed with Emotionally Unstable Personality Disorder (EUPD). Forty nine (68.1%) had a sole diagnosis of EUPD and an additional 12 (16.7%) had EUPD with co-morbid personality disorders. We therefore grouped all personality diagnoses together for analysis rather than by cluster.

There were 2,202 incidents of any aggression, including 710 incidents of physical aggression against people. More than half of the sample (61%, $n=308$) engaged in any aggressive behaviour during the three month follow-up period and 35.8% ($n=181$) engaged in physical aggression.

3.2 HCR-20 risk ratings

HCR-20 total score ranged from 7 to 38. The mean HCR-20 total, C5 scale and R5 scale scores were significantly higher for patients who had engaged in each form of aggression compared with those who had not (see Table 2). For patients whose SJ was available nearly half (46.9%, $n=159$) were classified as moderate risk; 27.7% ($n=94$) as low risk and 25.4% ($n=86$) as high risk. Mean HCR-20 scores differed significantly across the SJ-defined risk levels, $F(2,336)=23.46, p<0.001$. A Tukey post-hoc test revealed that mean HCR-20 scores were significantly higher in the moderate ($M=28.90, SD=4.77, p<0.001$) and high risk ($M=29.52, SD=5.45, p<0.001$) groups compared with the low risk group ($M=24.66, SD=6.30$). There was no significant difference between mean HCR-20 scores for those classified as moderate and high risk ($p=0.674$).

[INSERT TABLE 2 ABOUT HERE]

3.3 Predictive validity of HCR-20

AUC values ranged from 0.474 to 0.760 (see Table 3). HCR-20 total and the SJ were significant predictors for both types of aggression. A large effect size (>0.75) was present solely for the prediction of any aggression by the C5 scale. With the exception of the SJ, AUC values were larger for the prediction of any aggression than for physical aggression. However, 99% confidence intervals for these two outcomes overlapped and thus were not significantly better for any aggression.

[INSERT TABLE 3 ABOUT HERE]

3.3.1 Gender

There were no significant differences in the proportion of male and female patients who exhibited any aggression ($n=220$, 63.0% vs. $n=88$, 56.4%) or physical aggression ($n=131$, 37.5% vs. $n=50$, 32.1%). There was no gender difference in mean total HCR-20 score. Women, however, had significantly higher H10 scale (14.10 vs. 13.51; $t[365]=-2.18$, $p=0.030$) and R5 scale (6.96 vs. 6.43 $t[503]=-2.19$, $p=0.029$) scores. Risk assigned by the SJ also differed significantly between men and women ($\chi^2[2, N=339]=21.59$, $p<0.001$); the standardised residuals revealed that women were more likely to be classified as high risk. Women were more likely than men to be under 40 years of age ($\chi^2[1, N=505]=5.43$, $p=0.020$) and less likely to be non-Caucasian ($\chi^2[2, N=505]=11.15$, $p=0.004$). Significant gender differences in terms of diagnostic group ($\chi^2[5, N=505]=130.09$, $p<0.001$) were due to over-representation of personality disorder diagnosis among women, and of schizophrenia, developmental and organic diagnoses among men. Men also had a significantly longer mean length of time between admission and HCR-20 assessment ($t[485]=2.18$, $p=0.030$), and were more likely to be detained under a forensic section ($\chi^2[2, N=505]=7.37$, $p=0.025$). As a result of these findings, ethnicity, age, diagnosis, legal status and time between admission and assessment were controlled for in the subsequent rocreg analyses. With the exception of the C5 scale, AUC values were higher for women than for men (see Table 4). The rocreg analysis indicated that these differences were significant in just under half of the comparisons. Except for the SJ, AUC values were higher for the prediction of any aggression than for physical aggression for both men and women.

[INSERT TABLE 4 ABOUT HERE]

Item-outcome analysis (see data supplement tables 5 and 6) revealed that, for the prediction of any aggression and physical aggression, items C5(4) (impulsivity) and R5(3) (lack of personal

support) were the strongest predictors among women; item C5(4) produced a large effect size for the prediction of any aggression of 0.823. For men, C5(4) and R5(4) (noncompliance with remediation attempts) were the strongest predictors of any aggression and C5(3) (active symptoms of major mental illness) and C5(4) were the best predictors of physical aggression; none of the AUC values reached the threshold for a large effect size in this group.

[INSERT TABLE 5 ABOUT HERE]

[INSERT TABLE 6 ABOUT HERE]

3.3.2 Diagnosis

Rates of any aggression ($\chi^2[4, N=399]=18.74, p<0.001$) and physical aggression ($\chi^2[4, N=399]=26.83, p<0.001$) differed between diagnostic groups. People with an organic diagnosis were more likely to manifest both types of aggression; there were no significant differences between any of the other diagnoses. Mean HCR-20 total score differed significantly across groups ($F[4, 395]=5.46, <0.001$). A post hoc Tukey test revealed that the Schizophrenia & Personality Disorder group had a significantly higher score than all other groups, with the exception of the Organic group. Whilst the overall test showed that the risk level assigned by the SJ differed significantly between groups ($\chi^2[8, N=257]=16.72, p=0.033$), none of the standardised residuals reached the generally accepted significance level of ± 1.96 (Sheskin, 2003). Nevertheless, the Schizophrenia & Personality Disorder group contained the largest proportion of people classified as high risk (40.6%) and the Developmental group the smallest proportion (17.6%). The Schizophrenia group had the largest proportion of people rated as low risk (36%) and the Organic group had the smallest proportion (10%). Gender ($\chi^2[4,$

$N=400$]=131.70, $p<0.001$), ethnicity ($\chi^2[8, N=400]=30.81, p<0.001$), age ($\chi^2[4, N=400]=43.56, p<0.001$) and mean time between admission and assessment ($F[4, 395]=2.67, p<0.032$) differed as a function of diagnoses and were controlled for in the subsequent analyses. Examination of standardised residuals and post-hoc Tukey tests revealed that i) the Schizophrenia group were more likely to be male, non-Caucasian and have a longer time between admission and assessment, ii) those with a personality disorder were more likely to be female, have an unknown ethnicity and be aged <40 ; they were less likely to be non-Caucasian, iii) the Developmental group were less likely to be female or aged ≥ 40 and had a shorter mean time between admission and assessment, and iv) those with an organic diagnosis were less likely to be female or aged <40 . Rocreg indicated that predictive efficacy of the HCR-20 differed in three instances: any aggression and physical aggression as predicted by the C5 scale, and physical aggression as predicted by the HCR-20 total (see Table 5). However, the C5 scale predicted any aggression for all diagnostic groups, except Organic, at better than chance levels. For the prediction of physical aggression, HCR-20 total was no better than chance for the Developmental and Organic groups and, for both groups the best prediction was from the R5 subscale. For predictions based on the HCR-20 total, H10 scale and C5 scale scores, the largest AUC values were obtained for the Schizophrenia group. With the exception of the SJ, AUC values were larger for the prediction of any aggression than physical aggression. Results of the item-outcome analyses are presented in supplementary tables 1 and 2. For all diagnostic groups, C5(4) (impulsivity) was one of the strongest predictors of both outcomes and produced large effect sizes in most instances, the largest being 0.829 for the prediction of any aggression in the Organic group. R5(2) (exposure to destabilizers) was one of the best predictors of both outcomes for the Organic, Developmental and Schizophrenia & Personality Disorder groups. For the Schizophrenia group, C5(3) (active

symptoms of mental illness) was a strong predictor of any aggression and R5(4) (noncompliance with remediation attempts) was a strong predictor of physical aggression. In those with a personality disorder R5(4) was among the best predictors of any aggression and R5(1) (plans lack feasibility) was a strong predictor of physical aggression. Additionally, H10(9) (personality disorder) produced a large effect size for the prediction of any aggression in the Organic group and H10(6) (major mental illness) produced AUC values significantly smaller than 0.5 for both outcomes in the Organic, Developmental and Schizophrenia & Personality Disorder groups.

[INSERT TABLE 7 ABOUT HERE]

3.3.3 Ethnicity

Neither incidence of aggression or risk levels assigned by the SJ differed significantly between the two groups. The only significant difference in mean HCR-20 scores was observed for the R5 scale ($t[325]=-1.99, p=0.047$), on which Caucasian people scored significantly higher. Non-Caucasians were less likely to be female ($\chi^2[1, N=327]=4.63, p=0.031$), more likely to reside in medium security ($\chi^2[1, N=327]=3.94, p=0.047$) and be detained under a forensic section ($\chi^2[2, N=8.87]=3.94, p=0.012$), more likely to have a diagnosis of schizophrenia and less likely to hold other or multiple diagnoses ($\chi^2[5, N=327]=31.42, p<0.001$). Gender, diagnosis, legal status and security level were therefore controlled for in subsequent analyses. The rocreg analyses provided no evidence that the predictive ability of the HCR-20 differed significantly as a function of ethnicity (see Table 6). However, the AUC values for the non-Caucasian group were larger than for the Caucasian group except for prediction of any aggression and physical aggression by the R5 scale. With the exception of the SJAUC values were again larger for the

prediction of any aggression compared with physical aggression. C5(2) (negative attitudes) and C5(4) (impulsivity) were the only significant predictors of any aggression among Caucasians and C5(4) was the only significant predictor of physical aggression; none of the AUC values produced a large effect size. C5(2) and C5(4) were also the only significant predictors of any aggression among the non-Caucasian group; physical aggression was predicted by H10(2) (young age at first violent incident) and C5(4) (see data supplement tables 5 and 6)..

[INSERT TABLE 8 ABOUT HERE]

3.3.4 Age

Those aged <40 scored significantly higher on the H10 scale ($t[420]=-6.62, p<0.001$) and those aged ≥ 40 had significantly higher C5 scale scores ($t[503]=2.21, p=0.028$). There were significant gender ($\chi^2[1, N=505]=5.43, p=0.020$) and diagnostic ($\chi^2[5, N=505]=44.22, p<0.001$) differences between groups; those aged <40 were more likely to be female, have a personality disorder or developmental diagnosis and were less likely to have an organic diagnosis than those aged ≥ 40 . Those aged <40 also had a significantly shorter mean time between admission and assessment ($t(289)=4.63, p<0.001$). Therefore, gender, diagnosis and time between admission and assessment were controlled for in the rocreg analyses. The H10 and C5 scales performed significantly differently as a function of age for both outcomes (see Table 6). AUC values obtained from the HCR-20 total and H10 scale were higher for people aged <40 years, while the opposite was true for the C5 and R5 scales. The H10 scale was only a significant predictor of any aggression for people aged <40 years. AUC values were again higher for the prediction of any aggression than physical aggression with the exception of the SJ. For those aged <40, C5(3)

(active symptoms of major mental illness) and C5(4) (impulsivity) were the best predictors of any aggression whilst R5(1) (plans lack feasibility) and R5(3) (lack of personal support) were the strongest predictors of physical aggression; none of the AUC values reached the threshold for a large effect size. The best predictors of any aggression in those aged ≥ 40 were C5(4) and C5(2) (negative attitudes); C5(2) produced a large effect size of 0.755. Physical aggression in this group was best predicted by C5(3) and C5(4); C5(4) had an effect size of 0.768 (see data supplement tables 5 and 6).

[INSERT TABLE 9 ABOUT HERE]

4. Discussion

This study of the predictive validity of the HCR-20 is the largest conducted in any inpatient setting to date. It showed that the HCR-20 total score is a significantly better predictor of any inpatient aggression, and specifically of person-directed physical aggression, for women than for men even after controlling for diagnosis, age, ethnicity, and time between admission and assessment. The HCR-20 predicted aggression better for people with diagnoses of schizophrenia and personality disorder, less well for organic and developmental disorders, and predicted better in younger than older participants. It performed equally well among Caucasian and non-Caucasian patients.

Overall, each one of the HCR-20 total, C5 and R5 scales, and the SJ were predictive of both any inpatient aggression and physical aggression; the historical scale was not predictive for either. These findings are consistent with previous research, although the AUC values are somewhat smaller than previously reported (Dolan and Doyle, 2000). AUC values for the SJ

were around 0.60 in this data whereas the AUC values for this predictor in previous studies have typically been the largest (O'Shea et al., 2013). One possible explanation of this difference is that the clinical HCR-20 template required a SJ regarding the likelihood that any future violence would *cause serious physical harm* as opposed to rating the *likelihood of violence itself* occurring, as was originally intended by the authors of the HCR-20 (Webster et al., 1997); hence, the aggressive outcome may not converge as closely with the prediction being made as in previous research. AUC values for the SJ were larger for the prediction of physical aggression towards others compared to any aggression and they may be expected to further increase for the prediction of more severe acts of physical aggression.

4.1 Gender

Our finding that the HCR-20 predicted aggression better in women than men is counterintuitive, and contrary to our study hypothesis, given that predictive validity is expected to be greatest in groups most like validation samples. The finding verifies, however, a meta-analysis in which effect sizes were greatest in samples with larger proportions of women (O'Shea et al., 2013). Interestingly, in the current study the predictive validity of the C5 scale for men is effectively equivalent to the total HCR-20 score for women. Item-outcome analysis revealed that more C5 items were significant predictors of both outcomes for men than for women. While the current study design precludes causal inferences about this finding, the data suggest some interesting possibilities. For example, de Vogel and de Ruiter (2005) reported that the most frequent additional risk factors listed by clinicians completing the HCR-20 differed by gender, suggesting that different factors drive aggression risk in women. Our finding that the HCR-20 total outperformed the C5 subscale for women but not men, suggests that features

beyond clinical presentation play a greater role in predicting risk, or influencing raters perception of risk, for women in the inpatient setting. The R5 scale was a significant predictor of aggression for women but the H10 scale was not, suggesting that it is largely these risk management factors that add incremental validity to the current clinical presentation in women. Item-outcome analysis supported this finding, revealing that more R5 items were significant predictors of both categories of aggression for women than men. While a limited number of R5 items predicted men's aggression they did not provide incremental predictive validity over the C5 scale.

Extrapolating from that finding to the clinical setting we suggest that, when making an overall summary judgement about the risk for inpatient aggression in women, due consideration must be afforded to the R5 and C5 items to maximise its predictive power. Conversely, accurate inpatient prediction in men might be better served by weighting the decision based on the presence of the crucial C5 items. A recent development has been the introduction of the Female Additional Manual (FAM; de Vogel and de Vries Robbé, 2013) which is intended to augment the HCR-20 for violence risk assessment of women. It contains nine additional items and additional guidelines for five of the existing HCR-20 (version 2) items. Development of the FAM was driven by observation of additional risk factors for women (de Vogel and de Ruiter, 2005) rather than by empirical data about the effects of gender on the performance of the HCR-20. Our data does not refute the need for a female-specific version of HCR-20; indeed it supports the central contention that a different range of risk factors predict women's aggression. However, it does suggest that, at least in terms of institutional aggression, practitioners need to be cognisant of the differential predictive properties of HCR-20 for women *and* men.

An alternative explanation for our finding of best performance in women is that their inpatient aggression is less well prevented, allowing them greater opportunity to manifest their potential for aggression. There is some evidence that mental health professionals view females as less dangerous than males and underestimate their risk (e.g. Wynn, 2002). This may lead to greater vigilance and earlier intervention among male patients, preventing a greater number of aggressive incidents, and less response to female patients' escalating aggression (Lam, et al., 2000). In support of this, research has shown that women receive PRN medication less frequently than men (Nicholls, et al., 2009) and are less likely to be physically restrained (Wynn, 2000; Stewart, et al., 2009) during aggressive incidents. Further, rates of engagement in self-injurious behaviours are higher in women than men (Claes, et al., 2007) and women are more likely to display aggression following suicidal behaviour (Nicholls, et al., 2009). It may be that greater resources are devoted to the management of self-harm, rather than aggression. It is impossible to statistically distinguish between the identification of false positives and effective risk management / prevention of aggression in the clinical setting (Dernevik et al., 2002) and so we cannot categorically reject this possibility. In future it may be possible to disentangle the impact of informed risk management on the predictive accuracy of risk assessment by investigating if measures such as de-escalation moderate the relationship between HCR-20 scores and aggression.

4.2 Diagnosis

Beyond the effect of gender our other specific study hypotheses were largely supported. The HCR-20's predictive validity was best in diagnostic groups most similar to validation samples, notably in those with schizophrenia and/or personality disorder rather than organic or

developmental disorders. Item-outcome analysis revealed that most C5 and R5 items significantly predicted aggression in the Schizophrenia and Personality Disorder groups, but fewer were significant predictors for individuals with developmental and organic diagnoses. We have already shown in a meta-analysis (O'Shea et al., 2013) that greatest effect sizes were in samples containing a higher proportion of people with schizophrenia though not for personality disorder. Here, the AUC values for the personality disorder group were similar to the schizophrenia group. Tengström et al (2006) reported similar AUC values for people with schizophrenia and personality disorder for antisocial behaviour and threats, but not for physical violence. It is possible that these differences stem from the composition of the personality disorders within the samples. Cluster A (odd or eccentric) and Cluster B (dramatic, emotional or erratic) personality disorders increase risk for violence (Johnson et al., 2000) while Cluster C (anxious or fearful) may actually be protective (Dean et al., 2006). In our study, only 8% of the personality disorder sample did not have at least one Cluster A or B disorder, compared to 37% in the sample used by Tengström et al (2006).

The HCR-20 total score was not a significant predictor of physical aggression towards others in developmental or organic disorders. However, the SJ was, suggesting that clinicians develop a good understanding of risk in these groups, and are rating risk accordingly, perhaps in spite of, rather than because of, the HCR-20. The AUC values for the R5 scale and the SJ were actually largest in these diagnostic groups. This suggests that the R5 items are most pertinent in the prediction of inpatient aggression in these groups and that these factors are being given more weight by clinicians when making their overall judgement.

4.3 Ethnicity

There were no differences in the AUC values as a function of ethnicity. This may be unexpected given that the HCR-20 was validated in predominantly Caucasian samples (Douglas, unpublished results) and previous meta-analyses have found larger effect sizes in samples containing higher proportions of Caucasians (O'Shea et al., 2013; Singh et al., 2011). While non-significant, the AUC values actually suggested better performance in non-Caucasians. Fujii et al (2005), also found no significant differences in predictive efficacy of the HCR-20 with ethnicity, but reported that the most predictive items did differ. Item-outcome analysis suggested that the most predictive items were also largely the same for these two groups; the exception being that H10(2) (young age at first violent incident) was a significant predictor of physical aggression in the non-Caucasian group.

4.4 Age

Item-outcome analysis revealed that consideration of both C5 and R5 items should maximise predictive validity for any aggression in both age groups. The H10 scale successfully predicted aggression in the younger group unlike in all other groups, and significantly better than for older patients. It may be that younger people have had less time to accumulate historical risk items and therefore the H10 scale will have more variance among this group. However, better relative performance appeared to reflect poor performance in older patients as much as its success as a predictor for younger ones. Further, the item-outcome analysis revealed that none of the individual historical items produced significant AUC values for either outcome. For the prediction of physical aggression, C5 items seem most pertinent to those over 40 years of age while the R5 items seem most relevant for the younger group. Singh et al (2011) found higher Diagnostic Odds Ratios, indicating superior predictive efficacy, in those over aged 40 years. This

is consistent with our clinical scale finding, but not the historical scale. As with women, it appears that a wider range of risk factors beyond current clinical features are salient to the prediction of inpatient physical aggression for younger people.

4.5 Relevance of individual scale items

Our item-outcome analysis revealed that not all items were significant predictors of aggression for all patients. Very few historical items predicted aggression. This likely reflects that psychiatric inpatients tend to score highly on historical items, resulting in insufficient variance for it to act as a meaningful predictor in this population (Belfrage, et al., 2000), rather than the H10 items being unrelated to aggression. That all the C5 and R5 scale items did not predict aggression for all groups may lead to a suggestion that resources might be saved by rating only those items with significant predictive power for the individual in question based on their group membership. However, non-predictive items based on their diagnostic group might well be predictive for their gender or age. For example, for males with personality disorder our results suggest that all C5 and R5 items would be predictive of any inpatient aggression based on either the individual's gender, diagnosis or both. Because individual patients sit in multiple categories they would be assessed on relatively few items that fail to predict aggression among people who share at least one important characteristic. We suggest that our findings should be used to inform the overall summary judgement for individual patients rather than to jettison individual items. Further, our item-outcome analysis suggests that it may be pertinent to focus risk-management strategies and interventions on targeting those items identified as particularly significant to a given group.

4.6 Rates of aggression

Base rates were slightly higher than in previous studies: 61% of patients in our sample had been aggressive during the three month follow-up period compared with 40-50% in other studies (Bowers et al., 2011; Dickens et al., 2013). We used a rich data source based on a simple recording mechanism where routine reports were flagged electronically thus optimising data capture. Additionally, the culture of identifying and recording aggression and violence in the study setting is strong.

4.7 Limitations

The 3-month period following assessment examined in the current study is one of the shorter follow-up periods in the current institutional-based HCR-20 literature; eight of twenty papers included in our previous meta-analysis had follow-up periods of 6 months or less (O'Shea, et al., 2013). This may have affected the predictive validity of the HCR-20 but was judged worthwhile as it rendered the sample sufficiently large to explore moderating influences. One factor that may have contributed to the lower AUC values in this study is that the HCR-20 risk assessments were completed by the patients' clinical teams as part of routine practice. This means that the people completing the risk assessment are also responsible for identifying, managing and preventing aggressive behaviour. Consequently, aggression rarely escalates to the severe levels due to effective interventions. This may also partly explain why predictive validity was better for any aggression over physical aggression; the presence of staff skilled in the de-escalation of violence is likely to result in prevention of some incidents, making this category inherently more difficult to predict. False prediction and successful prevention cannot be

disentangled in a ROC analysis. This becomes even more problematic with use of the OAS to code aggression since the OAS measures only objective outcome, rather than for example intent, degree of planning, or commitment to do harm (Yudofsky et al., 1986). In contrast, the HCR-20 authors indicate that “the resulting damage to a victim is not the defining feature of a violent act. Rather it is the act itself” (Webster et al., 1997; p. 24) and represents a flaw in our ability to recognise, record and understand the complexity of violence (Bowers et al., 2002). A further limitation of the current research is the reliance on pre-existing, routinely collected information. Whilst this allowed data collection relating to a much larger sample of patients than has previously been used, we were unable to verify diagnoses through structured methods. However, diagnoses would have been made or verified by the responsible clinician upon admission in accordance with ICD-10 guidelines. Also, we were missing around a third of data relating to ethnicity, resulting in the need to group all non-Caucasians patients together for analysis and possibly obscuring significant predictors of aggression among different ethnic groups. This may be inevitable as ethnicity is a self-defined variable and patients may choose not to select a category. Finally, although the study setting provides secure care for people with intellectual disability we did not include them in the sampling frame for this study. The reason for this was to minimize the level of complex co-morbidity in the sample. The predictive validity of the HCR-20 for this group will be reported separately in due course.

4.8 Implications and Future Considerations

The overall findings are consistent with previous research and indicate that the HCR-20 total score, dynamic scales and SJ are significant predictors of inpatient aggression. However, AUC values were smaller than have been observed in previous studies, particularly for the SJ.

Consideration of the items identified as significant predictors for the individual in question based on clinical and demographic characteristics when forming the SJ may improve its predictive efficacy. The most robust finding regarding group differences was that AUC values were greater in women compared with men, suggesting that the HCR-20 has superior efficacy among this group. This appears to be related to the role of future risk-management items as important predictors of female inpatient aggression. This supports the case for development of specialised female-specific risk assessment tools that focus on the full range of factors that predict these outcomes in women. However, the predictive ability of any new tools for women should be tested against the HCR-20 used with female only samples to ensure fair comparison. Future development of tools for assessing male inpatient risk would fruitfully focus on clinical factors in order to maximise their predictive validity. The fewest number of predictive items were found for patients with organic and developmental diagnoses suggesting a potential for improved prediction among these groups through identification and addition of relevant items. Finally, our item-outcome analysis may help clinical teams prioritise risk management interventions and treatment targets for their own patient group.

References

- Belfrage, H., Fransson, G., Strand, S., 2000. Prediction of violence using the HCR-20: A prospective study in two maximum-security correctional institutions. *The Journal of Forensic Psychiatry* 11, 167–175.
- Bowers, L., Stewart, D., Papadopoulos, C., Dack, C., Ross, J., Khanom, H., Jeffery, D., 2011. Inpatient violence and aggression: a literature review: Report from the Conflict and Containment Reduction Research Programme. Institute of Psychiatry, Kings College London.
- Buchanan, A., 2013. Violence risk assessment in clinical settings: Being sure about being sure. *Behavioral Sciences and the Law* 31, 74-80.
- Campbell, M. A., French, S., Gendreau, P., 2009. The prediction of violence in adult offenders: a meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior* 36, 567-590.
- Claes, L., Vandereycken, W., Vertommen, H., 2007. Self-injury in female versus male psychiatric patients: A comparison of characteristics, psychopathology and aggression regulation. *Personality and Individual Differences* 42, 611-621.
- Dean, K., Walsh, E., Moran, P., Tyrer, P., Creed, F., Byford, S., Burns, T., Murray, R., Fahy, T., 2006. Violence in women with psychosis in the community: prospective study. *The British Journal of Psychiatry* 188, 264-270.
- Dernevik, M., Grann, M., Johansson, S., 2002. Violent behaviour in forensic psychiatric patients: risk assessment and different risk-management levels using the HCR-20. *Psychology Crime and Law* 8, 93-111.
- de Vogel, V., de Ruiter, C., 2005. The HCR-20 in personality disordered female offenders: a comparison with a matched sample of males. *Clinical Psychology and Psychotherapy* 12, 226-240.

- de Vogel, V., & de Vries Robbé, M. (2013). Working with Women. Towards a more gender sensitive violence risk assessment (pp. 224-241). In Logan, C., Johnstone, L. (Eds), *Managing Clinical Risk: A guide to effective practice*. London: Routledge
- Dickens, G. L., Picchioni, M., Long, C., 2013. Aggression in specialist secure and forensic inpatient mental health care: incidence across care pathways. *Journal of Forensic Practice* 15, 206-217.
- Dolan, M., Doyle, M., 2000. Violence risk prediction. *The British Journal of Psychiatry* 177, 303-311.
- Douglas, K. S., Guy, L. S., Reeves, K. A., Weir, J., 2002-2008. HCR-20 violence risk assessment acheme: overview and annotated bibliography (<http://kdouglas.wordpress.com/hcr-20/>).
- Douglas, K.S., Hart, S.D., Webster, C.D. & Belfrage, H. 2013. HCR-20 (Version 3): assessing risk of violence – user guide. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Doyle, M., Dolan, M. C., 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., Tokioka, A. B., Lichten, A. I., Hishinuma, E., 2005. Ethnic differences in prediction of violence risk with the HCR-20 among psychiatric inpatients. *Psychiatric Services* 56, 711-716.
- Guy, L. S., Douglas, K. S., Hendry, M. C., 2010. The role of psychopathic personality disorder in violence risk assessments using the HCR-20. *Journal of Personality Disorders* 24, 551-580.
- Johnson, J. G., Cohen, P., Smailes, E., Kasen, S., Oldham, J. M., Skodol, A. E., Brook, J. S., 2000. Adolescent personality disorders associated with violence and criminal behavior during adolescence and early adulthood. *American Journal of Psychiatry* 157, 1406-1412.
- Khiroya, R., Weaver, T., Maden, T., 2009. Use and perceived utility of structured violence risk assessments in English medium secure forensic units. *Psychiatric Bulletin* 33, 129-132

- Lam, J. N., McNeil, D. E., Binder, R. L., 2000. The relationship between patients' gender and violence leading to staff injuries. *Psychiatric Services* 51, 1167-1170.
- McKenzie, B., Curr, H., 2005. Predicting violence in a medium secure setting: a study using the historical and clinical scales of the HCR-20. *The British Journal of Forensic Practice* 7, 22-28.
- Mossman, D., 1994. Assessing predictions of violence: being accurate about accuracy. *Journal of Consulting and Clinical Psychology* 62, 783-792.
- National Institute for Mental Health in England., 2004. *Mental Health Policy Implementation Guide*. (www.nimhe.org.uk).
- Nicholls, T. L., Brink, J., Greaves, G., Lussier, P., Verdun-Jones, S., 2009. Forensic psychiatric inpatients and aggression: an exploration of incidence, prevalence, severity, and interventions by gender. *International Journal of Law and Psychiatry* 32, 23-30.
- O'Shea, L. E., Mitchell, A. E., Picchioni, M. M., Dickens, G. L., 2013. Moderators of the predictive efficacy of the Historical, Clinical and Risk Management-20 for aggression in psychiatric facilities: systematic review and meta-analysis. *Aggression and Violent Behavior* 18, 255-270.
- Sheskin, D. J., 2003. *Handbook of Parametric and Nonparametric Statistical Procedure*, 3rd ed. CRC Press.
- Singh, J. P., Grann, M., Fazel, S., 2011. A comparative study of violence risk assessment tools: a systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review* 31, 499-513.
- Stewart, D., Bowers, L., Simpson, A., Ryan, C., Tziggili, M., 2009. Manual restraint of adult psychiatric inpatients: a literature review. *Journal of Psychiatric and Mental Health Nursing* 16, 749-757.
- Tengström, A., Hodgins, S., Müller-Isberner, R., Jöckel, D., Freese, R., Özokyay, K., Sommer, J., 2006. Predicting violent and antisocial behavior in hospital using the HCR-20: the effect of diagnoses on predictive accuracy. *International Journal of Forensic Mental Health* 5, 39-53.
- Warren, J. I., South, S. C., Burnette, M. L., Rogers, A., Friend, R., Bale, R., Van Patten, I., 2005. Understanding the risk factors for violence and criminality in women: the concurrent

- validity of the PCL-R and HCR-20. *International Journal of Law and Psychiatry* 28, 269-289.
- Webster, C. D., Douglas, K. S., Eaves, D., Hart, S. D., 1997. HCR-20: Assessing risk for violence version 2. Mental Health, Law and Policy Institute, Simon Fraser University.
- Webster, C. D., Eaves, D., Douglas, K. S., Wintrup, A., 1995. The HCR-20 scheme: the assessment of dangerousness and risk. Simon Fraser University and British Columbia Forensic Psychiatric Services Commission.
- Webster, C. D., Douglas, K. S., 2001. Purpose of the companion guide. In: Douglas, K. S., Webster, C. D., Hart, S. D., Eaves, D., Ogloff, J. R. P. (Eds.), HCR-20: Violence risk management companion guide. Mental Health, Law, and Policy Institute, Simon Fraser University and Department of Mental Health Law & Policy, University of South Florida, pp.1-12.
- Webster, C. D., Müller-Isberner, R., Fransson, G., 2002. Violence risk assessment: using structured clinical guides professionally. *International Journal of Forensic Mental Health* 1, 185-193.
- World Health Organisation, 1992. The ICD-10 Classification of Mental and Behavioural disorders. WHO.
- Wynn, R., 2002. Medicate, restrain or seclude? Strategies for dealing with violent and threatening behaviour in a Norwegian university psychiatric hospital. *Scandinavian Journal of Caring Sciences* 16, 287-291.
- Yudofsky, S. C., Silver, J. M., Jackson, W., Endicott, J., Williams, D., 1986. The Overt Aggression Scale for the objective rating of verbal and physical aggression. *American Journal of Psychiatry* 143, 35-39.