

Patient-rated impulsivity and aggression compared with clinician-rated risk in a forensic psychiatric sample: Predicting inpatient incidents

Julie Karsten¹  | Gerjonne J. Akkerman-Bouwsema² |
Loes A. Hagenauw² | Coby Gerlsma¹ | Marike Lancel^{1,2,3}

¹Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands

²Forensic Psychiatric Hospital Assen, Mental Health Services Drenthe, Assen, The Netherlands

³Psychiatric Sleep Center Assen, Mental Health Services Drenthe, Assen, The Netherlands

Correspondence

Julie Karsten, Faculty of Behavioural and Social Sciences, PO Box 72, 9700 AB Groningen, The Netherlands.
Email: j.karsten@rug.nl

Abstract

Background: Measures of impulsivity and aggression help to indicate risk of future violence or rule-breaking. Both clinician-rated risk assessment and self-report measures have been used but hardly ever compared in their ability to predict inpatient incidents.

Aims: To compare the self-report on the Barratt Impulsiveness Scale (BIS-11) and Buss-Perry Aggression Questionnaire (BPAQ) with the clinician-rated HKT-30, a Dutch adaptation of the Historical Clinical Risk Management-20, for their capacity to predict inpatient incidents.

Methods: All men newly admitted to a forensic psychiatric hospital were invited to participate in this study unless in intensive care. Tests of correlation were run between the BIS-11 and BPAQ scale scores and the HKT-30. Each was then tested separately for capacity to predict the number of aggressive and nonaggressive incidents while resident. Finally, scores of all rating scales were entered together into a negative binomial regression to compare their relative strengths in predicting later incidents.

Results: Patient and staff baseline impulsivity and aggression ratings correlated moderately well. All measures performed well in univariate analyses of relationship between

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. Criminal Behaviour and Mental Health published by John Wiley & Sons Ltd

baseline measures and later incidents. In final models, which included both patient and staff baseline ratings, the HKT-30 generally outperformed the self-report measures in the prediction of aggressive and nonaggressive incidents in both the first year and total length of stay.

Implications for clinical practice: Our findings suggest that some reliance may be placed on patient ratings of their own propensity for impulsive and/or violent acts, but, when used, they should remain combined with clinician-rated risk assessment for the time being. Future research should explore their utility in dialogue about treatment, and also the relative strength of staff response to each.

1 | INTRODUCTION

In forensic mental health settings, there is continuing concern about improving aids for identifying people at most risk of future violence, with particular emphasis on clinician ratings as in structured professional judgment (SPJ), using the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009) or the Historical Clinical Risk Management (HCR-20^{V3}; Douglas et al., 2014) or similar tools. Clinician-rated measures are thought to be less sensitive to bias than patient-rated measures, which may have a social desirable response tendency (Paulhus, 1984) or include unintentional self-deception (Ray, Hall, Poythress, Rivera-Hudson, & Lilienfeld, 2013). The validity of clinician-rated measures for predicting violent recidivism is reasonably well established (De Vogel, De Ruiter, Hildebrand, Bos, & Van de Ven, 2004; Lowder, Desmarais, Rade, Coffey, & Van Dorn, 2017), as well as for inpatient aggression (Marriott, O'Shea, Picchioni, & Dickens, 2017; Hogan & Olver, 2016).

Yet the use of self-report measures in general and forensic psychiatric care is increasing, as they are cheap, and administration requires little training. Further, self-report measures on impulsivity and aggression could increase insight into the extent to which these traits are present—on an affective or cognitive level. When using both clinician-rated and self-report measures, concurrence may be preferred in clinical practice, but discrepancy could be informative and used as a starting point for dialogue in treatment. Self-report measures of impulsivity and aggression might also be useful in the prediction of future violence, in particular inpatient aggression. To illustrate, the Barratt Impulsiveness Scale-11 (BIS-11; Patton, Stanford, & Barratt, 1995) has been found to be associated with institutional aggression (Dolan & Fullam, 2004; Wang & Diamond, 1999). Similarly, the Buss-Perry Aggression Questionnaire (BPAQ; Buss & Perry, 1992) was correlated with aggression on the ward among forensic psychiatric inpatients (Hornsveld, Muris, Kraaimaat, & Meesters, 2009). Thus, although these self-report measures show some promise as additions in forensic psychiatric practice, there has been little research comparison of self-reported impulsivity or aggression with clinician-rated measures.

In this study, we will compare two well-established self-report measures of impulsivity and aggression, the BIS-11 and the BPAQ, to a clinician-rated risk assessment instrument, the Historical Clinical Future-30 (HKT-30; Taskforce Risk Assessment Forensic Psychiatry, 2002), a Dutch adaptation of the Historical Clinical Risk Management-20 (HCR-20), its impulsivity item, and its hostility item. Furthermore, we will compare the BIS-11, the BPAQ, and the HKT-30 in their ability to predict the number of aggressive and non-aggressive inpatient incidents. We hypothesise a positive association between these measures and that both self-report measures and

the clinician-rated measure will independently contribute to the prediction of the number of incidents while in residence.

2 | METHODS

The project was approved by the local ethics committee (Hospital Ethics Committee, Isala Clinics in Zwolle, the Netherlands) and conducted in accordance with the Declaration of Helsinki.

2.1 | Sample

Data were drawn from a prospective observational study of male adult forensic psychiatric inpatients at the Forensic Psychiatric Hospital in the Netherlands. Every patient admitted between 2006 and 2016 was approached for participation 3 months after admission unless he fell into one or more of the following conditions: an expected stay of less than 1 year, insufficient understanding of the Dutch language, severe mental disability (established or estimated $IQ < 70$), or severe current psychiatric dysregulation requiring intensive care treatment including measures such as seclusion or restraint. Diagnoses according to DSM-IV-TR (American Psychiatric Association, 2000), established by clinicians, were collected from patient files.

2.2 | Measures

The *BIS-11* (Barratt, 1959) consists of 30 items scored on a 4-point Likert scale, ranging from 1 (*rarely/never*) to 4 (*almost always/always*). Higher scores reflect greater impulsivity. Factor analyses have resulted in three subscales: attentional impulsivity (8 items), motor impulsivity (11 items) and nonplanning impulsivity (11 items; Patton et al., 1995). In our study, internal consistency (IC), assessed with ordinal alpha (according to Gaderman, Guhn, & Zumbo, 2012), was good for the attentional impulsivity scale ($\alpha = .81$) and acceptable for motor impulsivity ($\alpha = .75$) and non-planning impulsivity scales ($\alpha = .71$).

The *BPAQ* (Buss & Perry, 1992) consists of 29 items scored on a 5-point Likert scale, ranging from 1 (*extremely uncharacteristic of me*) to 5 (*extremely characteristic of me*), higher scores reflecting more aggression. Items are divided over four subscales: physical aggression (9 items), verbal aggression (5 items), anger (7 items) and hostility (8 items), with the physical aggression ($\alpha = .90$) and hostility ($\alpha = .84$) scales having good IC, the anger scale acceptable IC ($\alpha = .75$), but the verbal aggression scale weak IC ($\alpha = .51$).

The *HKT-30* is a SPJ instrument based on the Historical Clinical Risk-management-20 (HCR-20; Webster et al., 1997), commonly used in the Netherlands and developed for the Dutch forensic system (Taskforce Risk Assessment Forensic Psychiatry, 2002). It consists of 11 items on the historical subscale, 13 items on the clinical subscale and 6 items on the risk management subscale. Items are scored on a 5-point Likert scale, with higher scores reflecting greater risk. Clinician-rated impulsivity and hostility are directly assessed as two items of the clinical subscale. The *HKT-30* total score has a moderate to good predictive value for violent recidivism (De Vries & Spreen, 2012; Hildebrand, Hesper, Spreen, & Nijman, 2005; Spreen et al., 2009). In our study, reported here, IC for the total score was $\alpha = .68$, in accordance with the wide range of behaviours measured.

2.2.1 | Incidents

The numbers of aggressive and nonaggressive incidents during the first year after assessment and during the total length of stay were collected from patient files and an incident registration system. To avoid criterion contamination, incidents in the first 3 months of stay, during which time the instruments were scored, were excluded from analyses. Aggressive incidents were defined as all behaviours causing harm to a person or object, verbal aggression included

any threats of physical aggression (but not merely shouting or disrespectful behaviour), nonverbal aggression which included clear threats of violence such as a cut-throat gesture, fire setting, or self-harming behaviour. Nonaggressive incidents included substance use, breaking hospital rules, not returning from hospital leave or returning too late, not following instructions from personnel, or disrespectful behaviour (e.g. name calling).

2.3 | Procedure

Three months after admission, patients were provided with verbal and written information about the study. After written informed consent, patients completed the BIS-11 and the BPAQ unaided, although a researcher was present in the room to oversee completion of the questionnaires. The HKT-30 was completed in the same time frame by their treating clinicians and added to the patients' files. These clinicians were all psychologists and psychiatrists who were trained in the use of the instrument and blind to the self-report questionnaires completed by the patients.

2.4 | Analyses

To assess convergent validity, the BIS-11 total score, the BPAQ total score, and their subscales were correlated to the HKT-30 total score, and its impulsivity and hostility items using Spearman's rho. Next, the BIS-11 and the BPAQ were used to predict the number of incidents in the first year and over the total length of stay. As the numbers of incidents are continuous count variables, Poisson regression analyses were considered, with the total number of incidents, the number of aggressive incidents and the number of nonaggressive incidents as dependent variables. However, we detected overdispersion in the count data; that is, the sample variance exceeded the mean; therefore, negative binomial regression analyses were performed instead. The binomial regression analyses were corrected for length of stay in the hospital, as longer stay increases the opportunity for incidents. First, the scales of the BIS-11 and BPAQ were entered into the model separately. Then, the scales of the BIS-11 and BPAQ were entered into the model simultaneously to assess their unique contribution to the prediction of the number of incidents. Last, the HKT-30 total score was added to the BIS-11 and BPAQ scales in the model. The reported adjusted rate ratio represents the factor by which the number of incidents increases with each point on the BIS-11 or BPAQ. One case was excluded from analyses concerning incidents as an outlier, as this particular patient had caused more than 100 incidents.

3 | RESULTS

3.1 | Sample characteristics

A total of 162 male patients completed the self-report questionnaires. Filling in the HKT-30 was not mandatory for clinicians, but the instrument was filled out for 115 patients, resulting in a final sample of 115 patients with all three measures available. Included patients had a mean age of 33.5 years ($SD=10.8$, range=19.1-61.9) at the time of assessment. They were charged with or convicted of various offences. Main offence distribution was sexual offences ($n=39$, 33.9%), assault ($n=44$, 38.3%), robbery ($n=16$, 13.9%), theft ($n=8$, 7.0%), arson ($n=7$, 6.1%), or homicide ($n=1$, 0.9%).

The majority of patients had at least one personality disorder or traits thereof ($n=104$, 90.4%), sometimes spanning multiple clusters ($n=35$, 30.4%). Most patients were diagnosed with at least one cluster B personality disorder or traits thereof ($n=100$, 87.0%), that is, antisocial ($n=73$, 63.5%), narcissistic ($n=51$, 44.4%), borderline ($n=34$, 29.6%) or histrionic ($n=3$, 2.6%). Some patients were diagnosed with a cluster C personality disorder or traits thereof ($n=37$, 32.2%), that is, avoidant ($n=30$, 26.1%), dependent ($n=16$, 13.9%) or obsessive compulsive ($n=3$, 2.6%). Cluster A personality disorders were less common ($n=4$, 3.5%); only 3 (2.6%) patients were diagnosed with schizoid personality disorder and

1 (0.9%) with paranoid personality disorder. One hundred and eight (93.9%) patients had at least one Axis I disorder, 81 (50.4%) two or more. The most common Axis I disorder was substance abuse disorder in environmentally enforced remission ($n=83$, 72.2%); 28 (24.3%) had a paraphilia, including paedophilia, 20 (17.4%) attention deficit hyperactivity disorder and 11 (9.6%) psychotic disorder. Ninety-eight (85.2%) had comorbid axis I and II conditions.

Means and standard deviations of the BIS-11, the BPAQ, the HKT-30 total score, and its impulsivity and hostility items are provided in Table 1. The subscales of the BIS-11 and the BPAQ were all correlated significantly, but within instrument Spearman correlations between subscales did not exceed $r=0.68$ and Spearman correlations between BIS-11 and BPAQ scales did not exceed $r=0.42$, indicating that multicollinearity was not an issue for analyses (Tabachnick & Fidell, 1996).

Correlations of the BIS-11 and the BPAQ with the HKT-30 total score and its impulsivity and hostility items are presented in Table 2. All scales of both the BIS-11 and the BPAQ were correlated with the clinician-rated HKT-30 total score, the strongest correlations being with the BPAQ. Of the BIS-11, the motor impulsivity scale, the non-planning scale and the total impulsivity score were related to clinician-rated impulsivity. Of the BPAQ, all scales and the total score were related to clinician-rated hostility. In addition, the physical aggression, verbal aggression and anger scales and the total score were correlated with clinician-rated impulsivity.

A total of 860 incidents were registered among 104 patients during a mean total length of stay of 25.1 months ($SD=12.9$ months); 213 of the incidents were aggressive and 647 were nonaggressive. About two-thirds of the incidents (565, 65.7%) took place in the first year after assessment, with 151 being aggressive and 414 nonaggressive.

First, associations between subscales and total scores of the two self-report measures were tested separately with the number of incidents. Negative binomial regression analyses with respect to self-rated impulsivity showed that higher total BIS-11 scores and higher scores on its nonplanning impulsivity subscale were associated with more aggressive and nonaggressive incidents, in the first year and in the total length of stay (Table 3). Nonplanning impulsivity especially was associated with incidents; the number of incidents increased by a factor of 1.05 to 1.11 for every point on the BIS-11 nonplanning scale. The motor impulsivity scale was associated with aggressive incidents in

TABLE 1 Means (M) and standard deviations (SD) of the BIS-11, BPAQ and HKT-30 (N=115)

	M (SD)
BIS	
Attentional	16.9 (4.4)
Motor	23.2 (5.1)
Non-planning	25.8 (4.9)
Total	65.9 (12.3)
BPAQ	
Physical	26.3 (9.7)
Verbal	14.3 (3.3)
Anger	19.1 (5.5)
Hostility	21.9 (7.0)
Total	81.6 (20.8)
HKT-30	
Impulsivity	1.7 (1.2)
Hostility	1.5 (1.0)
Total	57.5 (12.3)

Abbreviations: BIS-11: Barratt Impulsiveness Scale-11; BPAQ: Buss-Perry Aggression Questionnaire; HKT-30: Historical Clinical Future-30.

TABLE 2 Correlations(*r*) between the BIS-11, BPAQ and the HKT-30 (N=115)

	HKT-30		
	Impulsivity	Hostility	Total
	<i>r</i>	<i>r</i>	<i>r</i>
BIS-11			
Attentional	0.15	0.02	0.23*
Motor	0.20*	0.05	0.27**
Non-planning	0.19*	0.04	0.30**
Total	0.21*	0.04	0.31**
BPAQ			
Physical	0.39**	0.33**	0.46**
Verbal	0.32**	0.27**	0.36**
Anger	0.38**	0.35**	0.38**
Hostility	0.18	0.26**	0.24*
Total	0.38**	0.38**	0.45**

Abbreviations: BIS-11: Barratt Impulsiveness Scale-11; BPAQ: Buss-Perry Aggression Questionnaire; HKT-30: Historical Clinical Future-30.

Note.

* $p < .05$.; ** $p < .01$, two-tailed.

both the first year and in the total length of stay, and the attentional impulsivity scale was not associated with incidents over the first year after assessment but was related to non-aggressive incidents in the total length of stay.

Higher scores on self-rated aggression as measured by the BPAQ total score and all its subscales were also associated with a greater number of aggressive incidents in the first year after assessment and in the total length of stay. Higher scores on the physical aggression scale, but not anger or hostility subscales, predicted a higher number of non-aggressive incidents in the first year and in the total length of stay and thus just nudged association between total score and number of nonaggressive incidents into significance; higher score on the verbal aggression scale borderline predicted a higher number of non-aggressive incidents in the first year only.

Next, the subscales of the BIS-11 and BPAQ were added to the model simultaneously to test relationship their prediction of number of incidents in the first year and the total length of stay (Table 4). When thus combined, only higher scores on the BPAQ physical aggression scale were associated with higher numbers of aggressive and nonaggressive incidents in the first year and in the total length of stay. Scores indicated an increase of 1.03 to 1.05 in the number of incidents for every extra point on the BPAQ physical aggression scale. In addition, a lower score on the BPAQ anger was associated with a larger number of nonaggressive incidents during the first year.

Thirdly, the HKT-30 total score was added to the model with the BIS-11 and BPAQ subscales (Table 4). Higher scores on the clinician-rated HKT-30 were associated with higher numbers of all types of incidents in both the first year and the total length of stay. Regarding the self-report measures, results were in the same direction as when analysed without the HKT-30, although the BPAQ physical aggression score was no longer significantly associated with aggressive incidents in either time-frame or non-aggressive incidents in the total length of stay.

4 | DISCUSSION

In this study, with an inpatient population of male offender patients, we compared the relationships between a self-report measure of impulsivity (BIS-11) and another of aggression (BPAQ) to a clinician-rated measure of risk for their

TABLE 3 Adjusted rate ratios (95% CI) BIS-11 and BPAQ on incidents^a

	Incidents First Year			Incidents Total Stay		
	Total	Ag	NAG	Total	Ag	NAG
BIS-11						
Total	1.02** (1.01-1.04)	1.04** (1.01-1.06)	1.02* (1.00-1.04)	1.02** (1.01-1.04)	1.03** (1.01-1.05)	1.02** (1.01-1.03)
At	1.04 (1.00-1.09)	1.07 (1.00-1.14)	1.03 (0.99-1.08)	1.05** (1.02-1.09)	1.05 (0.99-1.12)	1.05** (1.01-1.09)
M	1.05* (1.01-1.09)	1.09** (1.03-1.15)	1.03 (1.00-1.07)	1.05** (1.01-1.08)	1.08** (1.02-1.13)	1.03 (1.00-1.07)
N-p	1.07** (1.02-1.11)	1.11** (1.04-1.18)	1.05* (1.01-1.10)	1.06*** (1.03-1.10)	1.08** (1.02-1.14)	1.06** (1.02-1.09)
BPAQ						
Total	1.02*** (1.01-1.03)	1.03*** (1.02-1.04)	1.01* (1.00-1.03)	1.02*** (1.01-1.03)	1.03*** (1.02-1.04)	1.01* (1.00-1.02)
P	1.06*** (1.03-1.08)	1.08*** (1.04-1.12)	1.05*** (1.02-1.07)	1.04*** (1.02-1.06)	1.08*** (1.05-1.12)	1.03** (1.01-1.05)
V	1.09** (1.03-1.16)	1.13** (1.04-1.22)	1.08* (1.00-1.16)	1.08** (1.03-1.14)	1.14** (1.06-1.23)	1.06 (0.99-1.12)
A	1.03 (0.99-1.07)	1.11*** (1.05-1.17)	1.01 (0.97-1.05)	1.04** (1.01-1.08)	1.12*** (1.07-1.18)	1.02 (0.99-1.05)
H	1.02 (0.99-1.05)	1.05* (1.01-1.10)	1.01 (0.98-1.04)	1.02 (1.00-1.05)	1.05* (1.01-1.09)	1.01 (0.99-1.04)

Abbreviations: A: BPAQ Anger scale; Ag: aggressive incidents; At: BIS-11 Attentional scale; BIS-11: Barratt Impulsiveness Scale-11; BPAQ: Buss-Perry Aggression Questionnaire; H: BPAQ Hostility scale; M: BIS-11 Motor scale; NAG: nonaggressive incidents; N-p: BIS-11 nonplanning scale; P: BPAQ Physical scale; V: BPAQ Verbal scale.

^aAdjusted for length of stay.

Note.

* $p < .05$; ** $p < .01$; *** $p < .001$.

TABLE 4 Adjusted rate ratios (95% CI) BIS-11, BPAQ and HKT-30 on incidents^a

	Model: BIS-11 and BPAQ						Model: BIS-11, BPAQ and HKT-30					
	Incidents First Year			Incidents Total Stay			Incidents First Year			Incidents Total Stay		
	Total	Ag	NAG	Total	Ag	NAG	Total	Ag	NAG	Total	Ag	NAG
BIS-11												
At	1.00 (0.95-1.06)	0.95 (0.87-1.04)	1.02 (0.96-1.08)	1.01 (0.97-1.06)	0.94 (0.87-1.02)	1.04 (0.99-1.09)	1.00 (0.95-1.06)	0.96 (0.86-1.04)	1.02 (0.96-1.08)	1.01 (0.97-1.06)	0.95 (0.88-1.02)	1.03 (0.98-1.08)
M	1.01 (0.96-1.06)	1.06 (0.97-1.15)	0.99 (0.94-1.05)	0.99 (0.95-1.05)	1.06 (0.98-1.14)	0.98 (0.94-1.03)	1.01 (0.96-1.06)	1.06 (0.98-1.14)	0.99 (0.94-1.05)	1.00 (0.96-1.04)	1.06 (0.99-1.14)	0.98 (0.94-1.02)
N-P	1.03 (0.99-1.09)	1.04 (0.95-1.13)	1.03 (0.98-1.09)	1.03 (0.99-1.08)	1.02 (0.94-1.10)	1.04 (0.99-1.09)	1.03 (0.99-1.08)	1.02 (0.94-1.11)	1.03 (0.98-1.08)	1.03 (0.99-1.07)	1.00 (0.94-1.08)	1.04 (1.00-1.08)
BPAQ												
P	1.06*** (1.03-1.09)	1.05* (1.00-1.11)	1.05** (1.02-1.09)	1.03* (1.01-1.06)	1.05* (1.01-1.10)	1.03* (1.00-1.05)	1.05** (1.02-1.08)	1.04 (0.99-1.09)	1.05** (1.02-1.08)	1.03* (1.00-1.07)	1.04 (1.00-1.08)	1.02 (0.99-1.04)
V	1.03 (0.95-1.11)	1.00 (0.89-1.012)	1.03 (0.94-1.13)	1.02 (0.95-1.10)	1.01 (0.91-1.12)	1.02 (0.94-1.11)	1.01 (0.94-1.09)	0.99 (0.89-1.11)	1.01 (0.92-1.10)	1.01 (0.94-1.07)	1.00 (0.90-1.10)	1.00 (0.92-1.07)
A	0.96 (0.91-1.01)	1.04 (0.95-1.14)	0.93* (0.88-0.99)	0.99 (0.95-1.04)	1.07 (0.99-1.16)	0.97 (0.92-1.02)	0.95* (0.90-1.00)	1.03 (0.95-1.13)	0.93** (0.88-0.98)	0.98 (0.94-1.02)	1.07 (0.99-1.15)	0.96 (0.92-1.01)
H	1.00 (0.97-1.04)	0.98 (0.93-1.04)	1.01 (0.97-1.05)	1.00 (0.97-1.03)	0.97 (0.92-1.02)	1.00 (0.97-1.03)	1.01 (0.97-1.04)	0.98 (0.93-1.04)	1.01 (0.97-1.05)	1.00 (0.97-1.02)	0.97 (0.92-1.01)	1.00 (0.97-1.03)
HKT-30												
							1.03** (1.01-1.05)	1.04** (1.02-1.07)	1.02* (1.01-1.04)	1.04*** (1.02-1.05)	1.05*** (1.02-1.07)	1.03*** (1.02-1.05)

Abbreviations: A: BPAQ Anger scale; Ag: aggressive incidents; At: BIS-11 Attentional scale; H: BPAQ Hostility scale; M: BIS-11 Motor scale; NAG: nonaggressive incidents; N-p: BIS-11 Nonplanning scale; P: BPAQ Physical scale; V: BPAQ Verbal scale.

^aAdjusted for length of stay.
 Note. *p<.05.; **p<.01.; ***p<.001.

association with subsequent incidents while still resident in the unit. The BIS-11 total score, the BIS-11 motor impulsivity scale and the BPAQ total score and all its subscales were modestly correlated with their clinician-rated counterparts as rated on the HKT-30.

In bivariate analyses, both the BIS-11 and the BPAQ were associated with later aggressive and nonaggressive incidents, while in residence, in both the first year and during the total duration of hospitalisation. This broad finding is in line with previous research both for self-reported impulsivity (Bousardt, Hoogendoorn, Noorthoorn, Hummelen, & Nijman, 2016) and self-reported aggression (Hornsveld et al., 2009) while resident in a hospital. By combining two self-report measures in our analyses, we were able to determine which aspects of impulsivity and aggression may help most as pointers to future violence or other antisocial behaviours. Combined, the self-reported BPAQ physical aggression, more than verbal aggression, anger, hostility, or any kind of impulsivity as measured by the BIS-11 predicted a whole range of inappropriate behaviour during hospitalisation. Our findings add to the current literature by indicating early markers of risk for both early to intermediate stages of inpatient care and also in the total duration of stay. Moreover, by adding the clinician-rated measure in the analyses, we further demonstrated that, in contrast to our expectation all three instruments would uniquely contribute to the prediction of incidents, even though the clinician-rated SPJ-measure generally outperformed both self-rated instruments for both timeframes.

The finding that the HKT-30 better predicted incidents than the two self-report measures may have several explanations. First, clinicians may indeed be better at assessing risk than patients themselves. Limited insight into psychological issues, level of risk of violence and factors contributing to it in forensic psychiatric populations have each been considered a risk factor for future violence (Douglas et al., 2014). Secondly, compared with the two self-reports measured, the HKT-30 has a broader scope, including a wide range of risk factors, such as substance abuse, social skills and coping skills, all related aggressive behaviour. Furthermore, the HKT-30 includes historical risk factors in addition to dynamic risk factors, such as previous convictions, being a former victim of abuse and behavioural problems in childhood, and the BIS-11 and the BPAQ may be considered to measure two dynamic risk factors only. Both historical and dynamic risk factors have been found to predict inpatient incidents (Vitacco, Gonsalves, Tomony, Smith, & Lishner, 2012). In short, the HKT-30 could be considered the more extensive instrument both in the number and type of measured risk factors. However, the finding that the BPAQ physical aggression scale and, to a lesser extent, the anger scale still contributed after adding the HKT-30 as a predictor of incidents, indicates that patient ratings on these scales add unique information not included in the HKT-30. To illustrate, the BPAQ physical aggression scale includes items about the inclination to use physical aggression in certain situations, which may not have happened in the hospital and therefore may not have been observed yet by clinicians, but is related to future risk. Using the BPAQ, in particular, the physical aggression scale, in addition to a clinician-rated measure of risk, may therefore be useful in forensic inpatient care.

Mean scores on the BIS-11 and BPAQ were lower than might be expected in an inpatient offender population, and, indeed, patient scores in this study were similar to scores found in the general population in previous research (Buss & Perry, 1992; Patton et al., 1995). Relatively low scores on the BIS-11 and BPAQ in forensic populations are not uncommon and have previously been thought to be the result of underreporting due to social desirability or limited insight (Haden & Shiva, 2008; Pettersen, Nunes, & Cortoni, 2016; Spaans, Molendijk, de Beurs, Rinne, & Spinhoven, 2017). Despite this possible bias in scores, we nevertheless found that the scales of the BIS-11 and BPAQ, especially the BPAQ physical aggression scale, can be employed to predict undesirable behaviour while hospitalised. A longitudinal study to explore possible changes in patient perception of their potential risk would be interesting in this context and of potential importance if much reliance is to be placed on this at different times. In certain admission phases, for example, one could hypothesise that an increase in self-rated risk indicator scores might better reflect improving insight than increasing risk.

4.1 | Strengths and limitations

A particular strength of our study lies in its step-wise analysis. First, we showed that two self-report measures correlate closely with clinician ratings on similar scales, made blind to the patients' ratings. Secondly, we showed that both

the patient ratings and the clinician ratings separately relate prospectively to later observed aggressive behaviours, but thirdly that the clinician ratings were apparently superior to the patient ratings for predicting incidents. Nevertheless, although significant, the correlations between the instruments were only small to moderate. This is to be expected, as the three instruments were designed to measure slightly differing constructs. Another drawback may lie in the fact that patients who participated in this study did so voluntarily, without compensation, with a possible implication that the sample was biased towards inclusion of more prosocial patients. We do not, however, consider this problematic. Given the high percentage of patients involved in incidents, however, it seems unlikely that self-selection resulted in a healthier, less problematic sample. Although no data were available on the patients who could not or would not participate in this study, scores on the BIS-11 and the BPAQ were comparable to those found in other forensic samples (Hornsveld et al., 2009; Stanford et al., 2009). A third possible limitation is that assessments took place in the same environment as treatment. Although clinicians and patients were blind to the individual outcomes on the self-report measures, patients scoring high on the BIS-11 and the BPAQ may possibly be more likely to discuss impulsivity and aggression as risk factors in treatment, which may have resulted in reduced numbers of incidents. Clinicians, having scored the risk assessment instrument themselves, were aware of the individual outcomes on the clinician-rated HKT-30, and as a result may have adjusted patient management or treatment. However, the number of incidents was still high, with almost all patients being involved in infractions grave enough to warrant reporting. It therefore seems unlikely that the number of incidents was influenced too much by the assessments. Though a small effect of assessments on undesirable outcomes may be favourable for studies on the predictive value of the instruments, this may be problematic from a clinical point of view. In clinical practice, clinicians are expected and should act on assessed risk in order to prevent problematic behaviour.

4.2 | Implications

Self-reported impulsivity and aggression scores were (a) comparable to previous findings in the general population, (b) only modestly correlated to clinician-rated impulsivity and hostility and (c) outperformed in their accuracy of predicting aggressive inpatient incidents by a clinician-rated instrument. Thus, insofar as self-rating tools are used in assessing risk of future violence or breaches of conditions of residence or leave, they should always be supplemented by sound staff ratings, notwithstanding the extra time this may take. As only some of the self-rating scales performed well, clinical practice and future research could focus on these. In addition, future research could focus on (a) the extent to which staff and patients communicate about such ratings, (b) or how patients direct communications about their own perceptions for risk and how these affect both sets of ratings, and (c) on the extent to which adjustment of management strategies follows from staff ratings alone or from patients flagging their own concerns.

CONFLICTS OF INTEREST

None.

FUNDING AND OTHER ACKNOWLEDGEMENTS

None.

ORCID

Julie Karsten  <https://orcid.org/0000-0001-5727-0245>

REFERENCES

- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: DSM_IV-TR*. Washington, DC: American Psychiatric Association.
- Bousardt, A. M., Hoogendoorn, A. W., Noorthoorn, E. O., Hummelen, J. W., & Nijman, H. L. (2016). Predicting inpatient aggression by self-reported impulsivity in forensic psychiatric patients. *Criminal Behaviour and Mental Health: CBMH*, 26(3), 161–173. <https://doi.org/10.1002/cbm.1955>
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, 63(3), 452–459. <https://doi.org/10.1037/0022-3514.63.3.452>
- 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. *The International Journal of Forensic Mental Health*, 3(2), 149–165. <https://doi.org/10.1080/14999013.2004.10471204>
- De Vries, K., & Spreen, M. (2012). Can the clinical items of the risk assessment instrument HKT-30 predict to what extent forensic patients will break agreements during the resocialisation phase of treatment? [De HKT-30 als instrument voor beslismomenten binnen een tbs-behandeling]. *Tijdschrift Voor Psychiatrie*, 54(5), 429–438.
- Dolan, M., & Fullam, R. (2004). Behavioural and psychometric measures of impulsivity in a personality disordered population. *Journal of Forensic Psychiatry & Psychology*, 15(3), 426–450. <https://doi.org/10.1080/14789940410001721048>
- Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., Guy, L. S., & Wilson, C. M. (2014). Historical-clinical-risk management-20, version 3 (HCR-20V3): Development and overview. *The International Journal of Forensic Mental Health*, 13(2), 93–108. <https://doi.org/10.1080/14999013.2014.906519>
- Gaderman, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 17(3). Retrieved from <http://pareonline.net/getvn.asp?v=17&n=3>
- Haden, S. C., & Shiva, A. (2008). Trait impulsivity in a forensic inpatient sample: an evaluation of the Barratt Impulsiveness Scale. *Behavioral Sciences & the Law*, 26(6), 675–690. <https://doi.org/10.1002/bsl.820>
- Hildebrand, M., Hesper, B. L., Spreen, M., & Nijman, H. L. I. (2005). *De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie. een onderzoek naar de betrouwbaarheid en predictieve validiteit van de HCR-20, HKT-30 en PCL-R*. Utrecht: Expertisecentrum Forensische Psychiatrie.
- Hornsveld, R. H. J., Muris, P., Kraaimaat, F. W., & Meesters, C. (2009). Psychometric properties of the aggression questionnaire in Dutch violent forensic psychiatric patients and secondary vocational students. *Assessment*, 16(2), 181–192. <https://doi.org/10.1177/1073191108325894>
- Lowder, E. M., Desmarais, S. L., Rade, C. B., Coffey, T., & Van Dorn, R. A. (2017). Models of protection against recidivism in involved adults with mental illnesses. *Criminal Justice and Behavior*, 44(7), 893–911. <https://doi.org/10.1177/0093854817710966>, <https://doi.org/10.1177/0093854817710966>
- Marriott, R., O'Shea, L. E., Picchioni, M. M., & Dickens, G. L. (2017). Predictive validity of the Short-Term Assessment of Risk and Treatability (START) for multiple adverse outcomes: The effect of diagnosis. *Psychiatry Research*, 256, 435–443. <https://doi.org/10.1016/j.psychres.2017.07.009>
- Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology*, 51(6), 768–774. doi: [http://dx.doi.org.proxy-ub.rug.nl/10.1002/1097-4679\(199511\)51:6<768::AID-JCLP2270510607>3.0.CO;2-1](http://dx.doi.org.proxy-ub.rug.nl/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1)
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598–609. <https://doi.org/10.1037/0022-3514.46.3.598>
- Pettersen, C., Nunes, K. L., & Cortoni, F. (2016). Does the factor structure of the aggression questionnaire hold for sexual offenders? *Criminal Justice and Behavior*, 43(6), 811–829. <https://doi.org/10.1177/0093854815618692>
- Spaans, M., Molendijk, M. L., de Beurs, E., Rinne, T., & Spinhoven, P. (2017). Self-reported personality traits in forensic populations: a meta-analysis. *Psychology, Crime & Law*, 23(1), 56–78. <https://doi.org/10.1080/1068316X.2016.1220555>
- Spreen, M., Ter Horst, P. R. M., Bogaerts, S., Lammers, S. M. M., Hochstenbach, J., Soe-Agnie, S. E., & Reubsat, R. J. (2009). Risicotaxatie aan de hand van de HCR-20 en de HKT-30: Een vergelijking tussen beide instrumenten. In T. I. Oei, & M. S. Groenhuijsen (Eds.), *Forensische psychiatrie en haar grensgebieden. actualiteit, geschiedenis en toekomst* (pp. 459–476). Kluwer: Alphen aan de Rijn.
- Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: an update and review. *Personality and Individual Differences*, 47(5), 385–395. <https://doi.org/10.1016/j.paid.2009.04.008>
- Taskforce Risk Assessment Forensic Psychiatry (2002). *Manual HKT-30: Risk assessment in forensic psychiatry*. Den Haag: Dienst Justitiële Inrichtingen.
- Vitacco, M. J., Gonsalves, V., Tomony, J., Smith, B. E. R., & Lishner, D. A. (2012). Can standardized measures of risk predict inpatient violence? Combining static and dynamic variables to improve accuracy. *Criminal Justice and Behavior*, 39(5), 589–606. <https://doi.org/10.1177/0093854812436786>

Wang, E. W., & Diamond, P. M. (1999). Empirically identifying factors related to violence risk in corrections. *Behavioral Sciences & the Law*, 17(3), 377–389. [https://doi.org/10.1002/\(SICI\)1099-0798\(199907/09\)17:3<377::AID-BSL351>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1099-0798(199907/09)17:3<377::AID-BSL351>3.0.CO;2-M)

How to cite this article: Karsten J, Akkerman-Bouwsema GJ, Hagenauw LA, Gerlsma C, Lancel M. Patient-rated impulsivity and aggression compared with clinician-rated risk in a forensic psychiatric sample: Predicting inpatient incidents. *Crim Behav Ment Health*. 2019;29:296–307. <https://doi.org/10.1002/cbm.2131>