International Journal of Law and Psychiatry 66 (2019) 101448

Contents lists available at ScienceDirect

International Journal of Law and Psychiatry

journal homepage: www.elsevier.com/locate/ijlawpsy

Inpatient violence in forensic psychiatry: Does change in dynamic risk indicators of the IFTE help predict short term inpatient violence?

Erwin Schuringa[a,*], Marinus Spreen[a,b], Stefan Bogaerts[c,d]

[a] Forensic Psychiatric Centre Dr. S. van Mesdag, Post box 30.002, Groningen 9700 RC, the Netherlands
[b] NHL Stenden University of Applied Sciences, Leeuwarden, the Netherlands
[c] Department of Developmental Psychology, Tilburg University, Tilburg, the Netherlands
[d] Fivoor Science & Treatment Innovation, the Netherlands

ARTICLE INFO

Keywords:
IFTE
Inpatient violence
Change
Forensic psychiatry
Routine outcome monitoring

ABSTRACT

Inpatient violence is a form of recidivism in forensic psychiatric treatment and is stated as an adverse outcome of treatment and a predictor for recidivism after release from the institution. Dynamic Risk Indicators (DRI) are critical key indicators that can predict inpatient violence, but little is known about the effects of change in DRI during forensic psychiatric treatment on the prediction of inpatient violence. This study examines the effects of change in DRI on the prediction of short-term inpatient violence using the Instrument for Forensic Treatment Evaluation (IFTE).

A group of 96 patients is followed from entering a high secure forensic hospital until their fifth measurement approximately three years later. The outcome measure is defined as any inpatient violence six months after measurement five. Repeated measures are used to study whether there was a difference in change in DRI between the group of patients who did or did not committed inpatient violence. Binary logistic regression is used to establish the extent to which changes in DRI add to the predictive power of the last measurement.

At the group level, the extent of change in DRI did not discriminate between the two patient groups. A large part of the 96 patients already scored low on DRI when entering the hospital and did not (need to) change. At all five measurements violent patients had significant higher scores on DRI than nonviolent patients. Logistic regressions showed that the last measurement predicts inpatient violence sufficiently, the change in DRI during the first four measurements did not contribute to this prediction.

The change in dynamic risk indicators does not help to predict short term inpatient violence. The last measurement is the most practical predictor for short term inpatient violence, but because of the dynamic nature of these indicators it is necessary to frequently monitor these indicators to detect imminent risks.

1. Introduction

Rehabilitation programs in forensic psychiatry designed according to the three principles of the Risk-Need-Responsivity (RNR) model are evaluated as most effective to reduce recidivism (Andrews & Bonta, 2010; Andrews, Bonta, & Hoge, 1990). The most intensive treatment should be given to high-risk patients (Risk principle), patient's specific criminogenic needs must be addressed (Need principle), and a patient's learning style, motivation and competences must be considered among other personal characteristics for obtaining an effective treatment (Responsivity principle). Criminogenic needs are commonly considered as risk factors that are supposed to change positively by specific interventions during treatment (Andrews & Bonta, 2010). Whenever the change in criminogenic needs is considered sufficient, the risk of reoffending after release is considered low (Van der Veeken, Bogaerts, & Lucieer, 2018).

Andrews and Bonta (2010) distinguished between the Big Four (criminal history, pro-criminal attitudes, pro-criminal associates and antisocial personality patterns) and the Moderate Four (family/marital relationships, social achievement, substance abuse, leisure /recreation) criminogenic needs. The Big Four Needs have been found to be directly associated with recidivism after release, while the Moderate Four Needs are indirectly associated with recidivism. The most important goal of forensic treatment is changing these criminogenic needs, except for the criminal history which is unchangeable (McGrath & Thompson, 2012).

Change in criminogenic needs can be measured by repeated measurements. According to Wilson, Desmarais, Nicholls, Hart, and Brink (2013) repeated measurements have several advantages. First, to

* Corresponding author.
E-mail address: E.Schuringa@fpcvanmesdag.nl (E. Schuringa).

https://doi.org/10.1016/j.ijlp.2019.05.002
Received 18 January 2019; Received in revised form 12 April 2019; Accepted 9 May 2019
Available online 13 June 2019

monitor changes in a patient's risk and treatment needs, and second, to make better informed treatment decisions to direct treatment and prevent (inpatient) violence.

In forensic psychiatry, clinical items from risk assessments instruments may be used to monitor behavioral change (Chagigiorgis, Michel, Seto, Laprade, & Ahmed, 2013; Lewis, Olver, & Wong, 2013). Several studies showed that repeated measurements using clinical items from risk assessment instruments contain better and more valuable information than single time-point measurements (Hochstetler, Peters, & DeLisi, 2016; Labrecque, Smith, Lovins, & Latessa, 2014). Furthermore, improvements in criminogenic needs are associated with lower rates of recidivism after treatment (Cohen, Lowenkamp, & VanBeschaoten, 2016; de Vries Robbé, de Vogel, Douglas, & Nijman, 2015; Mooney & Daffern, 2013; Serin, Lloyd, Helmus, Derkzen, & Luong, 2013). In a 24-month period Raynor (2007) found that individuals decreasing in Level of Service Inventory - Revised (LSI-R) total score were less likely to reoffend (42%) than those with increasing total LSI-R scores (67%). In another study among a group of high-risk forensic psychiatric patients using the Instrument of Forensic Treatment Evaluation (IFTE; Schuringa, Spreen, & Bogaerts, 2014), patients with low scores on the factors Protective behavior and Resocialization skills and high scores on the factor Problematic behavior displayed significant positive treatment progress during a 3-year period follow-up period (Van der Veeken et al., 2018).

During intramural forensic psychiatric treatment the staff must always be aware and alert on the risk of inpatient violence (Jeandarme et al., 2016). Inpatient violence is a form of recidivism, which occurs frequently in forensic psychiatry with severe emotional and physical consequences for both victims (co-patients and/or personnel) and perpetrators (Dack, Ross, Papadopoulos, Stewart, & Bowers, 2013; O'Shea, Picchioni, & Dickens, 2016; Schuringa, Heininga, Spreen, & Bogaerts, 2016). Inpatient violence is also a strong predictor for recidivism after release (Daffern et al., 2007; French & Gendrau, 2006). Inpatient violence is an adverse outcome as well as a signal for increased risk of post-treatment recidivism.

Dynamic criminogenic needs, which are part of the risk assessment instruments Historical Clinical Future - Revised (HKT-R; Spreen, Brand, ter Horst, & Bogaerts, 2014; Bogaerts et al., 2018) and Historical Clinical Risk - 20 version 3 (HCR-20v3; Douglas, Hart, Webster, & Belfrage, 2013) have shown to be associated with inpatient violence (e.g., Desmarais, Nicholls, Wilson, & Brink, 2012; O'Shea & Dickens, 2015; van der Veeken, Lucieer, & Bogaerts, 2016). However, both instruments are not originally intended for predicting short term (< six months) violence.

In a group of regular psychiatric patients, Abderhalden et al. (2008) found a substantially reduced level of short-term inpatient aggression (< 100 days) and coercive measures by applying a structured risk assessment instrument, the Staff Observation Aggression Scale – Revised (SOAS-R; Nijman et al., 1999). The Short Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Middleton, 2004) is one of the most studied instruments to predict short term inpatient violence. However, most studies in which the START is used are characterized by single-time point measurements (e.g., Desmarais et al., 2012; O'Shea et al., 2016; O'Shea, Picchioni, McCarthy, Mason, & Dickens, 2015). An exception is the study of Whittington et al. (2014) who analyzed multiple measurements of the dynamic START indicators and reported that an increased risk score is associated with increased likelihood of inpatient violence.

In sum, criminogenic needs can be best measured over time instead of single-time points measurement. Moreover, it's important to register inpatient violence as this is an important predictor for post-release recidivism. The question however is whether the change in criminogenic needs on its own, the direction and variation in scores or only the most recent measurement does predict inpatient violence. Two studies of the IFTE, a Dutch forensic routine outcome measurement instrument derived from the HKT-R, have shown sufficient predictive power of some individual dynamic IFTE indicators of the factor Problematic behavior (impulsive, antisocial, and hostile behavior, compliance to rules and antisocial associates) for short-term (six months) inpatient violence (Schuringa et al., 2016; Van der Veeken et al., 2016.

In this paper, we explore the extent to which inpatient violence can be assessed by the level and change of Problematic behavior as measured by the IFTE. Some individual indicators of the factor Problematic behavior are combined into one factor, called the Dynamic Risk Indicators (DRI). Based on the discussed literature we expect that predicting short term inpatient violence from one single measurement (the most recent) without considering trends in the earlier measurements will be less strong than when taking these trends into account.

2. Methods

2.1. Procedure

The IFTE data used in this study were extracted from the Routine Outcome Monitoring (ROM) system of the Dutch maximum-security Forensic Psychiatric Centre Dr. S. van Mesdag (hereafter: Mesdag). IFTE measurements in the period April 2010 until July 2016 were included.

Permission for this study was given by the institution's director and the institutions committee of behaviorists, which is in accordance with the declaration of Helsinki (World Medical Association, 2013). According to the Dutch law on medical research in humans, patients in this study did not need to give permission because it concerns a retrospective study on electronic files. Permission of a medical ethical committee was therefore not required (www.ccmo.nl; The Central Committee on Research involving Human Subjects). This study was conducted according to the guidelines for Good Clinical Practice in mind (GCP; Pieterse, 2015).

2.2. Instrument

The Instrument for Forensic Treatment Evaluation is a multidisciplinary Routine Outcome Monitoring instrument. The IFTE is filled out in approximately 10 min every six months by all members of a patient's treatment team independently. The IFTE contains 22 indicators, comprising all 14 clinical criminogenic need indicators of the Dutch risk assessment instrument HKT-R (Spreen et al., 2014), three indicators based on the Atascadero Skills Profile (ASP; Vess, 2001), and five indicators designed in consultation with psychologists and psychiatrists. The 22 IFTE indicators are divided into three factors, namely Protective behaviors, Problematic behaviors and Resocialization Skills. Indicators of the IFTE are measured on a 17-point scale.

A distinguishing feature of the IFTE from other ROM instruments, such as the START (Webster, Martin, Brink, Nicholls, & Middleton, 2004), is the standardized way a multidisciplinary evaluation is applied to one patient. Each individual score of the different disciplines involved in the treatment of the patient is scored before the meeting, instead of a consensus score during the meeting. Consensus scores can be biased by group dynamic processes during meetings, while with the IFTE all raters, fill out the IFTE beforehand and independently and are instructed to only score observed behavior. Additionally, it is possible for raters to score 'not enough information' for indicators which were not observed during treatment. The enlarged 17-point scale is much more sensitive for measuring behavioral change, which is recommended by Serin et al. (2013), and Hildebrand and De Ruiter (2012).

A standard IFTE report consists of the mean score of all raters on all indicators individually and on the three factors. The mean score is seen as the best depiction of the observed behavior in different situation. A measurement of agreement (between 0 and 1) is calculated between the raters per indicator. This measurement is an indication of how close the observations of the raters are to each other and thus if the patient shows the same behavior with different therapists and thus in different

situations. A low measurement of agreement is informative for the treatment meeting, because different therapists can discuss the reason of the difference in observed behavior and can learn from each other's interventions.

In sum, the IFTE collects multidisciplinary forensic relevant information in an efficient manner and is sensitive for change, which makes the IFTE very suitable for repeated measures.

2.3. Participants

The Mesdag is a maximum-security forensic hospital for mentally disordered offenders, who are hospitalized under the judicial measure of a 'terbeschikkingsstelling-order' (tbs-order; Entrustment Act), which is a "Provision in the Dutch criminal code that allows for a period of treatment following a prison sentence for mentally disordered offenders" (van Marle, 2002, p. 83). The tbs-order is not an additional punishment on top of a prison sentence, but a measure to protect society against further offences. A tbs-order is reviewed every one or two years by the court and is prolonged if a court deems a patient still at risk of reoffending (van Marle, 2002). Patients with a tbs-order are held not (completely) accountable for the crime they have committed because of a mental condition which played a role while committing the crime. Furthermore, the crime committed must have a minimum penalty of at least 4 years. Treatment in the Mesdag is voluntary, but the confinement is not.

A sample of 306 patients was extracted from the ROM database. Because the IFTE was introduced in 2010, for a substantial part of the patients the first measurement did not always take place at admission. In this study, only patients were included who were at the beginning of their treatment. Patients were included who had their first IFTE measurement within 24 months of hospitalization ($M = 7.99$ months, $SD = 6.44$, range: 3–24) and five consecutive measurements within 38 months ($M = 27.77$, $SD = 4.61$, range: 21–38). Having in mind that the average stay of a tbs-order was about 8.4 years (101 months) (Nagtegaal, Horste, & Schönberger, 2011), a measurement within 24 months was considered as a baseline measurement. The resulting group consisted of 96 male patients.

3. Measurements

3.1. Dynamic risk indicators

In a previous study of the IFTE conducted by Schuringa et al. (2016), some individual indicators of the factor Problematic behavior (impulsive behavior, antisocial behaviors, hostile behavior, manipulative behavior, compliance to rules, antisocial associates and drug use), showed to be significant discriminative between patients who committed short term inpatient violence and those who did not (4–8 months; Cohen's d from -0.51 till 1.08). The remaining indicators of the factor Problematic behavior, i.e., psychotic symptoms and sexual deviant behavior did not show any discriminative power, and therefore were not included in the current study. The significant discriminative indicators of the factor Problematic behavior were aggregated into a sum score, denoted by the variable Dynamic Risk Indicators (DRI).

3.2. Outcome measure

The outcome variable was inpatient violence that occurred within a four to nine month follow up period ($M = 6.22$, $SD = 0.89$) after the fifth IFTE measurement and denoted Inpatient Violence, where 1 means having caused one or more violent incidents and 0 having caused no violent incident. Inpatient violence was defined as intentional behavior, which could or did physically harm a person or animal, and/or a form of (verbal) aggression, which was extremely intimidating or threatening (Troquete et al., 2013). Violent incidents were retrospectively coded from the reports of the sixth IFTE treatment evaluation, which covers all relevant behaviors of a patient in the period at risk.

3.3. Statistical analysis

3.3.1. Internal consistency, descriptive and AUC-values of the DRI

Internal consistency of the DRI scale was established by Cronbach's alpha as well as item-total correlations. Each of the five measurements, DRI were univariately described by mean, standard deviation and range; for the total group as well as for inpatient violent and non-violent patients separately. The differences in mean were tested using Mann-Whitney tests ($p < .05$) and Cohen's d. Where $d = 0.2$ is a small effect, $d = 0.5$ medium and $d = 0.8$ large. This was also done for the difference in DRI between measurement five and measurement one (Δ DRI). Area Under the Curve-values (AUC) were calculated through Receiver Operant Characteristic-analyses with Inpatient Violence as dichotomous outcome. An AUC-value between 0.60 and 0.70 is considered moderate, between 0.71 and 0.80 acceptable, between 0.81 and 0.90 is excellent and larger than 0.91 is outstanding (Hosmer & Lemeshow, 2000).

3.3.2. Repeated measures analysis and comparison of treatment period and follow-up period of violent and non-violent patients

A repeated measures design (General Linear Model) with the five IFTE measurements as within subject factor and Inpatient Violence as between subject factor was used to explore whether there was a significant change over time in DRI and whether violent and non-violent patients differed with respect to DRI change.

Treatment period and follow-up periods are tested using Mann-Whitney test with inpatient violence as outcome variable.

3.3.3. Binary logistic regression

To establish to what extent DRI measurement five (DRI_5) predicted inpatient violence after measurement 5, on its own and to what extent a change in DRI (Δ DRI) between measurement 1 (DRI_1) and measurement 5 was additional predictive, binary logistic regressions were performed. In the first model, only the independent variable DRI_5 (which is the sum score on DRI for measurement 5), was submitted. In model 2 only the independent variable the change of DRI (Δ DRI) was examined. In model 3 Δ DRI and DRI_5 were submitted together to analyze whether Δ DRI added to the predictive power of DRI_5. Also, the interaction between the entered variables was explored and removed from the model in case of non-significance. The models were compared by the log likelihood test. The percentage of correctly classified patients and the numbers needed to detain (NND) were calculated (Fleminger, 1997). NND displays the number of patients which should be detained, to prevent one violent occurrence.

4. Results

4.1. Participants

Table 1 displays some characteristics of the violent and non-violent group.

4.2. Internal consistency, descriptive and AUC-values of DRI

Cronbach's alpha of DRI at measurement 5 was acceptable being $\alpha = 0.83$ and an item-total correlation ranging from 0.44 to 0.76, which was also acceptable. Twenty-seven percent (N = 26) of the patients had caused a violent incident after measurement 5. Table 2 displays the DRI scores for the total group, violent group and non-violent group and the AUC-values for inpatient violence for each single measurement. For all measurements, there was a significant difference in DRI between violent and non-violent patients.

The AUC-values of all measurements were not significantly different from each other. The mean change in DRI was almost zero for both

Table 1
Characteristics of the violent and non-violent groups.

	Violent (N = 27) *M (SD)*	Non-violent (N = 70) *M (SD)*
Age	35.23 (8.60)	38.39 (11.15)
Number of Diagnosis	3.11 (1.34)	3.71 (1.33)
Axis 1 of DSM IV-TR[a]		
Schizophrenia or other psychotic disorder	11 (42%)	37 (53%)
Mood and anxiety disorder	3 (11%)	7 (10%)
Development disorder	6 (22%)	12 (17%)
Substance abuse	37	109
Pedophilia/paraphilia	1 (4%)	16 (23%)
Other	4 (15%)	12 (17%)
Number of patients with at least one substance (ab)use–related diagnosis	21 (81%)	55 (79%)
Axis 2		
Cluster A Personality disorder	0	2 (3%)
Cluster B Personality disorder	12 (44%)	28 (40%)
Cluster C Personality disorder	0	1 (1%)
Personality disorder NOSc	3 (11%)	16 (23%)
Mental retardation	3 (11%)	25 (36%)
Other	4 (15%)	3 (4%)
Index offences		
Homicide	8 (30%)	21 (30%)
Violence	11 (41%)	16 (23%)
Sexual offence	4 (15%)	24 (34%)
Theft with and without violence	1 (4%)	3 (4%)
Arson	2 (7%)	6 (9%)

[a] DSM-IV-TR, APA, 2000.

groups, although when considering the range of change of Δ DRI (−5.06 to 5.26) was rather high but this change did not predict inpatient violence by itself (AUC = 0.45).

4.3. Repeated measures analysis

Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 37.01$, $p < .001$, therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.83$). The results showed that DRI did not differ significantly over measurements, $F(3.3, 312.78) = 1.004$, $p = .40$. No interaction effect was found between inpatient violence and measurements, $F(3.5, 329.08) = 0.169$, $p = .94$. However, there was a significant effect between groups $F(1,94) = 703,73$, $p = .001$ implying that on average DRI did not change over time, but a significant constant difference in level of DRI between both groups at all measurements was observed (see Fig. 1).

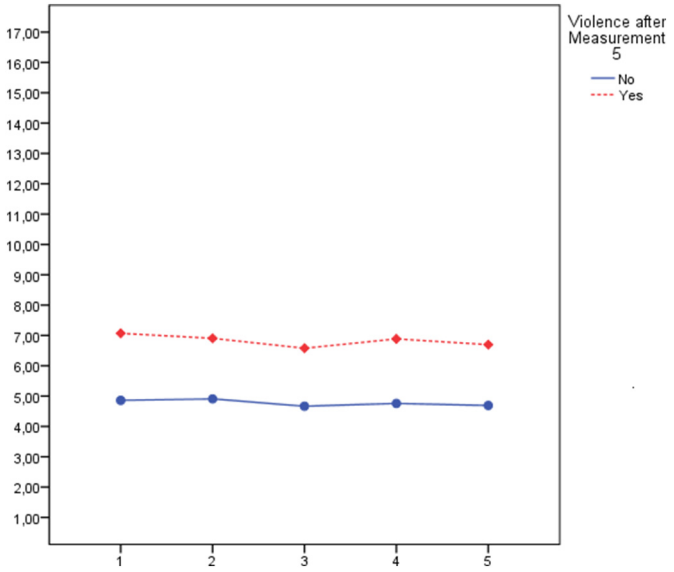

Fig. 1. Mean DRI for 5 measurements for violent and non-violent patients.

There is no difference between the violent and non-violent group on treatment duration until measurement 5 ($U = 1.017,00$, $z = 0.884$, $p = .376$) and there is no difference in follow-up period ($U = 968,00$, $z = 0.557$, $p = .577$).

4.4. Binary logistic regression

Results of the different binary logistic regressions are displayed in Table 3.

In model 1 DRI was significant as an univariate predictor, while the change in DRI was not univariately significant. In model 3 the addition of Δ DRI to model 1 did not add sufficient explained variance on the prediction of inpatient violence by the last measurement. There was no significant difference between model 1 and model 3 ($\chi^2(2) = \Delta$-2LL = 1.781; $p = .410$). There was no interaction effect in Model 3. Model 1 classified 77,1% of the patients correctly and had an odds ratio of 1.64 and the NND = 2.17. In comparison, the NND without any information but the rate of inpatient violence was 3.70 (1/27%).

Table 2
Means, effect sizes and AUC values for the different time points.

	Total M (SD) (Range) N = 96	Violent M (SD) (Range) N = 26	Non-Violent M (SD) (Range) N = 70	Difference Violent- Non-violent[a] (Z) Df = 94	Effect Size (Cohen's d)	AUC (95% CI)
DRI_1	5.46 (2.66) (1.19–13.29)	7.07 (3.15) (1.19–13.29)	4.85 (2.18) (1.61–10.14)	2.22** (3.39)	0.82	0.73** (0.60–0.85)
DRI_2	5.45 (2.49) (1.66–13.00)	6.91 (3.07) (2.07–13.00)	4.91 (2.01) (1.66–10.33)	2.00** (2.89)	0.77	0.69** (0.56–0.83)
DRI_3	5.18 (2.28) (1.57–11.05)	6.58 (2.45) (2.70–11.05)	4.67 (2.00) (1.57–10.26)	1.91** (3.41)	0.85	0.73** (0.61–0.84)
DRI_4	5.33 (2.31) (1.38–10.68)	6.89 (2.96) (1.71–10.68)	4.75 (1.71) (1.38–8.81)	2.14** (3.15)	0.89	0.71** (0.57–0.85)
DRI_5	5.23 (2.11) (1.32–10.87)	6.70 (2.50) (2.93–10.87)	4.69 (1.65) (1.32–8.44)	2.01** (3.46)	0.95	0.73** (0.62–0.85)
Δ DRI	−0.22 (2.07) (−5.06–5.26)	−0.37 (2.29) (−4.00–5.26)	−0.16 (2.00) (−5.06–4.64)	0.21 (−0.78)	0.10	0.45 (0.32–0.58)

[a] Mann-Whitney Test.
** $p < .05$.

Table 3
Logistic regression with Inpatient Violence after measurement 5 as dependent variable.

Model 1	B (S.E.)	Wald	Df	Sig.	Exp(B)	95% C.I.for EXP(B)
DRI_5	0.492 (0.134)	13.571	1	0.000	1.636	1.259–2.126
Constant	−3.748 (0.818)	21.007	1	0.000	0.024	
R^2 = 0.242 (Nagelkerke), $\chi^2(1) = 17.484$, $p < .00$; %correct = 77.1%; HL-test: $\chi^2(8) = 9.516$, $p = .30$; −2LL = 94.660						

Model 2	B (S.E.)	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)
Δ DRI	−0.049 (0.113)	0.185	1	0.668	0.953	0.754–1.189
Constant	−1.004 (0.233)	18.602	1	0.000	0.367	
R^2 = 0.003 (Nagelkerke), $\chi^2(1) = 0.186$, $p = .67$; %correct = 72.9%; HL-test: $\chi^2(8) = 4.898$, $p = .77$; −2LL = 111.958						

Model 3	B (S.E.)	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)
DRI_5	0.537 (0.143)	14.125	1	0.000	1.711	1.293–2.264
Δ DRI	−0.154 (0.116)	1.753	1	0.186	0.857	0.682–1.077
Constant	−4.036 (0.878)	21.151	1	0.000	0.018	
R^2 = 0.264 (Nagelkerke), $\chi^2(2) = 19.265$, p < .00; %correct = 81.3%; HL-test: $\chi^2(8) = 8.097$, $p = .42$; −2LL = 92.879						

5. Discussion

The goal of this study was to determine the influence of inpatient treatment history in the prediction of short-term inpatient violence: Must we consider the change in criminogenic needs as measured by the DRI scale or is the last measurement alone sufficient? The hypothesis was: the last DRI measurement (DRI_5) to which the change between first and last DRI measurement (Δ DRI) is added has more predictive power than only the last measurement. This hypothesis accounts for the development a patient makes during treatment and takes the baseline level into account (Beggs & Grace, 2011; Olver, Nicholaichuk, Kinston, & Wong, 2014). The predictive power of the last measurement was acceptable (AUC = 0.73) and comparable to structured risk assessment instruments (Ramesh, Igoumenou, Vazquez Montes, & Fazel, 2018) but the hypothesis was not confirmed by this study. The change in DRI did not add to the predictive power of the last measurement. Although the odds ratio of DRI_5 increased when the change in DRI was considered (Exp(B) = 1.64 vs. Exp(B) = 1.71), there was no significant difference between the two models.

The change in DRI in the first three years of treatment was on average very low (M = −0.22, see Fig. 1), which almost implies that the IFTE is not measuring change and repeated measures are not beneficial. This is unexpected since most change is expected at the beginning of treatment (Wooditch, Tang, & Taxman, 2014). Inspecting the range of the change (Δ DRI: −5.06–5.26), some patients changed positively, others negatively and some remained stable resulting in a mean change of almost zero. Schuringa, Spreen, and Bogaerts (2018) showed that within a cross sectional selection of patients, a large group of patients had a low level on the factor Problematic behavior, which consists of the DRI and two extra indicators, which were not involved in this study (psychotic symptoms and sexual deviant behavior). This large group of patients with a low level of Problematic behavior does not need to change, anymore. In the current study, the mean and range of DRI at every measurement also showed that a group of patients already scored low on DRI and therefore did not need to change. This low level of DRI could be explained by the characteristics of the high-security institution. On the one hand the security-measures prevent patients from displaying problematic behavior. On the other hand, the institution supplies patients with all kind of means, like food, a bed, a shower, medical care, medication, mental support, meaningful daytime activities, and structure so that there is no 'need' to display DRI. Nonetheless, there is still a small group of patients displaying behaviors that score high on DRI and a group that commits inpatient violence. The large group who does not need to change anymore could be the reason why change in DRI does not contribute to the prediction of violence in this study.

5.1. Clinical implications

Typically, in risk assessment of inpatient violence, the last measurement is used (O'Shea et al., 2016), this procedure is validated by the results of this study. To keep the risk assessment up to date, continuous monitoring of dynamic risk indicators is recommended. By continuous monitoring who is and is not at risk, management measures can be deployed more efficient, interventions can be evaluated and adapted if necessary. Treatment effects are therefore closely monitored, which adheres to the responsivity principle of the RNR-model.

5.2. Strengths, limitations and future research

A strong characteristic of this study was the naturalistic way of data collection (ecological validity). The IFTE was filled out in everyday use in a treatment setting, by therapists who were involved in the treatment of the patient and was not scored by trained researchers based on file information. Therefore, these data represent real-life observations (Lens, Pemberton, & Bogaerts, 2013; Wilson et al., 2013). The treatment period was held relatively stable for all patients and at the beginning of the treatment period, in comparison to some pre- and post-treatment assessment studies, which can have irregular or unknown treatment periods (Beggs & Grace, 2011; De Vries Robbé et al., 2015).

The group of patients used in this study consisted of various diagnosis and crimes committed. Although an earlier study showed that the IFTE has predictive power for different kind of diagnostic patient groups (Schuringa et al., 2018). Maybe, patients who committed a non-violent crime prior to admission, like a sexual assault or theft, are less likely to commit inpatient violence. In future research, a group of patients should be selected which have shown to be aggressive in the past, for example by selecting patients with violent crimes.

To study the effect of change on criminogenic needs on inpatient violence a study which takes into (or, out of) account the large group that does not change, but also does not need to change would be beneficial. A way of doing this is using a cut-off to divide the group in a high risk and low risk group (Raynor, 2007; Schuringa et al., 2018). Patients can either start high or low and end high or low. This leads to four groups, a low-low group, a high-high group, a high-low group and a low-high group. Change on DRI might only be important if this change means that someone is transferring from one risk category to the other one. Patients within the same risk category could have different violence rates according to their change or lack of it. It could be possible that patients changing from low risk to high risk are less likely to

commit inpatient violence than patients who stayed high (Hochstetler et al., 2016). But it is also conceivable that for a patient who is at high risk and remains at high risk, sufficient risk management actions are already taken to prevent inpatient violence. While a patient changing from low risk to high risk is often not noticed by treatment teams (Kahneman, 2011), especially if they are not using ROM tools (Waller & Turner, 2016), so they pose a higher risk of inpatient violence than patient who were at risk the whole time.

6. Conclusion

The sum of dynamic risk indicators of the DRI is dynamic and has predictive power for short-term inpatient violence. The change in these indicators, however in this study, does not contribute to a more sophisticated prediction of short-term inpatient violence. The last measurement is the most efficient predictor for short-term inpatient violence, but because of the dynamic nature of these indicators it is necessary to frequently monitor these indicators to detect imminent risks.

Ethics Approval and consent to participate

No ethics committee approval was necessary since this study was performed retrospectively on electronic patient files and patients were not subjected to actions or rules of conduct imposed on them. Patients in this study did not have to give permission because the research was conducted on electronic files. Permission for this study was given by the director and committee of behaviorists of the institution.

Availability of data and materials

The datasets generated and analyzed during the current study are not publicly available due to the sensitivity of the data but are available from the corresponding author on reasonable request.

Declarations of interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

Not applicable.

References

Abderhalden, C., Needham, I., Dassen, T., Halfens, R., Haug, H.-J., & Fischer, J. E. (2008). Structured risk assessment and violence in acute psychiatric wards: randomised controlled trial. *The British Journal of Psychiatry, 193*, 44–50. https://doi.org/10.1192/bjp.bp.107.045534.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author text rev.

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy and Law, 16*, 39–55. https://doi.org/10.1037/a0018362.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19–52. https://doi.org/10.1177/0093854890017001004.

Beggs, S. M., & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: a comparative validity study. *Journal of Consulting and Clinical Psychology, 79*, 182–192. https://doi.org/10.1037/a0022900.

Bogaerts, S., Spreen, M., Ter Horst, P., & Gerlsma, C. (2018). Predictive validity of the HKT-R Risk Assessment for two and five-year recidivism in a cohort of Dutch Forensic Psychiatric Patients. *International Journal of offender Therapy and Comparative Criminology, 62*, 2259–2270. https://doi.org/10.1177/0306624X17717128.

Chagigiorgis, H., Michel, S. F., Seto, M. C., Laprade, K., & Ahmed, A. G. (2013). Assessing short-term, dynamic changes in risk: The predictive validity of the Brockville Risk Checklist. *International Journal of Forensic Mental Health, 12*, 274–286. https://doi.org/10.1080/14999013.2013.857740.

Cohen, T. H., Lowenkamp, C. T., & VanBeschaoten, S. W. (2016). Examining changes in offender risk characteristics and recidivism outcomes: A research summary. *Criminology & Public Policy, 15*, 263–296. https://doi.org/10.1111/1745-9133.12190.

Dack, C., Ross, J., Papadopoulos, C., Stewart, D., & Bowers, L. (2013). A review and meta-analysis of the patient factors associated with psychiatric in-patient aggression. *Acta Psychiatrica Scandinavica, 127*, 255–268. https://doi.org/10.1111/acps.12053.

Daffern, M., Jones, L., Howelss, K., Shine, J., Mikton, C., & Tunbridge, V. (2007). Editorial refining the definition of offence paralleling behavior. *Criminal Behaviour and Mental Health, 17*, 265–273. https://doi.org/10.1002/cbm.671.

De Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. (2015). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting reductions in post discharge community recidivism. *Law and Human Behavior, 39*, 53–61. https://doi.org/10.1037/lhb0000089.

Desmarais, S. L., Nicholls, T. L., Wilson, C. M., & Brink, J. (2012). Using dynamic risk and protective factors to predict inpatient aggression: Reliability and validity of START Assessments. *Psychological Assessment, 24*(3), 685–700. https://doi.org/10.1037/a0026668.

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.

Fleminger, S. (1997). Numbers needed to detain. *British Journal of Psychiatry, 171*, 287. https://doi.org/10.1192/bjp.171.3.287a.

French, S. A., & Gendrau, P. (2006). Reducing prison misconducts. *Criminal Justice and behavior, 33*, 185–218. https://doi.org/10.1177/0093854805284406.

Hildebrand, M., & de Ruiter, C. (2012). Psychopathic traits and change on indicators of dynamic risk factors during inpatient forensic psychiatric treatment. *International Journal of Law and Psychiatry, 35*, 276–288. https://doi.org/10.1016/j.ijlp.2012.04.001.

Hochstetler, A., Peters, D. J., & DeLisi, M. (2016). Classifying risk development and predicting parolee recidivism with growth mixture models. *American Journal of Criminal Justice, 41*, 602–620. https://doi.org/10.1007/s12103-015-9320-8.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

Jeandarme, I., Wittouck, C., Vander Laenen, F., Pouls, C., Oei, T. I., & Bogaerts, S. (2016). Risk factors associated with inpatient violence during medium security treatment. *Journal of Interpersonal Violence,* 1–26. https://doi.org/10.1177/0886260516670884.

Kahneman, D. (2011). *Thinking, fast and slow.* London: Penguin Books.

Labrecque, R. M., Smith, P., Lovins, B. K., & Latessa, E. J. (2014). The importance of reassessment: How changes in the LSI-R risk score can improve the prediction of recidivism. *Journal of Offender Rehabilitation, 53*, 116–128. https://doi.org/10.1080/10509674.2013.868389.

Lens, K. M. E., Pemberton, A., & Bogaerts, S. (2013). Heterogeneity in victim participation: A new perspective on delivering a Victim Impact Statement. *European Journal of Criminology, 10*, 479–495. https://doi.org/10.1177/1477370812469859.

Lewis, K., Olver, M. E., & Wong, S. C. (2013). The Violence Risk Scale: predictive validity and linking changes in risk with violent recidivism in a sample of high-risk offenders with psychopathic traits. *Assessment, 20*, 150–164. https://doi.org/10.1177/1073191112441242.

van Marle, H. J. C. (2002). The Dutch Entrustment Act (TBS): Its principles and innovations. *International Journal of Forensic Mental Health, 1*, 83–92. https://doi.org/10.1080/14999013.2002.10471163.

McGrath, A., & Thompson, A. P. (2012). The relative predictive validity of the static and dynamic domain scores in risk-need assessment of juvenile offenders. *Criminal Justice and Behavior, 39*, 250–263. https://doi.org/10.1177/0093854811431917.

Mooney, J. L., & Daffern, M. (2013). The offence analogue and offence reduction behaviour rating guide as a supplement to violence risk assessment in incarcerated offenders. *International Journal of Forensic Mental Health, 12*, 255–264. https://doi.org/10.1080/14999013.2013.867421.

Nagtegaal, M. H., Horste, R. P., & Schönberger, H. J. M. (2011). *Inzicht in de verblijfsduur van tbs-gestelden.* Insight in duration of tbs-order personsMeppel: Boom Juridische uitgevers.

Nijman, H. L. I., Muris, P., Merckelbach, H. L. G. J., Palmstierna, T., Wistedt, B., Vos, A. M., ... Allertz, W. (1999). The staff observation aggression scale–revised (SOAS-R). *Aggressive Behavior, 25*, 197–209. https://doi.org/10.1002/(SICI)1098-2337(1999)25:3<197::AID-AB4>3.0.CO;2-C.

O'Shea, L. E., & Dickens, G. L. (2015). Predictive validity of the Short Term Assessment of Risk and Treatability (START) for aggression and self-harm in a secure mental health service: Gender differences. *International Journal of Forensic Mental Health, 14*, 132–146. https://doi.org/10.1080/14999013.2015.1033112.

O'Shea, L. E., Picchioni, M. M., & Dickens, G. L. (2016). The predictive validity of the short-term assessment of risk and treatability (START) for multiple adverse outcomes in a secure psychiatric inpatient setting. *Assessment, 23*, 150–162. https://doi.org/10.1177/1073191115573301.

O'Shea, L. E., Picchioni, M. M., McCarthy, J., Mason, F. L., & Dickens, G. L. (2015). Predictive validity of the HCR-20 for inpatient aggression: the effect of intellectual disability on accuracy. *Journal of Intellectual Disability Research, 59*, 1042–1054. https://doi.org/10.1111/jir.12184.

Olver, M. E., Nicholaichuk, T. P., Kinston, D. A., & Wong, S. C. P. (2014). A multisite examination of sexual violence risk and therapeutic change. *Journal of Consulting and Clinical Psychology, 82*, 312–324.

Pieterse, H. (2015). *Richtsnoer voor good clinical practice (CPMP/ICH/135/95). Officiële Nederlandse vertaling [Guideline for good clinical practice (CPMP/ICH/135/95). Official Dutch translation].* Heerhugowaard: Profess Medical Consultancy.

Ramesh, T., Igoumenou, A., Vazquez Montes, M., & Fazel, S. (2018). Use of risk assessment instruments to predict violence in forensic psychiatric hospitals: a systematic review and meta-analysis. *European Psychiatry, 52*, 47–53. https://doi.org/10.1016/j.eurpsy.2018.02.007.

Raynor, P. (2007). Risk and need assessment in British probation: the contribution of LSI-R. *Psychology, Crime & Law, 13*, 125–138. https://doi.org/10.1080/10683160500337592.

Schuringa, E., Heininga, V. E., Spreen, M., & Bogaerts, S. (2016). Concurrent and predictive validity of the Instrument for Forensic Treatment Evaluation. *International Journal of Offender Therapy and Comparative Criminology, 62*, 1281–1299. https://doi.org/10.1177/0306624X16676100.

Schuringa, E., Spreen, M., & Bogaerts, S. (2014). Inter-rater and test-retest reliability, internal consistency, and factorial structure of the Instrument for Forensic Treatment Evaluation. *Journal of Forensic Psychology Practice, 14*, 124–144. https://doi.org/10.1080/15228932.2014.897536.

Schuringa, E., Spreen, M., & Bogaerts, S. (2018). Voorspellen van intramuraal geweld op korte termijn met het Instrument voor Forensische Behandel Evaluatie (IFBE), ROM-instrument in de tbs voor verschillende doelgroepen. [Predicting short term inpatient violence with the Instrument for Forensic Treatment Evaluation (IFTE), ROM-instrument in the tbs for different target groups]. *Tijdschrift voor Psychiatrie, 60*, 662–671.

Serin, R. C., Lloyd, C. D., Helmus, L., Derkzen, D. M., & Luong, D. (2013). Does intra-individual change predict offender recidivism? Searching for the holy grail in assessing offender change. *Aggression and Violent Behavior, 18*, 32–53. https://doi.org/10.1016/j.avb.2012.09.002.

Spreen, M., Brand, E., ter Horst, P., & Bogaerts, S. (2014). *Handleiding HKT-R. [Manual of the HKT-RJ, 2014, Stichting FPC Dr. S. van Mesdag; Groningen*.

Troquete, N. A. C., Van den Brink, R. H. S., Beintema, H., Mulder, T., van Os, T. W. D. P., Schoevers, R. A., & Wiersma, D. (2013). Risk assessment and shared care planning in out-patient forensic psychiatry: cluster randomised controlled trial. *British Journal of Psychiatry, 202*(5), 365–371. https://doi.org/10.1192/bjp.bp.112.113043.

Van der Veeken, F. C. A., Bogaerts, S., & Lucieer, J. (2018). The Instrument for Forensic Treatment Evaluation: Reliability, factorial structure, and sensitivity to measure behavioral changes. *Journal of Forensic Psychology Research and Practice, 18*(3), 229–253. https://doi.org/10.1080/24732850.2018.1468675.

Van der Veeken, F. C. A., Lucieer, J., & Bogaerts, S. (2016). Routine outcome monitoring and clinical decision-making in forensic psychiatry based on the Instrument for Forensic Treatment Evaluation. *PLoS ONE, 11*(8), e0160787. https://doi.org/10.1371/journal. pone.0160787.

Vess, J. (2001). Development and implementation of a functional skills measure for forensic psychiatric inpatients. *Journal of Forensic Psychiatry, 12*, 592–609. https://doi.org/10.1080/09585180110092001.

Waller, G., & Turner, H. (2016). Therapist drift redux: Why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Research and Therapy, 77*, 129–137. https://doi.org/10.1016/j.brat.2015.12.005.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Middleton, C. (2004). *Short Term Assessment of Risk and Treatability (START): An evaluation and planning guide.* Hamilton, Ontario–Port CoquitlamBritish Columbia: St. Joseph's Healthcare–Forensic Psychiatric Services Commission.

Whittington, R., Bjørngaard, J. H., Brown, A., Nathan, R., Noblett, S., & Quinn, B. (2014). Dynamic relationship between multiple START assessments and violent incidents over time: a prospective cohort study. *BMC Psychiatry, 14*, 1–7. https://doi.org/10.1186/s12888-014-0323-7.

Wilson, C. M., Desmarais, S. L., Nicholls, T. L., Hart, S. D., & Brink, J. (2013). Predictive validity of dynamic factors: Assessing violence risk in forensic psychiatric inpatients. *Law and Human Behavior, 37*, 377–388 10.1037lhb0000025.

Wooditch, A., Tang, L. L., & Taxman, F. S. (2014). Which criminogenic need changes are most important in promoting desistance from crime and substance use? *Criminal Justice and behavior, 41*, 276–299. https://doi.org/10.1177/0093854813503543.

World Medical Association (2013). World Medical Association: Ethical principles for medical research involving human subjects. *JAMA, 310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053.