AUGUST 8, 2018

UNIVERSITY OF TWENTE.

MASTER THESIS
VALIDATING THE HKT-R OVER TIME
A NEW PERSPECTIVE ON THE VALIDITY AND PREDICTIVE VALUE OF THE HISTORISCH KLINISCH TOEKOMST – REVISED

JONAS MÖLLER
Positive Psychology & Technology

Supervisor dr. P.M. ten Klooster

Second Supervisor dr. M.L. Noordzij

Abstract

Introduction. In the Netherlands, the Historisch Klinisch Toekomst – Revised (HKT-R) is one of the most popular instruments to assess the risk of recidivism into violent behavior among offenders with a psychiatric diagnosis. The HKT-R contains three scales including historical items that are unchangeable such as demographics, clinical and future-related items focused on the present and the future which are therefore changeable or so-called dynamic. In this study, the validity and predictive value of the HKT-R were tested in cross-sectional analyses as well as its' applicability to predict change in aggressive behavior over the course of a treatment period.

Methods. The study had a naturalistic prospective design. 50 patients in closed forensic centers were included in the study. Participants' risk of recidivating was assessed during admission, halfway through treatment and before discharge, using the HKT-R. Concurrently, aggressive behavior was assessed using the Social Dysfunction Aggression Scale-9 (SDAS-9). Correlational, linear regression analyses, computation of change scores and linear mixed model analyses were used to examine the validity and predictive value of the HKT-R.

Results. The clinical scale of the HKT-R was the only one to correlate significantly with SDAS-9 score for all measurements (r_1= .44, p= .001, r_2= .59, p< .001, r_3= .67, p< .001), showing adequate validity. In cross-sectional analyses the future-related scale only correlated significantly halfway through treatment and before discharge while the historical scale correlated only during admission significantly with the SDAS-9 score. The only significant correlation for change scores was found when the change score of the clinical scale for the entire treatment period was included (r_{1-3}= .38, p= .006). According to the linear mixed models, the clinical and the future-related scale covaried significantly with the SDAS-9 score over time.

Discussion. The clinical scale of the HKT-R displayed reasonable construct validity as well as some predictive value in predicting the SDAS-9 score. The future-related score should be examined in further research since it did not show stable construct or predictive validity. The historical scale showed moderate construct validity only during admission. In contrast to previous follow-up studies, the historical score was the weakest predictor for aggressive behavior in terms of the SDAS-9 score which might be explained through differences in the set-up of the studies. Additionally, this study indicated that the clinical scores corresponded to some kind of change in participants' behavior during treatment. Overall, this research does not support the suitability of the HKT-R to predict changes in aggressive behavior. On the other hand, it supports the idea that aggressive behavior in psychiatric patients might be reduced through treatment interventions and that this change might be, at least partly, displayed in the clinical scale.

Table of Contents

List of Tables

List of Abbreviations

ADHD	Attention Deficit Hyperactivity Disorder
ASPD	Antisocial Personality Disorder
AUC	Area Under Curve
DSM-IV/5	Diagnostic and Statistic Manual of Mental Disorders
FPC	Forensic Psychiatric Center
HCR-20	Historical Clinical Management-20
HKT-30	Historisch Klinisch Toekomst-30
HKT-R	Historisch Klinisch Toekomst-Revised
PCL-R	Psychopathy Checklist-Revised
PD	Personality Disorder
ROC	Receiver Operating Characteristics
SCJ	Structured Clinical Judgement
SDAS-9/11	Social Dysfunction and Aggression Scale
Tbs	terbeschikkingstelling

1. Introduction

In the past, research has shown a correlation between psychiatric disorders and violent offending (Fazel et al., 2009; Yu, Geddes, & Fazel, 2012). Among others, it is indicated that there might be a link between personality disorders and violence as well as anti-social behavior. Yu, Geddes and Fazel (2012) found that out of 9,578 individuals with a personality disorder (PD) 10.7% (1,024) had engaged in violent behavior while out of the general population sample only 1.2% had done so. Other research supported the finding that individuals with a PD are more likely to engage in violent behavior compared to individuals without a PD (Logan & Blackburn, 2009). However, the correlation differs depending on the kind of personality disorder. Noticeably, individuals diagnosed with an antisocial personality disorder (ASPD) were shown to be far more likely to commit violence compared to other PD's (Yu, Geddes, & Fazel, 2012). Also, Hiscoke et al. (2003) indicated that individuals with an ASPD are more likely to reoffend. Apart from personality disorders, some research indicates a correlation between psychotic disorders and violent behavior (Fazel et al., 2009). Still, Fazel et al. (2009) indicated that the correlation becomes explicitly weaker when adjusting for substance abuse. Research regarding recidivism in individuals with a psychotic disorder has offered mixed results (Bonta, Law, & Hanson, 1998; Yu & Fazel, 2011). Altogether, research has shown that psychiatric disorders, mainly personality disorders, are associated with violent behavior. Yet, research has not shown whether there is a causal relationship between the two (Gilbert, & Daffern, 2011). The motivation behind this study is to take a step towards the identification of factors that contribute causally to recidivism of violent behavior in offenders with psychiatric disorders. This shall be achieved by evaluating an assessment instrument throughout the course of a treatment period.

If there is the indication that a person committed a crime due to a psychiatric disorder a Dutch judge can consider to apply a so-called "terbeschikkingstelling" (tbs) sanction instead of a plain prison sentence or in addition to it (Art. 37a, Wetboek van Strafrecht, 2018). Tbs sanctions can only be imposed if certain premises are fulfilled. Most importantly, the offender must found to be impaired by a psychiatric disorder at the time of the conduct. Furthermore, the respective sentence for the committed offense must be equivalent to at least four years' imprisonment. Lastly, an expert witness must assess a high risk of recidivating into violent behavior (Philipse, 2005).

If imposed, offenders are randomly allocated to one of twelve closed forensic psychiatric centers in the Netherlands (FPC's; Harte, van Kalmhoudt, & Knüppe, 2010). Tbs sanctions are of indefinite duration. Every other year, a judge needs to either extend or end the sanction. His decision is based on risk of reoffending (Philipse, 2005).

The tbs sanction has two main objectives. The first, protecting the society, does not differ from usual prison sentences. Moreover, in FPC's, offenders are treated in order to prevent future recidivism (Harte, van Kalmhoudt, & Knüppe, 2010). Most offenders who are sentenced with a tbs sanction suffer from a psychotic disorder, a personality disorder, an autism spectrum disorder, a diminished intelligence or a combination of the above (Regioplan Beleidsonderzoek, 2006). Also, many have a substance or behavioral addiction (Harte, van Kalmhoudt, & Knüppe, 2010). When a judge is considering to extend the sanction there are a number of potential consequences he has to keep in mind. Often, the measure entails a long stay in an FPC which leads to high costs for the taxpayer. Additionally, there is a strong interference with the inmates' privacy. On the other hand, if and when an offender recidivates, it has a strong impact not only on the victim but also possibly on the society and politics. As a result, judges have to make a difficult and consequential decision. For this, they are supported by the advice of professionals such as psychologists or psychiatrists (Harte, & Breukink, 2010).

This emphasizes the important role of professionals in the assessment of an offender's mental state and the risk of recidivism. Not only do they advise a judge before the initial verdict, if a tbs sanction is applied, a risk assessment is to be performed at least once a year (Ministerie van Justitie, 2007). Especially, when considering to release offenders back into society risk assessments play a crucial role. De Kogel and den Hartogh (2005) showed that judges' decisions to end a tbs sanction are usually in accordance with the advice of professionals.

Altogether, that puts a lot of pressure on the recidivism risk assessment done by professionals. Traditionally, these assessments were only based on the subjective opinion the professional had formed (Lammers, 2007). This is particularly alarming because studies have shown that this manner of assessment is, on average, hardly more accurate than chance (Philipse, 2005; De Vogel, 2005). As a result, practitioners and scientists started to develop assessment instruments which are intended to predict the risk of recidivism among groups as well as individuals more objectively (Harte, & Breukink, 2010). Studies indicate that predictions on the basis of risk assessment tools are more accurate (De Vogel, 2004; De Vogel, 2005). Also, such structured instruments promote transparency and reduce bias (Harte, & Breukink, 2010) which led Hilterman (2001) to conclude that this way of judgement is more ethical in comparison to unstructured clinical judgements.

Today, there is a growing variety of instruments available with different features. Some predict recidivism indirectly because they are intended to indicate certain disorders that can be related to violent behavior like Hare's (1991) Psychopathy Checklist – Revised (PCL-R; Hare, et al., 2000). Others are designed to predict recidivism directly. One of the most popular tests is Webster et

al's (1997) Historical Clinical Risk Management-20 (HCR-20; Philipse et al., 1999). Based on this instrument, the Dutch Historisch Klinische Toekomst-30 (HKT-30; Werkgroep Risicotaxatie Forensische Psychiatrie, 2002) and its updated version Historisch Klinisch Toekomst-Revised (HKT-R; Bogaerts, et al., 2013) were developed.

Like the HCR-20 and the HKT-30, the HKT-R is composed of three distinct domains containing historical items, clinical items and future-related items. Historical items are focused on past events and behavior that occurred before the admission to the current facility (Spreen, et al., 2014). Therefore, the historical domain contains mostly items that describe static factors. Static factors are those that do not change over time such as age at first conviction or behavioral issues before twelfth birthday (Harte, & Breukink, 2010). Clinical items are scored with regard to the past year or the time span since the last assessment. Thus, they are focused on the behavior and events during the time of treatment. Lastly, there are future-related items that address the living situation of the patient after discharge as well as his skill-level in different, recidivism-related areas (Spreen et al., 2014).

Both, the clinical domain and the future-related domain contain items that describe mainly dynamic factors that, in contrast to static ones, are changeable such as attitudes or skill-levels (Harte, & Breukink, 2010). Those factors are particularly important to judge a patient's development during treatment. Yet, studies indicate that domains describing dynamic factors are not more accurate in predicting recidivism than domains with mainly static factors (Bogaerts, et al., 2017; Philipse, 2005). However, static factors cannot be used to assess a patient's development and thus the effect of treatments. Dynamic factors, on the contrary, can indicate a patient's progress during his treatment (Bonta, 1997). Therefore, they can be used to evaluate treatment interventions.

The HKT-R was validated in a study assessing the risk of recidivism among 347 former tbs patients who were released between 2004 and 2008. Predictive validity was determined using different methods. In ROC analyses, AUC values for the total score were .71 for recidivism after two years and .77 after five years (Bogaerts et al., 2013). These results indicate a reasonable predictive validity according to Sjöstedt & Grann (2002). Also, Bogaerts et al. (2013) stated that these AUC values are somewhat better than those found for other instruments by Singh et al. (2011). Furthermore, t-tests were applied to compare the mean score of recidivists with those of not-recidivists. The results showed that average domain scores as well as the total score differed significantly between the two groups. Among the different domains, the historical domain predicted recidivism better than the clinical as well as the future domain on group level (Bogaerts et al., 2013).

The biggest issue in today's recidivism treatment is that there is no theoretical rationale for it despite the growing variety of research in recent years (Bonta, Law, & Hanson, 1998; Canton et al.,

2004; Philipse, 2005; Hildebrand, et al., 2005; Fazel & Yu, 2009). Tbs treatment entails a number of different interventions including psychoeducation, skill trainings and societal work. However, tbs treatment needs to be more scientifically based (van Nieuwenhuizen et al., 2011). The usage of instruments that focus on dynamic factors to evaluate tbs treatments could help to identify underlying factors that have a causal relationship with recidivism. Thus, the assessment could help to form an explanation why (former) tbs patients recidivate. Understanding the mechanisms behind recidivism can go a long way to form interventions that address these mechanisms in order to reduce the risk of recidivism (Bogaerts et al., 2017).

The omission of a theoretical rationale for underlying causal correlations of violent recidivism can be ascribed to various issues associated with this kind of research. One issue here is that there is not one universal definition of recidivism. Some researchers operationalize recidivism as new conviction (Hildebrand et al., 2005), others as new encounters with the police or merely being mentioned as a suspect in a report (Canton et al., 2004). Jeandarme et al. (2015) showed impressively that the definition of recidivism strongly influences the outcome of the study. They found that out of 192 former psychiatric patients fifteen (7.81%) would be marked as recidivists in terms of a new conviction. This number increased to 75 (39.06%) when potentially indictable incidents were taken into account. As a result, it is difficult to make a meaningful comparison between studies which operationalized the construct differently. Another methodological obstacle is the choice between prospective and retrospective research designs. In retrospective studies, researchers analyze existing data and recidivism rates of tbs patients who have already been released. Often, data collected during tbs treatment is not very accurate or incomplete because it was not collected under research conditions (Harte, & Breukink, 2010). An advantage is that, if the data is available, the design is easy to implement. As a result, it is the most common type of research in risk assessment research. Prospective studies, on the other hand, are more sophisticated and more costly to implement because researchers collect the data of current tbs patients. Years after discharge, their recidivism data is collected and analyzed. However, it offers the advantage that researchers are able to collect the data very accurately. Moreover, at the point of data collection they are not biased because they have no knowledge of eventual recidivism rates. This is not necessarily the case for retrospective study designs.

An issue that impacts both research designs is the fact that their focus is only on the patients who are released into society. For obvious reasons, (usually) only patients who are predicted not to recidivate are released into society. Thus, there is a selection bias (Harte, & Breukink, 2010). Research has generally just examined the rate of so-called false negatives, patients predicted not to recidivate but do so, but not the rate of false positives which are the patients who are predicted to

recidivate but do not (Nijman, & Bulten, 2006). This is a reason why often results cannot be interpreted unambiguously as stated by Bregman and Wartna (2011). In their research, they found a strong significant decline of recidivism rates among former tbs patients since 1974. While among tbs patients released between 1984 and 1988 the recidivism rate for serious crimes was 36.4% after two years, the rate of those released between 2004 and 2008 was 17%. This trend was not only found for short-term recidivism, but there is also an indication that the long-term recidivism rate has decreased dramatically, as well. However, Bregman and Wartna (2011) concluded that their findings are no prove of effective tbs measures. Rather, the on average longer tbs sanctions as well as the increasing number of longstay patients suggest that the decrease of recidivism rates can be attributed to selection bias.

In this study, the validity of the HKT-R is examined in a different context. For the analyses, aggressive behavior is operationalized in terms of the Social Dysfunction and Aggression Scale-9 (SDAS-9) and used as a proxy for recidivism. In order to eliminate selection bias, this research is focused on aggressive behavior during the treatment period. However, it needs to be noted that this leads to a few other difficulties because it cannot be expected that violent behavior is the same during the treatment period under supervision as after discharge (Jeandarme et al., 2015). Contrary to existing literature, correlation between the domains of HKT-R and SDAS-9 are not only analyzed in a cross-sectional manner. Additionally, it is tested whether the HKT-R scores change significantly over time and whether these potential changes correlate to changes in aggressive behavior as measured by the SDAS-9. This is designed to indicate the predictive validity of the HKT-R over time and its usefulness to evaluate treatment interventions.

2. Methods

2.1 Participants

Participants were patients with different psychiatric disorders (N= 50) allocated to a closed psychiatric hospital in the Netherlands. All were diagnosed by professionals according to the DSM-IV (American Psychiatric Association, 2000) or DSM-5 (American Psychiatric Association, 2013). Participants were convicted for a crime at least once. However, no information regarding the index delict was available.

2.2 Materials

2.2.1 HKT-R

The Historisch Klinische Toekomst-Revised (HKT-R) is the updated version of the HKT-30, a popular instrument to assess recidivism risk among Dutch psychiatric patients who have been convicted for a violent offense. The structure of the test is largely based on the HCR-20 (Harte, & Breukink, 2010) and composed of three distinct domains which contain twelve historical, fourteen clinical and seven future-related items. The items are scored by professionals on a five-point Likert scale ranging from zero (= low risk) to four (= high risk). The HKT-R is developed to make a structured clinical judgement (SCJ). Thus in practice, the total score of the test cannot be used as sole predictor to assess a patient's risk to recidivate. The test is completed by two independent professionals who determine the total score by comparing their results in a so-called consensus dialog. The total score can be categorized as either low (0-42), moderate (43-54) or high (55 or higher). In order to form an SCJ, the total score, the patient's development and his individual context are taken into account. The SCJ is reflected in one of five categories: Low risk, low to moderate risk, moderate risk, moderate to high risk or high risk. Results on the basis of the HKT-R are valid for a year. After that, the test needs to be reapplied (Spreen et al., 2014).

Recently, Bogaerts et al. (2017) showed a reasonable predictive validity for the total score (AUC = .78), the SCJ (.78), the historical domain (.75) and the future-related domain (.71) as well as a marginal predictive validity for the clinical domain (.62 at admission; .63 at discharge) for recidivism after two years. Those numbers dropped for recidivism after five years. The historical domain remained reasonable (.74), while total score (.68), SCJ (.63) and clinical domain (.69 at admission; .62 at discharge) were marginal predictors. The future domain had low predictive value (.58). Past research indicates that the interrater reliability varied over domains but was shown to be moderate overall (ICC = .62). Internal consistency for all domains was shown to be good (Bogaerts et al., 2013).

2.2.2 SDAS-9

The Social Dysfunction and Aggression Scale-11 (SDAS-11; Wistedt et al., 1990) was used to measure aggressive behavior among participants. The instrument consists of two scales, one measuring outward aggression with nine items (SDAS-9). The other is applied to measure inward aggression with two items (SDAS-2). In this research, only the first scale was used. It is therefore referred to as SDAS-9. The SDAS-9 is scored on a five-point Likert scale which ranges from zero (= no incidents) to four (= very severe incidents). If a participant scored an eleven or higher on the outward aggression scale that is regarded as high aggressiveness. The test was administered by nurses to avoid possible source of error that is associated with self-rating. Wistedt et al. (1990) have found a reasonable inter-

observer reliability as well as good internal consistency for the outward scale. Moreover, they showed good convergent validity with two other scales, the Global Aggression Scale (GAS; r= .84, p< .001) and the Three-item Outward Aggression Scale (TOAS; r= .83, p< .001). Divergent validity was displayed through the inter-correlation with instruments measuring inward aggression. The SDAS-9 showed negative correlations with the SDAS-2 (r= -.23, p< .05) and the Three-item Inward Aggression Scale (TIAS; r= -.13).

2.3 Procedure

Inclusion criteria for participants was a diagnosis according to the DSM-IV (American Psychiatric Association, 2000) or DSM-5 (American Psychiatric Association, 2013) determined by a professional as well as at least one conviction for a criminal offense. Furthermore, patients with less than three completed measurements of the HKT-R or SDAS-9 test were excluded. The study had a naturalistic prospective research design. Patient ID's have been removed and substituted with participant numbers. As a result, the data is anonymized. The study was approved by a medical research ethics committee.

The data was collected by health professionals during routine check-ups within the facility between 2012 and 2017. The HKT-R was scored three times over the course of the treatment. It was taken by nurses and therapists, independently, during the patient's admission, midway through treatment and during discharge. The consensus score was determined in multidisciplinary meetings.

The SDAS-9 scores were assessed weekly. Average scores for the admission period (until four weeks after admission), midway through treatment (from two weeks before until two weeks after HKT-R was assessed) and discharge period (starting four weeks before HKT-R) were calculated. The basis for the score was the patient's everyday behavior on the ward. The consensus scores of the HKT-R and the mean scores of the SDAS-9 were used for the statistical analysis.

2.4 Statistical Analysis

Descriptive analyses were applied to show the participants' demographics as well as test outcomes. Cohen's d (1992) was calculated as effect size to determine the standardized difference between the test scores. The formula used to calculate Cohen's d was $d= (m_1 - m_2) / \sigma$. The effect size was categorized as .20 = small, .50 = medium or .80 = large (Cohen, 1992). Correlation between scores on each scale of the HKT-R and SDAS-9 scores were analyzed using Pearson Correlations. Pearson's correlation coefficient were categorized as < .30 = weak, .30 - .50 = moderate and > .50 = strong according to Cohen (1992). Linear regression was applied in order to predict the SDAS-9 scores on

the basis of the HKT-R scores. This was done by first analyzing each scale separately and then integrating them into a multivariate model to predict SDAS-9 scores, simultaneously.

Change scores for successive measurements and change from measurement one (admission) to three (discharge) were computed for all scales of the HKT-R as well as the SDAS-9. Then, Pearson Correlations were assessed between change scores of the HKT-R scales and change score of the SDAS-9. Linear regression was applied using the change scores of the HKT-R to predict the change scores of the SDAS-9. First, HKT-R scales were used as sole predictor and then integrated into one model, simultaneously.

Correlation between time point measurements of the SDAS-9 were assessed using Pearson Correlation. Moreover, the variance of scores for each measurement was calculated. The results were used to choose a covariance type for the linear mixed models. The model fit was confirmed using the -2 Log Likelihood test. Scores of each scale were inserted into the model to determine whether scores significantly changed with time entered as a fixed factor. Then, HKT-R scales were used as fixed covariates to determine the SDAS-9 score over time. Thereby, the advantage of mixed models in contrast to ANOVA's or simple regression analyses is the fact that autocorrelation is taken into account.

3. Results

Participants were predominantly male patients (n=44, 88%) with a mean age of 42.0 years (SD=10.88, range 51.6 years). Most patients were diagnosed with a psychotic disorder (n=16, 32%), autism spectrum disorder (n=12, 24%) or personality disorder (n=8, 16%). Other diagnoses included anxiety disorders (n=5, 10%), mood disorders (n=3, 6%), sexuality or gender identity disorders (n=3, 6%), ADHD (n=2, 4%) or somatoform disorder (n=1, 2%). Mean length of treatment was 709.02 days (SD=469.46, range 2693 days).

Mean scores of the SDAS-9 as well as the clinical scale and the future-related scale decreased throughout all three measurements (see Table 1). However, the mean future-related score showed the only notable decrease (d = .56). The decrease of mean SDAS-9 (d= .11) score as well as mean clinical score (d= .09) were small. On the historical domain, only one score by one patient changed across the three measurements. This can be accounted to filling in missing information and as such it was considered as no change for further analyses. As a result, there was no change for all H-scores. Lastly, when calculating the mean total scores it is evident that mean total scores were right around the cut-off score (55 points or higher) for high risk of recidivating.

Table 1

Mean scores SDAS-9 and HKT-R

	Mean scores measurement 1 (SD)	Mean scores measurement 2 (SD)	Mean scores measurement 3 (SD)
SDAS-9	2.35 (2.54)	2.14 (2.2)	1.77 (2.84)
HKT-R: Historical Scale	19.23 (9.91)	19.38 (10.08)	19.38 (10.08)
HKT-R: Clinical Scale	21.16 (7.75)	20.26 (7.57)	19.52 (9.56)
HKT-R: Future-related Scale	18.5 (5.49)	17.18 (5.37)	15.38 (5.7)
HKT-R: Total score	58.89 (18.12)	56.82 (18.27)	54.28 (19.76)

Note. SD= Standard deviation, all values for N= 50

The correlation between the historical scores and the SDAS-9 scores was moderate during the first time point (r= .36, p= .011). During the other time points, correlation between historical scores and SDAS-9 scores were weak and insignificant (p> .05; see Table 2).

The clinical scores (K_1) were moderately related to the SDAS-9 ($SDAS_1$) scores for the first measurement. The strength of correlation increased from r= .44, p= .011 for the first time point to r= .67, p< .001, a strong correlation for the third measurement of the clinical scores (K_3) and the SDAS-9 scores ($SDAS_3$).

There was no significant correlation between the first measurements of the future-related scores (T_1) and the SDAS-9 scores ($SDAS_1$; r= .25, p= .078). The results of T_2 - $SDAS_2$ displayed the strongest correlation between the two scales (r= .53, p< .001). The correlational strength of T_3 - $SDAS_3$ (r= .45, p= .001) decreased but remained stronger than T_1 - $SDAS_1$. Overall, the clinical scores showed the strongest correlations to the SDAS-9 scores. With K_3 - $SDAS_2$ (r= .69, p< .001), the single strongest correlation was displayed between (timely) non-corresponding scores.

Table 2

Pearson Correlations HKT-R scores – SDAS-9 scores for all measurements

	$SDAS_1$ scores	$SDAS_2$ scores	$SDAS_3$ scores
H_1 scores	.36*	.25	.16
H_2 scores	.36*	.25	.15
H_3 scores	.36*	.25	.15
K_1 scores	.44**	.56**	.40**
K_2 scores	.49**	.59**	.50**
K_3 scores	.44**	.69**	.67**
T_1 scores	.25	.51**	.40**
T_2 scores	.30*	.53**	.40**
T_3 scores	.24	.65**	.46**

*Note. *= significant for α= .05, **= significant for α= .01; all values for N= 50*

When integrated into a multiple linear regression model, the SDAS-9 scores could be predicted on the basis of the corresponding HKT-R scores for all three measurements. Moreover, for each time point, the K score was the only independent predictor of the SDAS-9 score. However, during the first measurement H_1 almost reached significance for an α= .05 level with p= .062 and T_2 almost reached significance during the second measurement with p= .069. These results resembled the trends of the correlational analyses where H_1 correlated significantly with $SDAS_1$ and T_2 with $SDAS_2$. Also, the fit of the regression model with the data increased steadily from time point one ($F(3,46)= 5.23$, p= .003) with an adjusted R^2 of .21, to time point two ($F(3,46)= 9.97$, p< .001) with an adjusted R^2 of .36, to time point three ($F(3,46)= 12.95$, p< .001) with an adjusted R^2 of .42. Also, $SDAS_3$ could be predicted on the basis of HKT_1 but model fit was noticeably weaker ($F(3,46)= 4.16$, p= .011) with an adjusted R^2 of .16. Moreover, none of the variables was a significant independent predictor in this model.

Change scores were computed for the SDAS-9 as well as the clinical and future-related scores in order to examine the predictive validity of the HKT-R over time. No change scores of the historical scale were included since there was no change in participants' scores. The only significant correlation between corresponding change scores of the clinical scale and the SDAS-9 change scores was the correlation for the whole treatment period (r= .38, p= .006). However, it is noteworthy that the correlation between the change score of the second and third measurement of the clinical scores (K_{2-3}) and the change score of the first and third measurement of the SDAS-9 scores ($SDAS_{1-3}$) was even stronger (r= .43, p= .002). There was no significant correlation between the change scores of the future-related scale and change scores of the SDAS-9. Moreover, many correlations were slightly

negative but the results were highly insignificant (see Table 3) indicating that there is rather no correlation in the population than a negative correlation.

Table 3

Correlation Change Scores Clinical and Future-related scale – Change Scores SDAS-9

	$SDAS_{1-2}$ scores	$SDAS_{2-3}$ scores	$SDAS_{1-3}$ scores
K_{1-2} scores	-.03	.11	.07
K_{2-3} scores	.28	.22	.43**
K_{1-3} scores	.18	.27	.38**
T_{1-2} scores	-.07	-.05	-.10
T_{2-3} scores	.30*	-.06	.20
T_{1-3} scores	.20	-.08	.10
$SDAS_{1-2}$ scores	1	-.32*	.58**
$SDAS_{2-3}$ scores	-.32*	1	.59**
$SDAS_{1-3}$ scores	.58**	.59**	1

*Note. *= significant for α= .05, **= significant for α= .01; all values for N=50*

When the clinical and the future-related change scores were integrated as predictors into a multiple linear regression model, $SDAS_{1-3}$ was predicted on the basis of K_{1-3} and T_{1-3}. A significant regression equation was found ($F(1,48)= 4.08$, $p= .023$) with an adjusted R^2 of .11. However, K_{1-3} was the only independent predictor while T_{1-3} was highly insignificant (see Table 4). Again, these results resemble the previous ones as K_{1-3} correlated significantly with $SDAS_{1-3}$ while T_{1-3} did not (see Table 3).

Table 4

Regression Analysis with K_{1-3} and T_{1-3} as Predictor of $SDAS_{1-3}$: Coefficients

	B	SE	Beta	Sig.
Intercept	.27	.45		.561
K_{1-3}	.14	.05	.38	.008
T_{1-3}	.03	.09	.04	.776

Note. Dependent variable: $SDAS_{1-3}$.

In order to choose a linear mixed model, correlations between successive measurements and their variance were assessed. Correlation between the measurements one and two of the SDAS-9 score ($r= .54$, $p<.001$) differed from correlation between two and three ($r= .6$, $p<.001$). Moreover, variance differed between the three measurements ($s^2_1= 6.46$, $s^2_2= 4.84$, $s^2_3=8.08$). In clinical and future-related scores, the differences were similar. Therefore, an unstructured linear mixed model

was chosen. After comparison, a model with random intercept displayed a superior model fit towards a model including a random intercept as well as a random slope. As a result, the former model was used. Only when the future-related domain was used as covariate for the SDAS-9 score both models showed similar fit according to the -2 Log Likelihood test.

Within a linear mixed model, time, as defined by the change of scores between the three measurements, was not a significant factor for the SDAS-9 scores ($F(2,50)= 1.21$, $p=.307$). Thus, according to the model, the SDAS-9 scores did not change significantly over time. As viewed in Table 1, there was no change in the historical scores. However, when entered into a linear mixed model time was also not a significant factor for the clinical scores ($F(2,50)= 1.38$, $p= .259$). Thus, according to the model, the clinical scores also did not change significantly over time. Time was a significant factor for the future-related scores ($F(2,50)=14.49$, $p< .001$). Thus, according to the model, the future-related scores decreased significantly over time. These results fit with those of Table 1 and the previously calculated Cohen's d which showed that the future-related scores were the only scores to decrease significantly.

When entered as a covariate, the clinical scores covaried significantly with the SDAS-9 scores ($F(1,91.73)= 50.54$, $p<.001$). More exactly, it was indicated that, on average, a change of one score in the clinical domain (possible maximum of 56 scores) of the HKT-R covaries with a change of .16 scores in the SDAS-9 (possible maximum of 36 scores). The estimates of fixed effects are depicted in Table 5.

Table 5

Mixed Model for Clinical Score as Covariate of SDAS-9: Estimates of Fixed Effects

						95% Confidence Interval	
Parameter	Estimate	SE	df	t	Sig.	Lower Bound	Upper Bound
Intercept	-1.19	.51	80.73	-2.34	.022	-2.20	-.17
K scores	.16	.02	91.73	7.11	.000	.11	.21

Note. Dependent Variable: SDAS-9 scores

Also, a change of the future-related scores (possible maximum of 28 scores) covaried significantly with a change in the SDAS-9 scores ($F(1,99.38)=16.98$, $B= .16$, $p<.001$) when analyzed in a separate model. The results lost their significance for an $\alpha<.05$ level when a random slope was added to the model ($F(1,1.72)=16.34$, $B= .12$, $p=.072$). This was also true when the clinical scale and the future-related scale were included as covariates into one model, simultaneously. While predictive value of the future-related domain was reduced ($F(1,111.76)=2.42$, $B= .06$, $p=.123$) the clinical domain remained a significant covariate ($F(1,119.57)=27.28$, $B=.14$, $p<.001$). When all three scales of

the HKT-R were included as covariates for the SDAS-9 the numbers did not change a lot. As expected, the historical scale remained highly insignificant (see Table 6).

Table 6

Mixed Model of all HKT-R Scores as Covariates of SDAS-9: Estimates of Fixed Effects

						95% Confidence Interval	
Parameter	Estimate	SE	df	t	Sig.	Lower Bound	Upper Bound
Intercept	-1.85	.66	66.70	-2.81	.006	-3.16	-.53
T scores	.06	.04	112.62	1.42	.159	-.03	.14
K scores	.14	.03	121.55	5.05	.000	.08	.20
H scores	.01	.02	52.03	.48	.634	-.04	.06

Note. Dependent Variable: SDAS-9 scores

An important note is that there were not enough iterations to achieve convergence for the linear mixed models which included the future-related score as covariate. This was attributed to the small sample size (N= 50) and the relative large number of parameters.

4. Discussion

Construct validity seemed to vary between the different scales of the HKT-R. The correlation between the clinical scale and the SDAS-9 score indicated reasonable validity. The historical score correlated considerably weaker with the SDAS-9 score. Moreover, the correlation was only significant for the first time point. Therefore, it was indicated that the historical scale possesses insufficient construct validity. The future-related scale needs further analyses as it displays moderate construct validity during time points two and three but not during time point one. Therefore, it was not shown to have stable construct validity.

All analyses showed that there was no change in the historical score across all three measurements confirming that the scale indeed contains only static items. However, this is hardly a surprise. Rather, it is interesting that the predictive validity of the historical scale decreased as the treatment period progressed. This does not necessarily prove that the treatment interventions caused a change within the participants regarding their aggressiveness. Still, it is an indicator that something changed during the treatment period that impacted the historical scale‘s predictive strength of the SDAS-9 score. There is not a lot of research evaluating treatment interventions and their effect on aggression reduction in psychiatric patients which could be compared to this study. Moreover, the studies that are available offer mixed results (Needham, 2004; Abderhalden et al., 2008; Fluttert et al., 2010). While Abderhalden et al. (2008) showed a reduction of aggressive

incidences in a three month period by applying aggression assessments for each patient, Needham et al. (2004) indicated that there was no significant change in participants' aggression after systematic risk assessments as well as training courses for aggression management.

The clinical scale displayed the strongest correlation as well as the strongest predictive value for the SDAS-9 score. In previous, retrospective, follow-up studies the historical scale was indicated to be the best predictor of violent recidivism (Bogaerts, et al., 2013). This discrepancy might be best explained by their differences in study design and operationalization of recidivism. While this study is focused on aggression during the treatment which could be defined as short-term aggressive behavior, Bogaerts et al's (2013) research measures long-term effects.

Moreover, the statistical power of the clinical score as a predictor improved steadily from measurement one to three. This indicates a change in participants' cognitions which is displayed in participants' clinical score and somehow corresponds to the SDAS-9 score. This is only partly supported by the change score regressions and the mixed model analyses. The mixed model analysis showed that the change in clinical scores was not significant but did covary positively with the SDAS-9 scores.

The future-related score was the only one to change significantly over time. Still, no change score correlated significantly with the corresponding SDAS-9 change score. Also, all analyses relying on the first measurement of the future-related scale were not significant. Significance was reached in the regression analysis for time points two and three. One could speculate that future-related items become more important over treatment time and therefore have more predictive value toward the discharge of a patient. However, this would be subject of research on item-level and/or qualitative research. Moreover, the future-related score did not remain significant when the clinical score was also included in the model, simultaneously. These results showed that the future-related score was no independent predictor of the SDAS-9 score.

Altogether, this research indicates that the HKT-R is rather limited in its ability to predict the SDAS-9 score as an operationalization of aggressive behavior. This notion is supported by the large discrepancies between mean SDAS-9 scores and the mean total scores of the HKT-R. While the mean SDAS-9 scores remain far below the cut-off score for high aggressiveness during all time points, the mean total scores of the HKT-R exceed the cut-off score for high risk of aggressive recidivism for two of three time points. Still, these discrepancies are only indicators since mean scores, especially with large standard deviations, cannot be used to give a definitive conclusion. Moreover, while the linear mixed models showed some promise that HKT-R could predict the SDAS-9 score over time, the change score analyses did the exact opposite. The SDAS-9 change score could only be predicted on

the basis of the clinical score and this was only the case when the analysis was applied to the entire treatment period. None of the other correlations was significant indicating that the HKT-R is not suitable to predict the SDAS-9 score over time. Still, these results might be impacted by the fact that there was no significant change in clinical scores as well as SDAS-9 scores over time. The linear mixed model analysis of SDAS-9 with the clinical score as covariate indicates that some change in SDAS-9 score might be predicted by change in the clinical score. However, even in a bigger sample size with significant change in the clinical as well as the SDAS-9 score over time, the predicted portion would presumably be small.

On the other hand, it is indicated that some change of cognition and aggressive behavior had taken place and that the clinical scale shows some usefulness to reflect this change. Moreover, the results can give new insights to a more philosophical topic. In past research, the successfulness of static items in predicting recidivism among this population has suggested that aggressive behavior can be predicted through unchangeable facts (Jeandarme et al., 2015). This logic diminishes the approach that aggressive behavior in psychiatric patients can be reduced through treatment interventions. The findings of this research suggest that change in aggressive behavior can occur which is better predicted through dynamic items than static ones. Although the results are merely indicators, they support the idea that aggressive behavior in psychiatric patients can be reduced independent of past behavior or events.

The main limitation of this research is its small sample size (N=50). This impacts the power of all analyses as well as the external validity of their results. As stated in the result section, there were not enough iterations for the algorithm to converge in analyses including the future-related scale as covariate. This is an indicator that the range of potential models could not be exhausted due to a small sample size which can lead to biased results (Gueorguieva & Krystal, 2004). Moreover, the average SDAS-9 score was very low among this sample. The mean score ranges from 2.35 during measurement one to 1.77 during measurement three. Considering the cut-off score of 11 points for high aggressiveness, these mean scores suggest a floor effect (Bortz & Döring, 2005). There are different possible explanations: First of all, the SDAS-9 might not be an adequate indicator of aggressive behavior. Other possibilities are biases in the assessment of the SDAS-9 score, an unrepresentative sample or the effect that patients behave significantly less aggressive the more they feel under supervision (Jeandarme et al., 2015). Regardless, a floor effect could explain the insignificance of change in the SDAS-9 score and therefore the inability to predict SDAS-9 on the basis of HKT-R accurately over time.

On the other side, there are a number of advantages and strengths associated with this research. First of all, the naturalistic prospective design of the study is superior to conventional

retrospective designs that are commonly used. Moreover, HKT-R scores were assessed by therapists and nurses independently. Then, a consensus score was determined in multi-disciplinary meetings. The SDAS-9 scores, on the other hand, were assessed by nurses. This way of assessment reduced the risk of self-confirmation bias. Additionally, aggressive behavior is measured during the treatment period instead of thereafter. Although this comes with the disadvantage that patients under supervision might act differently, it offers some advantages as well. Behavior can be monitored more closely than in a follow-up period where patients report only occasionally or where recidivism data is only reported if a judicial authority is involved (Jeandarme et al., 2015). Often, this leads to inaccurate recidivism data. Moreover, when recidivism rates are reported for incidences that occurred in a follow-up period, only patients who were released into society are subject of research. This selection bias described by various researchers (Nijman, & Bulten, 2006; Harte, & Breukink, 2010) is eliminated when aggressive behavior is monitored during the treatment period where all participants are included.

Most importantly, this research adds a new perspective to the current literature where tests are usually administered once to predict recidivism rates in a follow-up period (Hildebrand et al., 2005). As described above, this methodological approach does not take into account the change that can occur within a patient during treatment. This is also not in accordance with the daily practice where tests need to be applied recurrently (Bogaerts et al., 2013). In the manual of the HKT-R it is stated explicitly that professionals are instructed to take into account the patients development among other things to form a structured clinical judgement (Bogaerts et al., 2013). Although this research indicates that the predictive validity derived from the change in participants‘ scores is limited, it has not only methodological but also philosophical importance. Applying treatment to reduce aggressive and violent behavior only makes sense if it is assumed that these behaviors can be reduced or changed.

In line with that, it is important for future research to further explore the focus on change during treatment periods by, for instance, combining quantitative with qualitative measurements such as interviews. This is important for the evaluation of the HKT-R as well as other assessment instruments of aggressive behavior and recidivism in psychiatric patients. Also, research should combine short-term measurements during treatment with long-term measurements in follow-up periods in order to examine the extent treatment interventions are able to change or reduce aggressive behavior in the short- as well as long-term among psychiatric patients.

Correlation and regression analyses do not take autocorrelation into account. With the linear mixed model analyses, it was possible to account for this bias. However, all analyses were impaired by a small sample size. With a bigger sample size, more complex and potentially more accurate

models could be applied. For future research, it is also important to analyze the HKT-R on the item level. This way, it is possible to assess more psychometric properties in order to make a more complete evaluation of the HKT-R. For instance, item level factor analysis might help to identify underlying factors that are causally connected to violent behavior and recidivism. Since the clinical scale has shown some promise in its predictive validity it would be interesting to evaluate the scale on item level, in particular. The goal should be to identify items that have particular high predictive value as well as exploring ways to further improve the scale. Moreover, in all analyses, the future-related scale lost its significance as predictor of the SDAS-9 when the clinical scale was added to the model, simultaneously. Therefore, it might be interesting to compare both scales on item-level to examine whether items of the future-related scale are potentially only manifestations of factors that are formulated as items on the clinical scale. Lastly, it is important to validate the HKT-R against other instruments and with other samples. The low SDAS-9 mean scores suggest a floor effect which would have impacted the results strongly. A replication of those results is therefore necessary.

The analyses of change and development of psychiatric patients prone to violent recidivism is an important topic that demands further research. However, this research does not support the suitability of the HKT-R as preferred instrument to assess risk of violent recidivism among psychiatric patients. The clinical scale showed adequate construct and predictive validity while the historical and future-related scale did not show consistent construct or predictive validity. Therefore, it is recommended to rather use the HKT-R as a complementary instrument than as sole predictor of violent and aggressive behavior. Still, results of this research must be taken with caution due to a possible floor effect in the SDAS-9 measures. Moreover, in this study total test scores were used as outcome measures while in daily practice those scores only make a part of the structured clinical judgement that is the basis for decisions. Therefore, it is difficult to compare study results with the daily practice.

References

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

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders: DSM -IV-TR. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Artikel 37a, Wetboek van het Strafrecht. (2018). Retrieved on 02-03-2018 from http://wetten.overheid.nl/BWBR0001854/2018-01-01#BoekEerste_TiteldeelIIA_AfdelingTweede_Artikel37a

Bogaerts, S., Spreen, M., ter Horst, P., & Gerlsma, C. (2017). Predictive validity of the HKT-R Risk Assessment Tool: 2 and 5-year violent recidivism in a nationwide sample of Dutch forensic psychiatric patients. *International Journal of Offender Therapy and Comparative Criminology, 1-12*.

Bogaerts, S., Willems, M., Spreen, M., Schuringa, E. & Ter Horst, P. (2013). Van HKT-30 naar HKT-R. Tilburg: Tilburg University Press.

Bonta, J. (1997). Do we need theory for offender risk assessment? *Forum on Corrections Research, 9, 42-45*.

Bonta, J., Law, M., & Hanson, K. (1998). The Prediction of Criminal and Violent Recidivism Among Mentally Disordered Offenders: A Meta-Analysis. *Psychological Bulletin, 123(2), 123-142*.

Bortz & Döring (2005). Forschungsmethoden und Evaluation (S. 182). Heidelberg: Springer-Verlag.

Bregman, J., & Wartna, B. (2011). Recidive TBS 1974 – 2008. Ontwikkeling in de strafrechtelijke recidive van exterbeschikkinggestelden. Den Haag:WODC. Fact sheet 2011-6.

Canton, W.J., Veer, T.S. van der, Panhuis, P.J.A. van, Verheul, R. & Brink, W. van den (2004). De voorspellende waarde van risicotaxatie in de pro Justitia rapportage. Onderzoek naar de HKT -30 en de klinische inschatting. *Tijdschrift voor Psychiatrie, 46*(8), *525-536.*

Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112(1), 155-159.*

Fazel, S., Långström, N., Hjern, A., Grann, M., & Lichtenstein, P. (2009). Schizophrenia, substance use, and violent crime. *The Journal of the American Medical* Association, 301, 2016-2023.

Fazel, S., & Yu, R. (2009). Psychotic disorders and repeat offending: Systematic review and meta-analysis. *Schizophrenia Bulletin, doi:10.1093/schbul/sbp135*.

Fluttert, F., Meijel, B. van, Nijman, H., Bjørkly, S., & Grypdock, M. (2010). Preventing aggressive incidents and seclusions in forensic care by means of the 'Early Recognition Method'. *Journal of Clinical Nursing, 19, 1529-1537*. doi: 10.1111/j.1365-2702.2009.02986.x

Gilbert, F., & Daffern, M. (2011). Illuminating the Relationship Between Personality Disorders and Violence: Contributions of a General Aggresion Model. *Psychology of Violence, 1(3), 230-244*. DOI: 10.1037/a0024089.

Gueorguieva, R., & Krystal, J. (2004). Move over ANOVA: Progress in Analyzing Repeated-Measures Data and Its Reflection in Papers Published in the Archives of General Psychiatry. *Archives of General Psychiatry, 61, 310-317*.

Hare, R. D. (1991). The Hare Psychopathy Checklist-Revised. Toronto: Multi-Health Systems.

Hare, R., Clark, D., Grann, M., & Thornton, D. (2000). Psychopathy and Predictive Validity of the PCL-R: An International Perspective. *Behavioral Sciences and the Law 18, 623-645*.

Harte, J., & Breukink, M. (2010). Objectiviteit of schijnzekerheid? Kwaliteit, mogelijkheden en beperkingen van instrumenten voor risicotaxatie. *Tijdschrift voor Criminologie 52(1), 52-72*.

Harte, J., Kalmthout, W. van, & Knüppe, J. (2010). Forensisch Psychiatrisch Toezicht: Evaluatie van de testfase van een vernieuwde vorm van toezicht op tbs-gestelden. Amsterdam: VU Criminologie.

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.

Hilterman, E. (2001). Zekere onzekerheid of onzekere zekerheid. *Sancties, 303-308*.

Hiscoke, U. L., Langström, N., Ottosson, H., & Grann, M. (2003). Self-reported personality traits and disorders (DSM–IV) and risk of criminal recidivism: A prospective study. *Journal of Personality Disorders, 17, 293–305*.

Jeandarme, I., Pouls, C., Wittouck, C., van der Laenen, F., Ampe, M., Verelst, R., Degrauwe, S., Oei, T., & Bogaerts, S. (2015). Forensische pilootprojecten ‚medium security': Herval in delictgedrag na behandeling van geïnterneerden ressorterend onder CBM Gent. *Panopticon 36(3), 227 -247*.

Kogel, C. de, & Hartogh, V. den (2005). Contraire beëindiging van de TBS-maatregel. Aantal, aard en verband met recidive. Den Haag:WODC.

Lammers, S. (2007). Blijven stoute jongens stout? Taxatie van het recidiverisico in de tbs. *De Psycholoog, 42, 479-510*.

Logan, C., & Blackburn, R. (2009). Mental disorder in violent women in secure settings: Potential relevance to risk for future violence. *International Journal of Law & Psychiatry, 32, 31–38*.

Ministerie van Justitie. Dienst Justitiële Inrichtingen (2007). Circulaire. Toetsingskader verlof ter beschikking gestelden. Den Haag: Ministerie van Justitie.

Nijman, H., & Bulten, E. (2006). Risicotaxatie en behandelevaluatie in de forensische psychiatrie – Een overzicht van recente promoties. *Tijdschrift voor Criminologie, 48(1), 78-85*.

Needham, I., Abderhalden, C., Meer, R., Dassen, T., Haug, H., Halfens, R., & Fischer, J. (2004). The effectiveness of two interventions in the management of patient violence of acute mental inpatient settings: Report on a pilot study. Journal of Psychiatric and Mental Health Nursing, 11(5), 595-601.

Philipse, M. (2005). Predicting criminal recidivism. Empirical studies and clinical practice in forensic psychiatry. Nijmegen: Radboud Universiteit.

Philipse, M., Ruiter, C. de, Hildebrand, M. & Bouman, Y. (1999). HCR-20. Beoordelen van het risico van gewelddadig gedrag, versie 2. Nijmegen/Utrecht: Pompestichting/Forum Educatief.

Regioplan Beleidsonderzoek (2006). TBS in beeld. Een beschrijving van het Nederlandse TBS-stelsel. Bijlage bij eindrapport Tbs, vandaag over gisteren en morgen, Tijdelijke commissie onderzoek TBS. Den Haag: Sdu Uitgevers.

Singh, J.P., Serper, M., Reinharth, J. & Fazel, S. (2011). Structured Assessment of Violence Risk in Schizophrenia and Other Psychiatric Disorders: A Systematic Review of the Validity, Reliability and Item Content of 10 Available Instruments. *Schizophrenia Bulletin, 37, 5, 899 – 912*.

Sjöstedt, G. & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments*? International Journal of Forensic Mental Health, 1, 179-184*.

Spreen, M., Brand, E., ter Horst, P., & Bogaerts, S. (2014). Handleiding HKT-R. Historische, Klinische en Toekomstige – Revisie. Gronningen: Stichting FPC Dr. S. van Mesdag.

Vogel, V. de (2005). Structured risk assessment of (sexual) violence in forensic clinical practice. The HCR-20 and SVR-20 in Dutch forensic psychiatric patients. Diss. Universiteit van Amsterdam.

Vogel, V. de, Ruiter, C. de, Hildebrand, M., Bos, B. & Ven, P. van de (2004). Type of discharge and risk of recidivism measured by the HCR-20. A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health, 3*(2*), 149-165*.

Webster, C., Douglas, K., Eaves, D., & Hart, S. (1997). *HCR 20. Assessing Risk for Violence. Version 2.* B.C., Canada: Mental Health, Law and Policy Institute, Simon Fraser University, Burnaby.

Werkgroep Risicotaxatie Forensische Psychiatrie (2002). *Handleiding HKT-30 versie 2. Risicotaxatie in de forensische psychiatrie*. Den Haag: Ministerie van Justitie, Dienst Justitiele Inrichtingen.

Wistedt, B., Rasmussen, A., Pdersen, L., Malm, U., Träskman-Bendz, L., Wakelin, J., & Bech, P. (1990). The Development of an Observer-Scale for Measuring Social Dysfunction and Aggression. *Pharmacopsychiatry 23, 249-252*.

Yu, R., Geddes, J., & Fazel, S. (2012). Personality Disorders, Violence and Antisocial Behavior: A Systematic Review and Meta-Regression Analysis. *Journal of Personality Disorders, 26(5), 775-792*.

Appendix

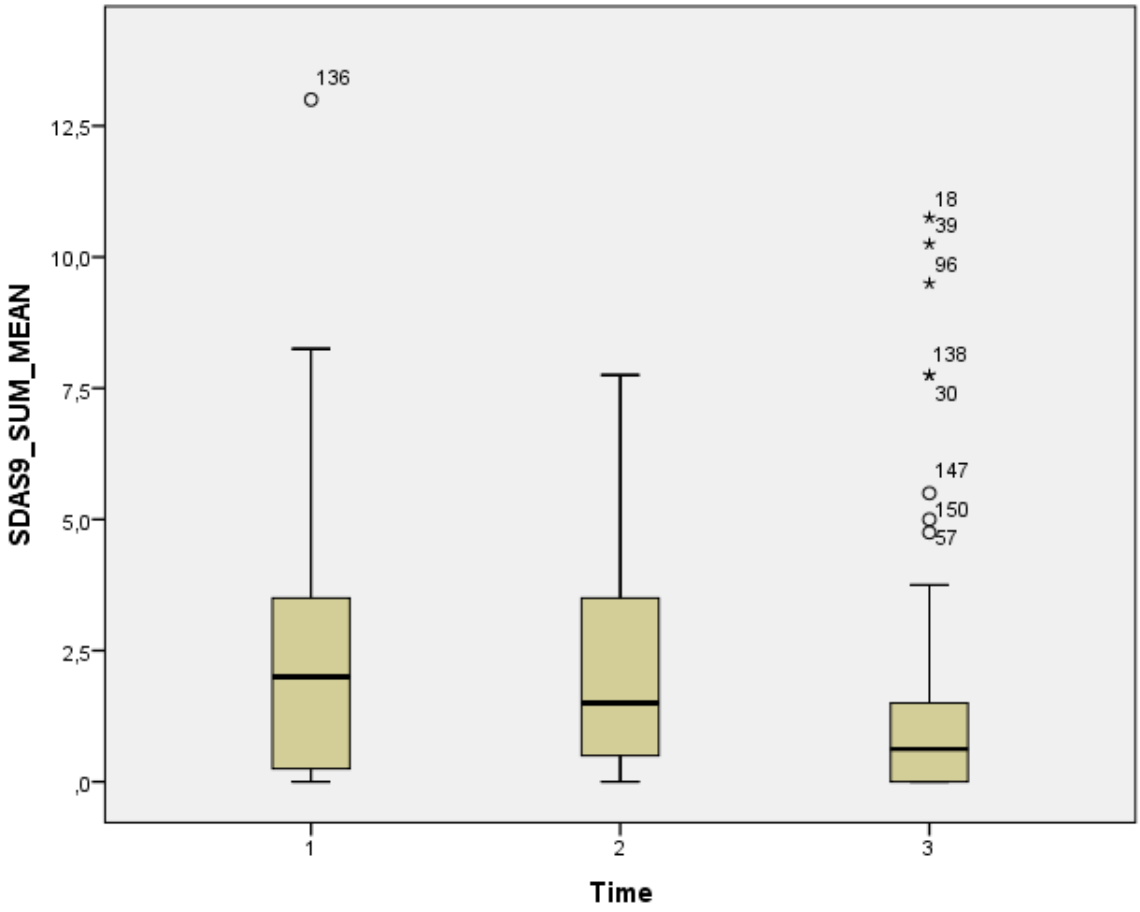

Figure 1. Variance in SDAS-9 mean score. The figures depicts the variance of the SDAS-9 mean score for all three measurements.

Table 7

Correlations and variances of the clinical and future-related scale

	K_1 scores	K_2 scores	K_3 scores	T_1 scores	T_2 scores	T_3 scores	σ
K_1 scores	1	.64**	.68**	.53**	.48**	.55**	60.02
K_2 scores	.64**	1	.83**	.59**	.61**	.61**	57.30
K_3 scores	.68**	.83**	1	.57**	.59**	.67**	91.43
T_1 scores	.53**	.59**	.57**	1	.88**	.72**	30.10
T_2 scores	.48**	.61**	.59**	.88**	1	.80**	28.81
T_3 scores	.55**	.61**	.67**	.72**	.80**	1	32.57

*Note. **= significant for α= .01; all values for N=50*

Table 8

Mixed Model for Future-related Domain as Covariate of SDAS-9 with random intercept: Type III Tests of Fixed Effects[a]

		Denominator		
	Numerator df	df	F	Sig.
Intercept	1	87.36	.64	.427
T scores	1	99.38	16.98	.000

Note. Dependent Variable: SDAS-9 scores.

Table 9

Mixed Model for Future-related Domain as Covariate of SDAS-9with random intercept: Estimates of Fixed Effects[a]

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	-.55	.69	87.36	-.8	.427	-1.93	.83
T scores	.16	.04	99.38	4.12	.000	.08	.24

Note. Dependent Variable: SDAS-9 scores.

Table 10

Mixed Model for Future-related Scores as Covariate of SDAS-9 with random intercept: Model fit

-2 Log Likelihood	645.54
Akaike's Information Criterion (AIC)	663.54
Hurvich and Tsai's Criterion (AICC)	664.83
Bozdogan's Criterion (CAIC)	699.64
Schwarz's Bayesian Criterion (BIC)	690.64

Note. The information criteria are displayed in smaller-is-better form; Dependent variable: SDAS-9 scores.

Table 11

Mixed Model for Future-related Domain as Covariate of SDAS-9 with random intercept and random slope: Type III Tests of Fixed Effects[a]

		Denominator		
	Numerator df	df	F	Sig.
Intercept	1	.64	.10	.822
T scores	1	1.72	16.34	.072

Note. Dependent Variable: SDAS-9 scores.

Table 12

Mixed Model for Future-related Domain as Covariate of SDAS-9 with random intercept and random slope: Estimates of Fixed Effects[a]

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	-.14	.44	.64	-.32	.822	-22.24	21.96
T scores	.13	.03	1.72	4.04	.072	-.04	.29

Note. Dependent Variable: SDAS-9 scores.

Table 13

Mixed Model for Future-related Domain as Covariate of SDAS-9 with random intercept and random slope: Model fit

-2 Log Likelihood	636.437
Akaike's Information Criterion (AIC)	656.437
Hurvich and Tsai's Criterion (AICC)	658.019
Bozdogan's Criterion (CAIC)	696.543
Schwarz's Bayesian Criterion (BIC)	686.543

Note. The information criteria are displayed in smaller-is-better form; Dependent variable: SDAS-9 scores.

Table 14

Mixed Model for Clinical Domain and Future-related Domain as Covariates of SDAS-9 with random intercept: Type III Tests of Fixed Effects[a]

	Numerator df	Denominator df	F	Sig.
Intercept	1	74.61	8.04	.006
T scores	1	111.76	2.42	.123
K scores	1	119.57	27.28	.000

Note. Dependent Variable: SDAS-9 scores.

Table 15

Mixed Model for Clinical Domain and Future-related Domain as Covariates of SDAS-9 with random intercept: Estimates of Fixed Effects[a]

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	-1.77	.62	74.61	-2.84	.006	-3.01	-.52
T scores	.06	.04	111.76	1.56	.123	-.02	.14
K scores	.14	.03	119.57	5.22	.000	.08	.2

Note. Dependent Variable: SDAS-9 scores.

Table 16

Mixed Model for Clinical Domain and Future-related Domain as Covariates of SDAS-9 with random intercept: Model fit

-2 Log Likelihood	624.106
Akaike's Information Criterion (AIC)	644.106
Hurvich and Tsai's Criterion (AICC)	645.689
Bozdogan's Criterion (CAIC)	684.212
Schwarz's Bayesian Criterion (BIC)	674.212

Note. The information criteria are displayed in smaller-is-better form; Dependent variable: SDAS-9 scores.