Routine outcome monitoring as a compass in forensic clinical treatment

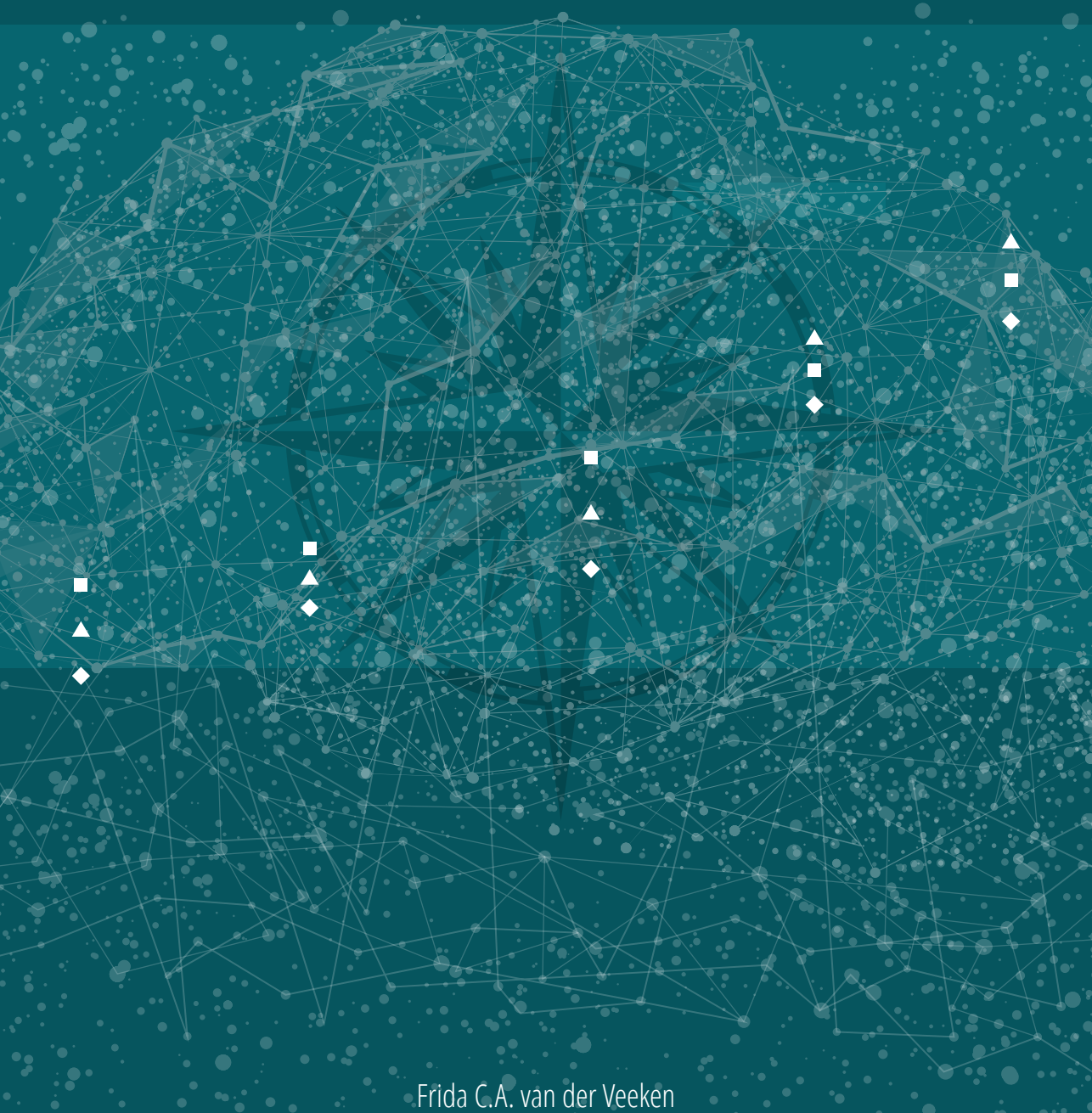
van der Veeken, F.C.A.

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Routine Outcome Monitoring

as a compass in forensic clinical treatment



Frida C.A. van der Veeken

Routine Outcome Monitoring as a Compass in Forensic Clinical Treatment

ISBN: 978 90 3610 550 7

This research was supported by Tilburg University, Department Developmental Psychology and Fivoor, Fivoor Academy of Research and Innovation Development.

Cover: Koen Geers
Layout and print: Haveka B.V. De Grafische Partner

Routine Outcome Monitoring as a Compass in Forensic Clinical Treatment

Proefschrift

ter verkrijging van de graad van doctor

aan Tilburg University

op gezag van de rector magnificus, prof. dr. E. H. L. Aarts,

in het openbaar te verdedigen ten overstaan van een

door het college voor promoties aangewezen commissie

in de Aula van de Universiteit op vrijdag 8 maart 2019 om 13.30 uur

door

Frida Cornelia Antonia van der Veeken,

geboren te Breda.

Promotor:	Prof. Dr. S. Bogaerts
Copromotor:	Dr. J. Lucieer

Promotiecommissie:

Prof. Dr. G. Meynen
Prof. Dr. F. Vander Laenen
Prof. Dr. I. S. J. G. Jeandarme
Dr. M. Spreen

TABLE OF CONTENTS

ABREVIATIONS 7

CHAPTER 1 9
Introduction

CHAPTER 2 33
Routine outcome monitoring in forensic psychiatry;
FPC de Kijvelanden and FPC2 Landen

CHAPTER 3 59
Patient profiles in Dutch forensic psychiatry based on risk factors,
pathology, and offense

CHAPTER 4 87
The Instrument of Forensic Treatment Evaluation: reliability,
factorial structure and sensitivity to measure behavioral change

CHAPTER 5 117
Routine outcome monitoring and clinical decision making in forensic
psychiatry based on the instrument for forensic treatment evaluation

CHAPTER 6 149
Forensic psychiatric treatment evaluation. The clinical evaluation
of treatment progress with multiple forensic routine outcome
monitoring measures

CHAPTER 7 175
General discussion

SAMENVATTING EN CONCLUSIE 199

APPENDICES 206
A. Excel format IFBE [In Dutch]
B. Survey [In Dutch]
C. Feedback guidelines [In Dutch]

CURRICULUM VITAE 210

LIST OF PUBLICATIONS AND PRESENTATIONS 211

DANKWOORD 213

ABREVIATIONS

AAID	The American Association on Intellectual and Developmental Disabilities
AIC	Akaike information criterion
APA	American Psychiatric Association
APA[1]	American Psychology Association
ASP	Atascadero Skills Profile
AVVB	Adaptieve Vragenlijst Verstandelijke Beperking [Adaptive Questionnaire for the Intellectually Disabled]
AUC	Area under the curve
BIC	Bayesian information criterion
CANSAS	Camberwell assessment of needs
CFA	Confirmatory factor analysis
DJI	Dienst Justitiële Inrichtingen [Office for judicial institutions]
DforZo	Directie forensiche zorg [Direction forensic care]
DROS	Dynamic Risk Outcome Scales
DSM	Diagnostic and statistical manual of mental disorders
EFA	Exploratory factor analysis
EFP	Expert group forensic psychiatry
EPA	Ernstig psychiatrische aandoening [severe psychiatric disorder]
FPC	Forensic psychiatric center
FQL	Forensic Inpatient Quality of Life Questionnaire: short version
GLM	Good Lives Model
HKT-30	Historical Clinical Future-30 items
HKT-R	Historical Clinical Future-Revised
HCR-20	Historical Clinical Risk-20 items
HCR-20^{V3}	Historical Clinical Risk-20 items version 3
HoNOS	Health of the Nations Outcome Scales
IDQOL	Intellectual Disabillity Quality of Life
ICC	Intraclass correlation coefficient
IFTE	Instrument for Forensic Treatment Evaluation
KMO	Kaiser-Meyer-Olkin measure

KFZ	Kwaliteit forensische zorg [Quality forensic care]
LCA	Latent class analysis
MANSA	Manchester Assessment of Quality of Life
MID	mild intellectual disorder
MJS	Ministry of Justice and Security
MJenV	Ministerie van Justitie en Veiligheid
N.A.	Non applicable
N.E.I.	Not enough information
NIFP	Nederlands Instituut voor Forensische Psychiatrie en Psychologie [Dutch Institute for Forensic Psychiatry and Psychology]
NJI	Nederlands Jeugd Instituut [Dutch Youth Institute]
NOS	Not otherwise specified
PA	Parallel analysis
PANSS	Positive and negative symptom scale
PCL-R	Psychopathy checklist-Revised
Pr	Practitioners report
Prob	Problem behavior
Prot	Protective behavior
PSD	Personality disorder
PSDS	Personality disorder with co-morbid substance use disorder
Reso	Resocialization skills
RNR	Risk Need Responsivity
ROC	Receiver operating characteristics
ROM	Routine Outcome Monitoring
SEO-r	Schaal voor Emotionele Ontwikkeling voor mensen met een verstandelijke beperking- revised, [Scale for Emotional Development-revised]
SPSS	Statistical package of the social sciences
Sr	Self report
SRZ-p	Social Reliance Scale
START	Short term assessment of risk and treatability
STIP-5	Semi-structured interview for personality functioning DSM-5
SUD	Substance use disorder
TBS	Ter beschikking stelling [disposal by order of the state]
UDS	Urine drugsscreening
VIM	Veilig incidenten melden [safe incident reporting programme]

CHAPTER ONE

General Introduction

Background

Forensic psychiatry generally ministers to a very heterogeneous population (Rice & Harris, 1997; Van Nieuwenhuizen et al., 2011) with patients whose mental health difficulties are directly or indirectly related to offending behavior (Van Marle, 2012; Shinkfield & Ogloff, 2015; Flora, Barbaree, Simpson, Noh, & McKenzie, 2012). As stated by Rice and Harris (1997), the categorization of mentally disordered offenders is a legal categorization defined by persons unfit to stand trial, persons found guilty but mentally ill, persons acquitted due to insanity, mentally disordered sex offenders, sexual predators, and prisoners transferred to mental health facilities (Robertson, Barnao, & Ward, 2011; Rice & Harris, 1997). Due to the judicial aspect of forensic care, its primary goal is to reduce the risk of violent reoffending and to enhance release (Olsson, Strand, Kristiansen, Sjöling, & Asplund, 2013). Treatment focuses on the reduction of risky behavior (Ter Horst, van Ham, Spreen, & Bogaerts, 2014).

Related to this perspective, recidivism or, rather, the absence of recidivism is the main treatment outcome in forensic psychiatric care (Yiend et al., 2011). Risk assessment, therefore, is an important part of clinical forensic psychiatry (Kwee, Schaafsma, & Hildebrand, 2009) and is now part of the treatment of forensic patients. While risk assessment items are predictive for general, sexual, and violent recidivism, they are also an important part of treatment indication and risk management throughout treatment (Hildebrand, Hesper, Spreen, & Nijman, 2005).

Since the last decade, routine outcome monitoring (ROM) has been used more and more often in Dutch mental healthcare. ROM is the routine assessment of clinically relevant treatment outcomes to assess treatment effectiveness. The routine assessment of clinically relevant treatment outcomes informs patients, therapists, management, and insurance companies about treatment changes and progress. Furthermore, ROM feedback offers a lot of information that can be used in the therapeutic patient-therapist relationship, which can improve treatment. ROM can enhance patient empowerment and makes shared decision-making possible. At first, ROM was assessed only in general healthcare, but subsequently, it was also assessed in mental healthcare (Buwalda, 2011).

ROM in forensic psychiatric care has received little attention nationwide and worldwide (Yiend et al., 2011), and is still in its infancy. However, ROM can serve important goals in forensic psychiatry. Firstly, ROM could serve decision-making processes in the case of leave modalities for forensic patients (Schuringa, Spreen, & Bogaerts, 2014), and secondly, ROM could serve the same goal as in other mental healthcare settings: treatment evaluation. ROM in mental health corresponds to the obligatory indicators of symptom functioning, daily functioning, and quality of life (Mulder et al., 2010). Outcome meas-

ures such as symptom remission could also be suitable in forensic psychiatry (De Beurs, & Barendregt, 2008).

Risk of recidivism, however, is an important treatment indicator. Therefore, risk-related items such as impulsivity and violent behavior could be important indicators next to the general mental healthcare indicators of psychiatric symptoms, daily functioning, and quality of life. As ROM in forensic psychiatry and the use of a "forensic" ROM tool have been little studied, the Instrument for Forensic Treatment Evaluation (IFTE) [Instrument voor Forensische Behandel Evaluatie (IFBE)]) has been developed (Schuringa et al., 2014). This thesis evaluates the use of ROM with the IFTE in forensic psychiatry. In this introduction, we will elaborate the Dutch forensic psychiatric framework, the main theoretical framework, and the importance and background of ROM, as well as ROM in forensic psychiatry and patient heterogeneity within the forensic psychiatric population. Finally, we will describe the main goals and the outline of this thesis.

Dutch Forensic Psychiatry

Article 39 of Dutch criminal law states that a person is not punishable when he or she has committed a crime for which he or she cannot be held accountable due to a deficient development or pathological disorder of the mental state (Wetboek van strafrecht, 2018, artikel 39). Offenders who have committed an offense when influenced by a psychiatric disorder can be admitted to a forensic psychiatric center (FPC) involuntarily by order of the state. This so-called TBS order (*Terbeschikkingstelling*: "*disposal to be involuntarily admitted to a forensic psychiatric hospital on behalf of the state*"; De Ruiter & Trestman, 2007) is a judicial measure imposed by a judge (Van Nieuwenhuizen et al., 2011) after a crime has been proven (De Ruiter & Trestman, 2007). When a deficient development or pathological disorder of the mental state was present during the offense, a person can be sentenced to a TBS order if: 1. The offense has a minimum penalty of four years or concerns a category as described by article 37a of Dutch criminal law; and 2. The safety of others, or general safety, is at high risk (wetboek van strafrecht, 2018, artikel 37, 37a).

When this study was performed, there were five classes in the Netherlands to verify a person's level of accountability: full accountability, slightly diminished accountability, diminished accountability, severely diminished accountability, and complete unaccountability (De Ruiter & Trestman, 2007). As of September 2016, the classification levels have been revised to include only three classes; full accountability, diminished accountability, and complete unaccountability (Nederlands Instituut voor Forensische Psychiatrie en Psychologie [Dutch Institute for forensic Psychiatry and Psychology] (NIFP), 2016).

A judge decides whether an offense was influenced by a pathological disorder on the basis of a report by behavioral experts (Dienst Justitiële Inrichtingen [Office for Judicial Institutions] (DJI) van het Ministerie van Justitie en Veiligheid [Ministry of Justice and Security] (MJenV), n.d.[1]). A person can also receive a prison sentence before he or she is committed to an FPC. A TBS order is imposed for two years and can be extended by a judge every one or two years (De Ruiter & Trestman, 2007). This means that a TBS order is imposed not only on non-culpable patients with a severe mental disorder, but also on patients whose offense is not only influenced by their mental disorder, such as by a personality disorder (Van Marle, 2002).

The scale of diminishing accountability, therefore, means that a TBS order can be imposed for a wide range of disorders (De Ruiter & Trestman, 2007). A generalizable study in the Netherlands, with 180 cases from 13 FPCs, illustrates the diversity of the Dutch forensic psychiatric population (Van Nieuwenhuizen et al., 2011). Patients differ in historical factors, offenses committed, and psychopathology. Most patients had experienced treatment in the past and have a criminal record. Patients differ widely on offense type and former DSM-IV-TR (American Psychiatric Association (APA), 2000) Axis I and Axis II diagnoses and current DSM-V diagnoses* (American Psychiatric Association, 2013), and comorbidity numbers are high. Many patients, for example, are diagnosed with both a substance use disorder and a cluster B personality disorder (Van Nieuwenhuizen et al., 2011).

The main goal of the TBS order is to protect society (De Ruiter & Trestman, 2007) and to rehabilitate patients into society. The goals are mainly achieved by patients' treatment and their sound rehabilitation (Van Nieuwenhuizen et al., 2011).

During their stay in a forensic institution, patients usually go through a diagnostic process at the beginning of their stay, and a first risk assessment is generally conducted in the first year of treatment.

Sociotherapy is an important part of daily forensic treatment (Van Nieuwenhuizen et al., 2011); it provides structure and support, motivates patients, and helps patients to find fitting activities. Many patients have activities such as work and sport and attend educational activities in the FPCs. FPCs offer different treatment programs such as cognitive behavioral therapy, psycho-education, skills training, non-verbal therapy, and pharmacotherapy if necessary. Rehabilitation and reintegration into society is realized gradually, with different leave modules. Risk assessment and risk management are important components of forensic psychiatric treatment, and leave approval (Van Nieuwenhuizen et al., 2011) is in line with one of the most used theoretical models: the Risk-Need-Responsivity (RNR) model (Bonta & Andrews, 2007).

*In this thesis, diagnoses in the studies are derived from the DSM-IV-tr (APA, 2000).

Theoretical Framework

The Risk-Need-Responsivity (RNR) principles emerged after the "nothing works movement," which refers to the period after 1970, in which people were convinced that offenders could not be rehabilitated (Bonta & Andrews, 2007). The RNR model, described by Andrews, Bonta and Hoge in 1990, comprises three main principles and is now one of the most used rehabilitation models in forensic psychiatric treatment.

The Risk principle implies that more intensive treatment should be offered to patients with a higher risk of recidivism than to patients with a lower risk of recidivism. Treatment intensity, in other words, should be matched to risk level (Bonta & Andrews, 2007).

The Need principle implies that treatment should focus on criminogenic needs (Bonta & Andrews, 2007). Criminogenic needs can be described as dynamic risk factors that are predictive of future recidivism but can be changed with an intervention (Andrews, Bonta, & Wormith, 2011).

The Responsivity principle states that the patients' rehabilitation opportunities should be maximized and that interventions should be tailored to patients' learning capacities, motivation, abilities, and strengths (Andrews et al., 2011). The Responsivity principle has two parts: general and specific responsivity. Bonta and Andrews (2007) claim that cognitive social learning methods should be used to influence behavior what refers to general responsivity. Specific responsivity emphasizes that a cognitive behavioral intervention should respond to a person's individual strengths, learning style, personality, motivation, readiness to change, mental status, circumstances, and demographic characteristics (Andrews et al., 2011; Bonta & Andrews, 2007).

The RNR model is derived from a "psychology of criminal conduct" (PCC; Andrews et al., 2011). The theory underlying the RNR model is the General Personality and Cognitive Social Learning (GPCSL) perspective on criminal behavior (Andrews et al., 2011). The GPCSL perspective reflects personality predisposition and claims that criminal behavior is learned and governed by a person's expectations. In an intervention, this means that behavior that is rewarded or that is expected to be rewarded is more likely to occur than behavior that is punished or thought to be punished (Bonta & Andrews, 2007).

The Risk principle requires risk assessment to assess the level of future offending risk, and the Need principle requires risk assessment to assess someone's criminogenic needs (Andrews et al., 2012). In response to the RNR model, studies have shown that there are eight central risk factors that are predictive of future violent behavior (Andrews et al., 2012). The big four risk factors are: history of criminal behavior, antisocial personality patterns, antisocial attitudes (values, beliefs, and cognitive emotional states), and antisocial associates. These are complemented by four moderate risk factors: home (family/marital),

school/work, leisure/recreation, and substance use (Andrews et al., 2012). Similar characteristics were found in an earlier review by Harris and Rice (1997), who found that the following factors were related to the risk of recidivism: age, history of criminality and violence, childhood antisocial behavior and aggression, psychopathic traits, interpersonal hostility, institutional rule breaking, anti-social values and sentiments, antisocial peer groups, substance use, poor social problem skills, and poor academic and life skills.

Concomitant with RNR model development, risk assessment has evolved from being a clinical judgment, with a clinician evaluating an offender's risk based on his/her experience and professional training, to being a structured clinical judgment, also referred to as fourth-generation risk assessment. This structured clinical judgment includes the evaluation of an offender's risk with well evaluated tools, which measure both historical and dynamic risk factors while considering a person's personal factors that are important to treatment (Bonta & Andrews, 2007).

The Historical Clinical Future-Revised (HKT-R; Spreen, Brand, Ter Horst, & Bogaerts, 2014) and the Historical Clinical Risk-20 items Version 3 (HCR-20^{V3}; Douglas, Hart, Webster, & Belfrage, 2013) are examples of fourth-generation risk assessment tools. The predecessor of the HKT-R, the Historical Clinical Future 30 items (HKT-30;[Historisch Klinisch Toekomst 30 items]; Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Task Force in Forensic Psychiatry], 2002) has been developed in the Netherlands based on the Dutch forensic situation (Blok, De Beurs, De Ranitz, & Rinne, 2010). Blok et al. (2010) compared Dutch risk assessment instruments and found moderate predictive values of the HKT-30 for serious or violent recidivism in two out of three studies. No significant predictive values were found in the third study conducted by Schönberger et al. (2008). Hildebrand, Hesper, Spreen, and Nijman (2005) found moderate to good predictive values for the total scale and historical scale (AUC = .77 - .87) and marginal predictive values for the clinical and future items and the risk estimate (AUC = .62 - .68) of the HKT-30 in a retrospective design.

Reviews show that treatment adhering to the RNR principles decreases recidivism rates (Andrews et al., 2011). Programs using the Risk principle are more effective when the Need and Responsivity factors are also enhanced (Dowden, Antonowicz, & Andrews, 2003; Andrews & Dowden, 2006). The ability of programs to reduce criminal behavior appears to be related to the extent to which they adhere to all RNR principles (Polaschek, 2012). Studies show that risk indicators, assessed with risk assessment instruments, can assess a decrease in risk factors. A longitudinal study by Morrissey, Beeley, and Milton (2014) shows a very small decrease in risk factors with the HCR-20 in high secure treatment; it is questionable, however, whether this change is of clinical value. De Vries Robbé, De Vogel, Douglas, and Nijman (2014) found lower recidivism rates for patients who showed change on dynamic risk and protective items on pre- and post-treatment measures in a Dutch clinical sample.

Routine Outcome Monitoring in mental healthcare

> Clinical expertise [...] entails the monitoring of patient progress (and of changes in the patient's circumstances—e.g., job loss, major illness) that may suggest the need to adjust the treatment (Lambert, Bergin, & Garfield, 2004). If progress is not proceeding adequately, the psychologist, psychiatrist or psychotherapist alters or addresses problematic aspects of the treatment (e.g., problems in the therapeutic relationship or in the implementation of the goals of the treatment) as appropriate. If insufficient progress remains a problem, the therapist considers alternative diagnoses and formulations, consultation, supervision, or referral. The clinical expert makes decisions about termination in timely ways by assessing patient progress in the context of the patient's life, treatment goals, resources, and relapse potential - (American Psychology Association, presidential task force, 2006, p. 276 - 277).

The APA presidential task force on evidence-based treatment reflects the importance of monitoring patient progress in treatment. The idea of assessing patient progress or treatment outcome was raised as an important part of clinical expertise and evidence-based practice (e.g., "the integration of the best available research with clinical expertise in the context of patient characteristics, culture and preferences") (APA presidential task force, 2006, p. 273).

Outcome monitoring was first mentioned in the medical world with Codman's "end result idea" in 1924 (Kortrijk, 2013, p. 18). The end result idea signified that clinicians should report their work and its results in order to share and to improve treatment methods (Codman, 2009). In 1988, Donabedian described quality of care assessment measured on the basis of structure, process, and outcome, with outcome being defined as the effect of care on health (Donabedian, 1988).

Ellwood (1988), finally, described the importance of implementing what was then called "outcome management" (Ellwood, 1988, p. 1551) in response to the restructured healthcare enterprise in America. A consequence of the restructured healthcare system was that patients, executives, and insurance companies had both critical views and high expectations. Patients and clinicians, however, remained uninformed as to what was the best treatment; insurance companies were skeptical about the efficacy of mental healthcare; and physicians required a tool to enable treatment evaluation, as did executives to support decision-making and specify what treatment method showed the best results.

Outcome management or ROM is a method for gaining insight into the effect of treatment and health outcomes (Ellwood, 1988). With standardized assessment of patient experience, functioning, well-being, and treatment, outcomes can be assessed routinely during

treatment. ROM can also serve the goal of benchmarking or testing policies, which is important for policymakers (Nugter & Buwalda, 2012).

In this thesis, however, ROM is described as an important tool in patient treatment, as the routine assessment of relevant treatment outcomes provides information on treatment progress to therapists as well as patients. ROM assessments at the start of treatment provide insight into patients' functioning at that moment (De Beurs et al., 2011). Routine assessments during treatment provide insight into patients' functioning and the progress they have made or failed to make on treatment goals. Treatment progress feedback informs therapists and patients on treatment progress and could enhance patient and therapist motivation (Slade, 2002). The timely assessment of treatment goals creates opportunities for adjusting the treatment plan in the case of stagnation or decrease (Mulder et al., 2011).

Since the introduction of ROM, models for ROM in mental healthcare have been developed for specific disorders, such as anxiety and mood disorders (De Beurs et al., 2011) together with validated instruments designed especially for ROM, such as the Health of the Nation Outcome Scales (HoNOS; Wing et al., 1998) to assess symptomatic and daily functioning, and the Camberwell assessment of needs (CANSAS; Slade, Loftus, Phelan, Thornicroft, & Wykes, 1999) to assess patients' treatment needs (Slade, Thornicroft, & Glover, 1999; Nugter & Buwalda, 2012). For a long time, however, outcome management was not used by clinicians (Slade et al., 1999[1]; Gilbody, House, & Sheldon, 2003; De Beurs & Zitman, 2007). Therefore, Slade et al. (1999[1]) introduced some feasibility characteristics a ROM tool would have to comply with: A ROM assessment should be brief (assessment and training should not take a lot of time), simple (the questionnaire should be easy to understand), relevant (the instrument should assess clinically relevant items), acceptable (easy to access, free), and of value (the use of normative data would have to make sense). A ROM tool should also have the sound psychometric properties of reliability, validity, and sensitivity to assess change (Slade et al., 1999[1]; Schoen, & Derksen, 2011).

Even though ROM was used on a limited scale (Slade, 2002[1]), some studies did show positive results of ROM and feedback (Lambert et al., 2003; Carlier et al., 2012) on treatment. A feedback system can increase positive outcomes and decrease deterioration by four to eight percent (Lambert et al., 2003). Studies on feedback and patient outcomes, however, have shown different results (De Jong, 2012). One review showed that a considerable part of studies found a positive effect of ROM on patients' mental state, but mainly for patients whose treatment did not seem to work at first hand (Carlier et al., 2012). ROM has a positive effect on the behavior of clinicians, in the sense of more adequate diagnoses and earlier treatment adjustments (Carlier et al., 2010), and on therapist-patient communication (Carlier et al., 2010).

Over the past two decades, after computer systems had been implemented and insurance companies began to exert pressure (Nugter, & Buwalda, 2012), the use of ROM has increased in Dutch mental healthcare (Van der Feltz-Cornelis, Volker, & de Heer, 2010; De Beurs & Zitman, 2007). The general key indicators for ROM measurements selected in Dutch general mental healthcare are symptom severity, daily functioning, and patient well-being or quality of life (Expertgroep ROM Forensische Psychiatrie [Expert group ROM Forensic Psychiatry], 2011). In 2002, a large ROM study was implemented in the Netherlands for patients with mood, anxiety, and somatoform disorders (De Beurs, & Zitman, 2007), and several ROM sets have been developed to evaluate treatment for different patient groups in mental healthcare, such as patients with severe depression (Schulte-Van Maaren et al., 2013), severe psychiatric disorders (Mulder et al., 2010), and children and adolescents (Boer, Markus, & Vermeiren, 2012). Several studies in the Netherlands have evaluated ROM in mental healthcare (Kortrijk, 2013; Buwalda, 2013; De Jong, 2012; Van der Lem, 2013). Specific forensic ROM indicators, however, have received less attention (Goethals & Van Marle, 2012).

Routine outcome monitoring in forensic psychiatry

As treatment in Dutch forensic psychiatry is indicated by psychopathology, offense, and risk and protective factors, treatment should also focus on these factors. According to the RNR model, the reduction of criminogenic needs and the consideration of responsivity factors are important treatment goals. As preventing recidivism is the main goal of treatment according to the RNR model, interventions should focus on reducing risk behavior, and change can occur by enhancing pro-social alternatives (Polaschek, 2012). ROM in forensic psychiatry, therefore, should monitor dynamic criminogenic needs, such as hostility, drug use, and impulsivity; pro- social behavior, such as coping skills and labor skills; and responsivity factors, such as strengths and motivation related to forensic treatment, together with the proclaimed ROM indicators for mental health: symptom functioning, daily functioning, and quality of life.

Study reports from the UK describing important treatment outcomes in forensic treatment claim that recidivism is the most important outcome (Yiend et al., 2011; Fitzpatrick et al., 2010), but this is an "ultimate" outcome, which can only be assessed after release and/or clinical treatment. While mental health, quality of life, social functioning, and psychosocial adjustment also are important treatment outcomes (Fitzpatrick et al., 2010), measurable risk factors could lead to future risk behavior when not assessed and untreated (Yiend et al., 2011), and change could be related to risk of recidivism (De Vries Robbé et al., 2014).

Tools assessing dynamic risk factors are interesting (Fitzpatrick et al., 2010) because they target risk factors that may change due to an intervention and are predictive of future recidivism. The use of an instrument assessing dynamic risk factors as a routine outcome measure, however, has not been widely studied (Fitzpatrick et al., 2010). Current risk assessment tools assessing dynamic risk factors include the Historical Clinical Risk-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) and its revised HCR-20^{V3} (Douglas et al., 2013) and the Historical Clinical Future-30 items (HKT-30, Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Task Force in Forensic Psychiatry], 2002) and its successor the HKT-R (Spreen et al., 2014).

However, these instruments have been developed to assess the risk of recidivism (Goethals & Van Marle, 2012) and have limited response categories. It is recommended, therefore, to assess dynamic risk factors in a tool developed for forensic treatment evaluations (Goethals & Van Marle, 2012) with a more dynamic scale that is sensitive to measuring change. ROM in forensic psychiatry could thus enhance dynamic risk factors and support desirable treatment outcomes, such as adequate employment (Yiend et al., 2011) and social functioning, that could aid rehabilitation (Fitzpatrick et al., 2010).

The Instrument for Forensic Treatment Evaluation

To evaluate forensic psychiatric treatment in a more standardized way and to compare evaluations with previous ones, the Instrument of Forensic Treatment Evaluation (IFTE) was developed in 2010 (Schuringa, 2010). The IFTE has been designed especially for forensic psychiatric treatment, with close consideration of the RNR principles, specifically for forensic treatment evaluation (Schuringa, Spreen, & Bogaerts, 2014).

The IFTE consists of 22 items and measures forensic clinical treatment outcomes divided over three factors: *problem behavior, protective behavior*, and *resocialization skills* (Schuringa et al., 2014). The instrument includes the 14 clinical dynamic items of the HKT-R. All Dutch clinical forensic treatment facilities are obliged to assess these 14 clinical HKT-R items annually (DJI, nd.). In addition to these 14 items, three more items were selected from the ASP-NV (ASP-NV; Vess, 2001): *skills to prevent substance use, skills to prevent physically aggressive behavior* and s*kills to prevent sexually deviant behavior* (Schuringa et al., 2014). The items of the IFTE were first based on the ASP-NV because the ASP-NV offers the opportunity to evaluate behavior. After clinical experience, items from the HKT-30, and later the HKT-R, were added after these had been validated in a Dutch population (Schuringa et al., 2014). Five additional items, finally, were included which were considered clinically important by clinicians: *manipulative behavior, balanced daytime activities, financial skills, sexually deviant behavior,* and *medication use* (Schuringa et al., 2014).

Schuringa et al., (2014) studied the inter-rater and test-retest reliability, internal consistency, factorial structure, sensitivity, and specificity of the IFTE. They displayed a moderate to good inter-rater reliability with an intra-class correlation coefficient on average measures (ICC = .65 - .92), and a moderate to good test-retest reliability with Cronbach's alpha (α > .61). The internal consistency was good for the three factors (α = .86 - .90).

The IFTE was designed to gather information from multiple disciplines in forensic psychiatry, assessing the IFTE individually and producing a composite score. Similar to the HKT-R items, the IFTE items have been designed with a five-point scale. Between every two Likert scales, however, three in-between scoring options are available, which gives clinicians the opportunity to make finer distinctions between two scales and gives the IFTE a more dynamic 17-point scale (Schuringa et al., 2014). While not all disciplines focus on all individual treatment indicators, all items also have the option of ticking the "N.E.I." (not enough information) box when a therapist does not have enough information to assess an item. As not all criminogenic needs are applicable to all patients, some items also have an "N.A." (non-applicable) option. When an item such as *medication use* is not applicable to a patient, the therapist can tick the N.A. box (Schuringa et al., 2014), thus excluding non-applicable criminogenic needs.

An individual IFTE report has been developed at the Dr. S. van Mesdag center to provide feedback to both patients and therapists. Patients' individual treatment and current functioning are evaluated in this report, with composite scores of the IFTE factors and items. The report also shows how many therapists have assessed an item and what the agreement between these raters is. The IFTE report shows if patients have shown a significant increase, stagnation, or decrease in problematic, protective, and resocialization behavior (Schuringa, Heininga, & Spreen, 2011).

The IFTE, in sum, offers therapists the opportunity to assess ROM in a forensic setting taking into consideration the theoretical framework of the RNR model. It is considered to be a dynamic tool that can assess behavioral change by using a 17-point scale. The IFTE also evaluates treatment on the basis of the evaluations of multiple therapists, giving a multidisciplinary view. With the IFTE assessments, a treatment team can gain insight into patients' treatment progress according to the RNR principles. The main question of this thesis, therefore, is: Is the IFTE an appropriate instrument with adequate psychometric qualities to be used as a ROM tool in a forensic inpatient setting, a setting that involves a highly heterogeneous population?

Patient heterogeneity

If an offense is associated with the offender's psychopathology, this is one of the judicial

reasons for referring a person to a forensic institution. As already mentioned in this chapter, the heterogeneity of forensic patients' characteristics is high and covers a wide range of treatment needs (Rice & Harris, 1997). Patients differ in their diagnoses, both on former axis I and axis II of the DSM-IV-TR (APA, 2000) and current DSM-V diagnoses (APA, 2013), co-morbidity rates, crimes committed, risk levels, and criminogenic needs (Van Nieuwenhuizen et al., 2011). Ogloff, Talevski, Lemphers, Wood, & Simmons (2015) also report that co-morbidity numbers are high in offending populations, as they often cope with psychotic, mood, and/or substance use disorders and co-morbid antisocial personality disorders. They report that co-occurring disorders can be related to offending behavior and functioning (Ogloff et al., 2015). Psychiatric Axis I disorders themselves, however, are not always directly linked to criminal behavior (Peterson, Skeem, Kennealy, Brav, & Zvonkovic, 2014).

As risk and protective factors, psychopathology, and offense types are related to each other, it is important to assess the influence of these factors and their combinations on treatment and treatment progress. Differing criminogenic needs can influence treatment progress in the case of low or high risk (Bonta & Andrews, 2007), but diagnoses could possibly also have an influence on treatment progress: patients with a substance use disorder and an antisocial personality disorder, for instance, could be less responsive to treatment (Ogloff et al., 2015).

Purpose of this study

In mental healthcare, ROM can be considered as daily practice (Kortrijk, 2013). The routine assessment of treatment progress makes it possible to monitor treatment goals in a standardized manner and to adjust patients' treatment plans if necessary. ROM can also support the therapist- patient relationship because ROM results can be discussed periodically in therapy sessions. Several ROM tools have been developed to achieve this goal in general mental healthcare.

In forensic psychiatry, however, ROM has only been studied to a limited extent, while we hypothesize that ROM could serve the same goal in forensic psychiatric treatment as it does in regular mental healthcare. The evaluation and monitoring of relevant forensic risk, protective, and resocialization factors could inform therapists during treatment. Treatment goals that have not been attained can be detected earlier, leading to a timely adjustment of a treatment plan. Input from patients and feedback to patients could improve patients' responsivity and motivation. Discussion of RNR-related factors could benefit both patients and therapists in a forensic psychiatric inpatient treatment.

The goal of this thesis is to study if the IFTE is applicable in forensic psychiatric inpatient treatment as a forensic ROM instrument. At first, a set of ROM tools will be assembled for the three main therapeutic environments in Dutch forensic mental healthcare: the social therapeutic environment, the supportive environment, and the intellectually disabled group of patients. As forensic psychiatry ministers to a very heterogeneous population, we want to construct patient profiles based on three patient characteristics: offense type, historical and dynamic risk factors, and psychopathology. This could provide us with a more specific and homogeneous image of different patient groups and makes it possible to study whether treatment progress differs between identifiable patient groups.

As the main focus of this thesis regards the use of the IFTE as a ROM instrument, the IFTE will be evaluated on the basis of the requirements such an instrument should possess. As mentioned, a ROM instrument should have sound psychometric properties with respect to reliability, validity, and sensitivity to assess change (Slade et al., 1999[1]; Schoen & Derksen, 2011) in clinically relevant outcomes. The reliability and sensitivity to assess change will be assessed for the benefit of the IFTE's clinical use.

Inpatient incidents are important predictors of future problematic behavior after treatment (Spreen et al., 2014), and leave modules during treatment are important steps in the resocialization of patients during forensic treatment (Jeffery & Woolpert, 1974; LeClair, & Guarino-Ghezi, 1991). To assess whether the IFTE can aid in forensic decision-making, its predictive validity will be assessed for negative treatment outcomes, such as inpatient incidents, and positive treatment outcomes, such as leave approval. Finally, we will study treatment progress assessed with the IFTE in the clinical forensic psychiatric population.

Problem and Significance

ROM studies in forensic psychiatry are scarce. This thesis is one of the first to study the use of a specific forensic ROM tool, the Instrument for Forensic Treatment Evaluation (IFTE), in a clinical forensic setting.

Research questions

- Which ROM tools can we apply in forensic psychiatry (**Chapter 2**)?
- Can we identify patient profiles based on clinical patient characteristics (**Chapter 3**)?
- Is ROM and, more specifically, the Instrument for Forensic Treatment Evaluation (IFTE), applicable in forensic psychiatric inpatient treatment?
- What are the psychometric qualities of the Instrument for Forensic Treatment Evaluation (**Chapter 4**)?

- What is the predictive validity of the Instrument for Forensic Treatment Evaluation (IFTE) for important forensic inpatient treatment outcomes (**Chapter 5**)?
- Can we assess patient progress with the Instrument for Forensic Treatment Evaluation (IFTE), and does this progress differ between identifiable patient profiles in forensic psychiatry (**Chapter 6**)?

Participants

All data are primary treatment information derived from electronic patient files. All patients resided either in FPC De Kijvelanden or in FPC 2landen and were sentenced to a TBS order and compulsory psychiatric treatment. Studies in this thesis were conducted at different time intervals, and participants, therefore, were described on the basis of age, offense type, and diagnosis in every study.

Study setting

This study has been conducted in FPC de Kijvelanden and FPC 2landen. Both institutions are closed settings with a high security level. Patients residing in both institutions are admitted with a tbs measure. Even though the security levels are high, both institutions are treatment facilities. FPC de Kijvelanden, has been founded in 1995 and is located in Poortugaal, Zuid-Holland. The FPC has several departments, including a department for very intensive care and a resocialisation unit outside of the clinic in Rotterdam. The clinic had a capacity of 174 beds in 2011, this number has been reduced to 138 beds during this study. FPC 2landen has been founded in 2009 as a reaction to the high number of tbs patients that could not be admitted to an FPC, in cooperation with FPC de Kijvelanden and Altrecht (Lucieer, 2015). FPC 2landen was located in the City of Utrecht and had a capacity of 55 beds. The clinic also had different departments within the clinic and a resocilialisation department adjacent to the clinic. However, several years after the opening of FPC 2landen, the number of patients with a tbs measure decreased over the years and patient outflow increased. Therefore the state secretary for Security and Justice decided that two FPC's had to be closed. Due to this decision, FPC 2landen was closed in 2015 and patients were admitted to either FPC de Kijvelanden or other institutions. De Kijvelanden also has a forensic psychiatric clinic and a forensic psychiatric department together with several outpatient treatment facilities and offers forensic sheltered housing. In January 2018 De Kijvelanden merged into Fivoor, together with the organisations Palier and Aventurijn.

General procedure

This thesis concerns a ROM study in forensic psychiatry and has been approved by the sci-

entific research committee of FPC De Kijvelanden. ROM was implemented in both FPC 2landen and FPC De Kijvelanden in September 2011. Therapists were asked to assess the IFTE in an Excel document appointed to a specific patient, with an abbreviated instruction (Appendix A, Appendix B). All IFTEs were scored in these Excel documents, which presented the opportunity of conducting an individual treatment evaluation report. A first ROM assessment was realized for approximately 40% of patients in FPC De Kijvelanden and 20% of patients in FPC 2landen in 2011. Mid 2012, ROM was implemented for all patients with a TBS order in both clinics.

After the first two assessments in 2011, ROM was conducted in preparation of a patient's individual treatment evaluation discussion for all patients in both clinics. All IFTEs, therefore, were assessed in preparation of these treatment evaluation discussions. Their main goal was to provide clinicians with a report which they could use to evaluate their patients' progress in these treatment evaluation discussions and to give therapists the opportunity to discuss these results with patients. All ROM measurements in this study, therefore, were primarily gathered for treatment benefits. Periods between assessments, therefore, depend on the frequency of these patient treatment evaluation discussions.

Different disciplines have been approached to assess a ROM questionnaire every treatment evaluation discussion. ROM was first assessed in Excel formats (Appendix A) or official questionnaires; after that, IFTE questionnaires were integrated into the electronic patient profiles (Appendix B). The assessments have been gathered for analysis purposes in this thesis.

Thesis outline

Chapter 2 elaborates on ROM instruments that are usable in a forensic psychiatric center for three identifiable treatment groups: social therapeutic, supportive, and mildly intellectually disabled groups.

Chapter 3 focuses on the recognition of more homogeneous groups as forensic psychiatry ministers to a very heterogeneous population. Psychiatric disorders, risk factors for recidivism, and the index offense offer important treatment information, and so these were gathered to assess latent class analysis in order to identify patients profiles. This thesis further focuses on the use of the IFTE as a forensic ROM tool at two FPCs. These chapters focus on some basic elements a ROM tool should possess.

Chapter 4 examines the psychometric values of the IFTE. Its test retest reliability, intra-rater reliability, and internal consistency have been studied, together with its factor structure and its ability to measure change.

Chapter 5 focuses on the predictive validity of the IFTE and its clinical use. In this study, the predictive validity for internal aggressive incidents and unreliable urine drug

screening tests or use of drugs was studied with IFTEs used in treatment. To assess whether the IFTE endorses current treatment decisions, its predictive validity was also assessed for current leave approvals, an essential part of treatment in an FPC to enhance rehabilitation.

Chapter 6, studies whether patients at an FPC show positive treatment change over time. As the IFTE used in this study assessed part of the whole treatment rather than the entire treatment for most patients, patients were divided into those who had assessments at the start of treatment and those whose assessments started during treatment in order to assess whether patients gain more progress at the start of treatment. After that, patients were also divided into those with high and low problematic scores, to assess whether those with more problematic scores at the first assessment would show more progress than those with better functioning scores.

Chapter 7, finally gives a comprehensive view of the results and discusses the findings an conclusion of this thesis as well as limitations and directions for future research.

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders – DSM-IV-TR.* Arlington, VA: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19-52.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2011). Does adding the good lives model contribute to effective crime prevention? *Criminal Justice and Behavior, 38,* 735 – 755. doi: 10.1177/0093854811406356

Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology, 50,* 88 – 99. doi: 10.1177/0306624X05282556

Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., & Wormith, J. S. (2012). Are the major risk/need factors predictive of both female and male reoffending? A test with the eight domains of the level of sercive/case management inventory. *International Journal of Offender Therapy and Comparative Criminology, 56,* 113 – 133. doi: 10.1177/0306624X10395716.

APA Presidential Task Force. (2006). Evidence based practice in psychology. The American Psychology Association, 61; 271 – 285. doi: 10.1037/0003-066X.61.4.271

Blok, G. T., De Beurs, E., De Ranitz, A. G. S., & Rinne, T. (2010). Psychometrische stand van zaken van risicotaxatie-instrumenten voor volwassenen in Nederland. [The current psychometric status of risk assessment scales for adults in the Netherlands]. *Tijdschrift voor Psychiatrie, 52,* 331 - 341.

Boer, F., Markus, M. T. H., & Vermeiren, R. R. J. M. (2012). Zeg niet: 'Sta stil, anders kan ik je niet meten!' ROM in de kinder- en jeugdpsychiatrie. [Don't say: Stand still, or I can't measure you! Routine outcome measurement in child and adolescent psychiatry] *Tijdschrift voor Psychiatrie, 2,* 167 – 171.

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and treatment* (User Report No. 2007-06). Ottawa, Ontario: Public Safety Canada. Retrieved from: https://cpoc.memberclicks.net/assets/Realignment/risk_need_2007-06_e.pdf

Buwalda C. S. "Praktijkboek ROM in de ggz, een leidraad voor gebruik en implementatie van meetinstrumenten", De Tijdstroom, Utrecht, 2011

Buwalda, V. J. A. (2013). Routine outcome monitoring in dutch psychiatry: Measurement, instruments, implementation and outcome. (Dissertation VU medisch centrum).

Carlier, I. V. E., Meuldijk, D., Van Vliet, I. M., Van Fenema, E., Van der Wee, N. J. A., & Zitman, F. G. (2010). Routine outcome monitoring and feedback on physical mental health status; evidence and theory. Journal of Evaluation in Clinical Practice, 18, 104-110. doi: 10.1111/j.1365-2753.2010.01543.x

Carlier, I. V. E., Meuldijk, D., Van Vliet, I, Van Fenema, E. M., Van der Wee, N. J. A., & Zitman, F.G. (2012). Empirische evidentie voor de effectiviteit van routine outcome monitoring; een literatuuronderzoek. [Empirical evidence for the effectiveness of Routine Outcome Monitoring. A study of the literature] *Tijdschrift voor Psychiatrie, 54*, 121-8.

Codman, E. A. (2009). The classic: The registry of bone sarcomas as an example of the endresult idea in hospital organization. *Clinical Orthopaedics and Relational Research, 467,* 2766 – 2770. doi: 10.1007/s11999-009-1048-7

De Beurs, E., & Barendregt, M. (2008). Mogelijkheden voor therapie-effectonderzoek in de tbs-secor. Wetenschappelijk onderzoek en documentatiecentrum. Retrieved from: https://www.wodc.nl/onderzoeksdatabase/methoden-voor-onderzoek-effectiviteit-tbs-behandeling.aspx

De Beurs, E., & Zitman, F. G. (2007). Routine outcome monitoring. Het meten van therapieeffect in de klinische praktijk met webbased software. *Maandblad Geestelijke Volksgezondheid, 62*; 13 – 28.

De Beurs, E., Den Hollander-Gijsman, M. E., Van Rood, Y. R., Van der Wee, N. J. A., Giltay, E. J., Van Noorden, ... Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands; practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology and Psychotherapy, 18,* 1-12.

De Jong, K. (2012). A chance for change: building an outcome monitoring feedback system for outpatient mental health care. (Dissertation Leiden University). Ridderkerk grafisch bedrijf, Ridderkerk.

De Ruiter C, & Trestman R. L. (2007). Prevalence and treatment of personality disorders in Dutch forensic mental health services. *Journal of the American Academy of Psychiatry and the Law, 35,* 92-97.

De Vries Robbé, M., De Vogel, V., Douglas, K. S., & Nijman, H. L. I. (2014). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting reductions in rostdischarge community recidivism. *Law and HumanBehavior*. Advance online publication. http://dx.doi.org/10.1037/lhb0000089

Dienst Justitiële Inrichtingen. (nd). *Risicotaxatie-instrumenten* [Risk assessment tools]. Retrieved from: https://www.forensischezorg.nl/beleid/risicotaxatie-instrumenten.

Dienst Justitiële Inrichtingen. (nd[1]). *Justitiabelen in forensische zorg* [Offenders in forensic care. Retrieved from: https://www.dji.nl/justitiabelen/justitiabelen_in_forensische_zorg/index.aspx

Donabedian, A. (1988). The quality of case. How can it be assessed? *Journal of the American Medical Association, 260,* 1743 – 1748.

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). HCR-20V3 Historical, Clinical, Risk Management (Version 3): Professional guidelines for evaluating risk of violence. Vancouver, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Dowden, C., Antonowicz, D., & Andrews, D. A. (2003). The effectiveness of relapse prevention with offenders: A meta-analysis. *International Journal of Offender Therapy and Comparative Criminology, 47,* 516-528. Doi: 10.1177/0306624X03253018

Ellwood, P. (1988). Outcomes management. A technology of patient experience. *New England Journal of Medicine, 318,* 1549-1556. doi:10.1056/NEJM198806093182329

Expertgroep forensische psychiatrie (2011). *ROM forensische psychiatrie: Adviesdocument. [ROM forensic psychiatry: consultation document].* GGZ Nederland.

Fitzpatrick, R., Chambers, J., Burns, T., Doll, H., Fazel, S., Jenkinson, C., . . . Yiend, J. (2010). A systematic review of outcome measures used in forensic mental health research with consensus panel opinion. *Health Technology assessment*, 18, 1 – 91. doi. 10.3310/hta14180

Flora, N., Barbaree, H., Simpson, A. I. F., Noh, S., & McKenzie, K. (2012). Pathways to forensic mental health care in Toronto: A comparison of European, African Caribbean, and other ethnoracial groups in Toronto. *La Revue Canadienne de Psychiatrie, 57*, 414 - 421. doi: 10.1177/070674371205700704

Gilbody, F. M., House, A. O., & Sheldon, T. A. (2003). Psychiatrists in the UK do not use outcome measures. National survey. *British Journal of Psychiatry*, *180*, 101 – 103. doi: 10.1192/bjp.180.2.101

Goethals, K., & Van Marle, H. J. C. (2012). Routine outcome monitoring in de forensische psychiatrie: Een lang verhaal in het kort. [Routine outcome monitoring in forensic psychiatry: A long story cut short] *Tijdschrift voor Psychiatrie, 2*, 179-183. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/issues/450/articles/9335

Harris, G. T., & Rice, M. E.. (1997). Risk appraisal and management of violent behavior. *Psychiatric Services, 48,* 1168 - 1176

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. [The value of structured risk assessment and diagnosis of psychopathy. A study into the reliability and predictive validity].* Expertise Centrum Forensische Psychiatrie, Utrecht.

Jeffery, R., & Woolpert, S. (1974). Work furlough as an alternative to incarceration: assessment of its effects on recidivism and social cost. *The Journal of Criminal Law and Criminology, 65*, 405-415. Retrieved from: http://scholarlycommons.law. northwestern.edu/cgi/viewcontent.cgi?article=5908&context=jclc

Kortrijk, H. E. (2013). Use of routine outcome monitoring data for evaluating assertive community treatment. (Dissertation, Erasmus University Rotterdam). Ipskamp drukkers, The Netherlands

Lambert, M. J., Whipple, J. L., Hawkins, E. D., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003) Is it time for clinicians to rouintely track patient outcome? A meta-analysis. *American Psychological Association, D12,* 288 – 301

LeClair, D. P., & Guarino-Ghezzi, S. (1991). Does incapacitation guarantee public safety? Lessons from the Massachusetts furlough and prerelease programs. *Justice Quarterly, 8*, 9 – 36. doi: 10.1080/07418829100090891

Lucieer, J. (2015). *Tender IV: FPC 2landen.* Oisterwijk: Wolf Legal Publishers

Morrissey, C., Beeley, C., & Milton, J. (2014). Longitudinal HCR-20 scores in a high-secure psychiatric hospital. *Criminal Behaviour and Mental Health, 24,* 169 - 180. doi: 10.1002/cbm.1893

Mulder, C. L., Van der Gaag, M., Bruggeman, R., Cahn, W., Delespaul, P. A. E., Dries, P., ... Van Os, J. (2010). Routine outcome monitoring voor patiënten met ernstige psychiatrische aandoeningen; een consensusdocument. [Routine outcome monitoring for patients with severe mental illness: a consensus document]. *Tijdschrift voor Psychiatrie, 52,* 169-179. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/articles_2927pdf.pdf Nederlands Instituut voor Forensische Psychiatrie en Psychologie. (2016). *Jaarbericht 2016.* Den Haag: NIFP

Nugter, M. A., & Buwalda, V. J. A. (2012). Achtergronden en gebruiksmogelijkheden van ROM in de GGZ. [Background and possible use of ROM in mental health care]. *Tijdschrift voor Psychiatrie, 54;* 111-120. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/TvP12-02_p111-120.pdf

Ogloff, J. R. P., Talevski, D., Lemphers, A., Wood, M., & Simmons, M. (2015). Co-occuring mental illness, substance us disorders, and antisocial personality disorder among clients of forensic mental health services. *Psychiatric Rehabilitation Journal, 38,* 16-23. doi: http://dx.doi.org/10.1037/prj0000088

Olsson, H., Strand, S., Kristiansen, L., Sjöling, M., & Asplund, K. (2013). Decreased risk for violence in patients admitted to forensic care, measured with the HCR-20. *Archives of Psychiatric Nursing, 27,* 191 - 197. doi: 10.1016/j.apnu.2013.03.004

Peterson, J. K., Skeem, J., Kennealy, P., Bray, B., & Zvonkovic, A. (2014). How often and how consistently do symptoms directly precede criminal behavior among offenders with mental illness? *Law and Human Behavior, 38,* 439 - 449. doi: 10.1037/ lhb0000 075

Polaschek, D. L. L. (2012). An appraisal of the risk-need-responsivity (RNR) model of offender rehabilitation and its application on correctional treatment. *Legal and Criminological Psychology, 17,* 1 -17. doi: 10.1111/j.2044-8333.2011.02038.x

Rice, M. E., & Harris, G. T. (1997). The treatment of mentally disordered offenders. *Psychology, Public Policy, and Law, 31,* 126 - 183. doi: 10.1037/1076-8971.3.1.126

Robertson, P., Barnao, M., & Ward, T. (2011). Rehabilitation frameworks in forensic mental health. *Agression and Violent Behavior, 16,* 472-484. doi: 10.1016/j.avb.2011.03.003

Schoen, R., & Derksen, J. (2011). Praktische vergelijking van uitkomstinstrumenten in de geestelijke gezondheidszorg. *Directieve therapie, 31*, 209 – 248. doi: 10.1007/s12433-011-0012-3

Schulte-van Maaren, Y. W. M., Carlier, I. V. E., Zitman, F. G., Van Hemert, A. M., De Waal, M. W. M., Van der Does, A. J. W., ... Giltay, E. (2013). Reference values for major depression questionnaires: The Leiden routine outcome monitoring study. *Journal of Affective Disorders, 149,* 342 – 349. doi:10.1016/ j.jad.2013.02.009

Schuringa, E. (2010). Routine outcome monitoring in het FPC dr. S. van Mesdag. GGzet Wetenschappelijk, 1, 27 - 35

Schuringa, E. (2011). Een voorbeeld uit het patiënt volg systeem van het FPC Dr. S. van Mesdag. *GGzet Wetenschappelijk, 2*, 70-77. [An example of the "patient following" system of the FPC Dr. S. Van Mesdag]. Retrieved from: http://static.mesdag .wdmc.eu/ niAe9OzHKU-x2JO03A57HQ/TW4xvBca2ENSZYBwBXk9Pg/WG -qCbngKkl-SiZk6ShJEGw/Een%20voorbeeldcasus%20uit%20het%20Patient%20Volg %20Systeem%20-%202011.pdf

Schuringa, E., Heininga, V., & Spreen, M. (2011). De N = 1 statistiek achter het patiënt volg systeem in het FPC Dr. S. van Mesdag. [The N=1 statistics behind the patient following system in the FPC Dr. S. van Mesdag] GGzet Wetenschappelijk, 15, 70-77.

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,* 127 - 144. doi: 10.1080/ 15228 932.2014.897536

Shinkfield, G., & Ogloff, J. (2015). Use and interpretation of routine outcome measures in forensic mental health. *International Journal of Mental Health Nursing, 24,* 11-18. doi: 10.1111/inm.12092

Slade M. Editiorial (2002). Routine outcome assessment in mental health services. *Psychological Medicine, 32,* 1339-43.

Slade, M. (2002[1]). What outcomes to measure in routine mental health services, and how to assess them: a systematic review. *Autstralian and New Zealand Journal of Psychiatry, 36,* 743 – 753.

Slade, M., Loftus, L., Phelan, M., Thornicroft, G., & Wykes, T. (1999) The Camberwell Assessment of Need. Gaskell, London

Slade M, Thornicroft, G, & Glover, G. (1999)[1]. The feasibility of routine outcome measures in mental health. *Social Psychiatry and Psychiatric Epidemiology, 35(5)*, 243-249

Spreen, M., Brand, E., Ter Horst, P., & Bogaerts, S. (2014). *Handleiding, Historische Klinische Toekomst-Revisie versie 2014 [Manual, Historical, Clinical, Future, Revised version 2014].* The Netherlands, Groningen: Stichting FPC Dr. S. van Mesdag.

Ter Horst, P., van Ham, M., Spreen, M., & Bogaerts, S. (2014). Behandelevaluatie en klinische besluitvorming met HKT-30-ROM. [Treatment evaluation and clinical decision making using HKT-30-ROM]. *Tijdschrift voor Psychiatrie*, 56(4), 228-36. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/56-2014-4 -artikel-ter-horst.pdf

Van-der Feltz-Cornelis, C. M., Volker, D., & De Heer, E. (2010). Routine outcome monitoring: ontwikkelinge in de GGZ. Trimbos Instituut, Utrecht.

Van der Lem, R. (2013). Are depression trials generalizable to clinical practice. Something clinicians always wanted to know about RCTs, but were afraid to ask. (Dissertation Leiden University). Ipskamp Drukkers BV, Enschede.

Van Marle, H. J. C. (2012). De stagering met profielen in de forensisch psychiatrische delictanalyse. [Staging with profiles in offence analysis: diagnostic stepts in forensic psychiatry]. *Tijdschrift voor Psychiatrie, 11,* 973 - 978. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/54-2012-11-artikel-VanMarle.pdf

Van Nieuwenhuizen. C. H., Bogaerts, S., De Ruijter. E. A. W., Bongers, I. L., Coppens, M., & Meijers, R. A. A. C. tbs-behandeling geprofileerd; een gestructureerde casussenanalyse [*Profiling TBS-treatment: A structured cases analysis]*. Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) Ministerie van Veiligheid en Justitie. 2011. Retrieved from http://www.wodc.nl/onderzoeksdatabase/voor studie-onderzoek-tbs.aspx

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

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

Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Task Force in Forensic Psychiatry]. (2002). *Handleiding HKT-30 versie 2. Risicotaxatie in de forensische psychiatrie* [Guide HKT-30 version 2. Risk Assessment in Forensic Psychiatry]. Den Haag, The Netherlands: Ministerie van Justitie, Dienst Justitiële Inrichtingen.

Wetboek van Strafrecht (1881). [code of criminal law]. http://wetten.overheid.nl/ BWBR0001854/2018-01-01#BoekEerste. Accessed April, 2018.

Wing, J., Beevor, A., Curtis, R., Park, S., Hadden, S., & Burns, A. (1998) Health of the Nation Outcome Scales (HoNOS). *Britisch Journal Psychiatry, 172*: 11-18.

Yiend, J., Chambers, J. C., Burns, T., Doll, H., Fazel, S., ... Fitzpatrick, R. (2011). Outcome measurement in forensic mental health research: An evaluation. *Psychology, Crime and Law, 17 (3),* 277 – 292. doi: 10.1080/10683160903203938

CHAPTER TWO

Routine Outcome Monitoring in Forensic Psychiatry: FPC 2landen and FPC De Kijvelanden

Frida C. A. van der Veeken
Stefan Bogaerts
Jacques Lucieer

This is an adapted version of the chapter pusblished as: Van der Veeken, F.C.A., Bogaerts, S., & Lucieer, J. (2012). *Routine Outcome Monitoring in forensic psychiatry: FPC 2landen and FPC De Kijvelanden*. In T. I. Oei & M. S. Groenhuijsen (Eds.), Progression in Forensic Psychiatry. Alphen aan den Rijn: Kluwer; 2012. p 355-375.

Introduction

In 1988, Ellwood introduced outcome management in response to the "chaotic" American healthcare system, in which medical care was expensive, diverse disciplines did not share insights into patients' well-being, and the increased number of patients and complexity of medical care jeopardized sound decision-making. The effects of decisions made by patients, physicians, and healthcare executives were unclear. Patients requested more information about their health progress and possibilities; management teams and insurance companies were interested in the costs of healthcare outcomes (Ellwood, 1988). Similar concerns were voiced in mental healthcare institutions in the Netherlands, and the pressure to get more information on treatment outcomes was increasing (De Beurs & Zitman, 2007).

Ellwood described outcome management as "a technology of patient experience designed to help patients, payers and providers to make rational medical care-related choices based on better insight into the effect of these choices on the patient's life" (Ellwood, p. 1551). Outcome management or routine outcome monitoring (ROM) can be used for different purposes, but an important goal of ROM is to obtain better insight into the treatment progress of individual patients and to make rational choices in treatment by systematic measurements of patients' disease, functioning, and well-being.

In this chapter, we discuss the development of ROM in two Forensic Psychiatric Centers (FPCs): FPC 2landen and FPC De Kijvelanden. Since ROM was first introduced in Dutch mental healthcare, several expert groups have discussed ROM, and three primary performance indicators related to treatment effectiveness have been specified: 1. Change in symptom severity (measurement of symptoms related to mental health problems), 2. Change in daily functioning, and 3. Change in quality of life (subjective experience of quality of life) (Expertgroep Forensische Psychiatrie [Forensic Psychiatry Expert Group] (EFP), 2011). These indicators can also be seen as important outcome measures for psychiatric treatment.

It is of great importance to select a limited set of highly relevant instruments (Fitzpatrick et al., 2010) that restrict the work load for patients and therapists (Stinckens, Smits, Claes, & Soenen, 2012). An extensive set of instruments is very time intensive, does not guarantee success (Stinckens et al., 2012), and may even produce a plethora of information that counteracts the feasibility of ROM (Nugter & Buwalda, 2012). ROM assessments should be conducted at relevant moments in treatment (Drieschner, Hesper, Marrozos, & Hout, 2011): at the start of treatment, during multidisciplinary treatment evaluation, and at the end of treatment. An assessment at the beginning of treatment can contribute to the treatment indication and plan and to the diagnostic process (Carlier et al., 2012); assessments during treatment and at the end may offer important information about treatment progress (Nugter & Buwalda, 2012; Bickman, Kelley, Breda, de Andrade, & Riemer, 2011).

The application of ROM in Dutch forensic psychiatry evolves more slowly than in other mental healthcare sectors. In FPCs, patients are hospitalized involuntarily (TBS order). A TBS order is imposed by the court when offenders committed a crime with a minimum sentence of four years as a consequence of a mental disorder, and with a high risk to reoffend (Van Nieuwenhuizen et al., 2011). The offenders' responsibility for the crime is then considered diminished to fully absent (De Ruiter & Trestman, 2007), and they will be admitted for a minimum of two years with or without prior imprisonment. Every one or two years, the court will review the necessity of admission and decide whether admission should continue or discontinue. The two main goals of a TBS order are to protect society and to rehabilitate patients back into society. Treatment largely focuses on the criminogenic needs. Risk assessment is very important in clinical forensic psychiatry (Robertson, Barnao, & Ward, 2011).

In the Netherlands, the risk of recidivism is assessed by weighing risk factors and protective factors. The Historical Clinical Future-30 items (HKT-30; Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Taskforce in Forensic Psychiatry], 2002), or the Historical Clinical Risk-20 items (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) (supplemented by 9 dynamic HKT-30 items)* (Nagtegaal, 2010) are assessed annually by order of the law. When psychopathy and/or sexually delinquent behavior are present, the Psychopathy Checklist-Revised (PCL-R) and/or the Sexual Violence Risk-20 (Sexual Violence Risk-20 items (SVR-20); Boer, Hart, Kropp & Webster, 1997)** should also be scored. With these instruments, the main risk factors and protective factors are covered. Other instruments are under development, but the added value of these instruments for different patient groups has not yet been demonstrated.

The HKT-30 was developed specifically for Dutch forensic psychiatry as an alternative to the HCR-20. The instrument was developed by the 13 Dutch FPCs and includes Historical (H), Clinical (K), and Future (T) factors (HKT-30; Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Taskforce in Forensic Psychiatry], 2002). The HKT-30 is an instrument that has more dynamic items than the HCR-20 (De Beurs & Zitman, 2007, Spreen et al., 2009; Lammers, 2007). The HKT-30 is currently under revision, a process that is funded by the Department of Correctional Institutions of the Ministry of Justice and Security*. The most important risk factors and inadequate protective factors for future risk of recidivism include a lack of problem insight, medication incompliance, substance abuse, impulsivity, hostility, inadequate coping skills, and problematic social networks (HKT-30; Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Taskforce in Forensic Psychiatry], 2002).

* The HKT-30 and HCR-20 have been revised, and FPCs currently use the HKT-R of HCR-20^{V3}. All FPCs must assess the 14 clinical HKT-R items. ** The SVR-20 has been replaced by the Static-99R, Stable-2007, and Acute-2007 in FPC De Kijvelanden.

The underlying rehabilitation theories of Routine Outcome Monitoring

A rehabilitation theory is a theory composed of values, principles, etiological assumptions, and clinical guidelines (Ward, Collie, & Bourke, 2009). The Risk-Need-Responsivity model (RNR) and the Good Lives Model (GLM) are rehabilitation theories that emphasize the importance of risk factors and protective (strength-based) factors (Andrews, Bonta, & Wormith, 2011; Robertson et al., 2011). The RNR model comprises three core principles (Bonta & Andrews, 2007). The Risk principle assumes that risk of recidivism can be reduced when treatment intensity is matched to risk of recidivism. The Need principle focuses on the criminogenic needs or dynamic risk factors. The Responsivity principle emphasizes the importance of personal strengths, pathology, and personality factors that strengthen the treatment effect. The RNR model also emphasizes the General Personality and Cognitive Social Learning (GPCSL) perspective on criminal behavior. Criminal behavior can be influenced by personal expectations of rewards and consequences, and it can be empowered by internal and external motivation and by the presence of a target (Bonta & Andrews, 2007).

The Good Lives Model (GLM) was developed by Ward and Stewart (2003) to address the limitations in the RNR model, such as limited treatment responsivity or lack of motivation (Robertson et al., 2011). The GLM focuses on the development of skills and sources in order to live a better life (Ward & Stewart, 2003), with the aid of a good lives plan and primary human goods such as happiness, knowledge, and friendship (Ward & Stewart, 2003; Ward & Gannon, 2006) that are attained in a socially accepted manner (Ward & Gannon, 2006). Secondary or instrumental goods are manners to gain and maintain primary human goods.

Risk factors are considered to be internal or external obstacles that counteract primary human goods. The goal of the GLM is to complement the RNR model (Ward & Stewart, 2003) and to show patients a different way of life, one not involving a criminal lifestyle (Ward & Gannon, 2006).

Routine Outcome Monitoring in forensic psychiatry: patient heterogeneity, general instruments, and specific patient populations

Besides its focus on the three performance indicators of daily functioning, problem reduction, and quality of life, an important fourth indicator in forensic psychiatry must be addressed: the reduction of future recidivism (EFP, 2011). In line with the RNR model and the GLM, both risk factors and protective factors can be emphasized. In Dutch forensic

psychiatry, ROM is still in its infancy, and valid and reliable studies on treatment effectiveness, therefore, are quite scarce (Drieschner, Hesper, & Marrozos, 2010). Reliable and valid insights into treatment in terms of progress, stagnation, or deterioration, however, is complicated by the heterogeneity of the population and the selection of instruments (general and specific). Below, we will elaborate on patient heterogeneity and general ROM instruments. Further sections describe instrument sets for social therapeutic patients, mildly intellectual disabled patients, and patients residing in a supportive treatment environment.

Patient heterogeneity

Although all forensic patients have been legally defined, the group of patients is very heterogeneous (Robertson et al., 2011). In FPC 2landen and FPC De Kijvelanden, three different main groups can be identified: the social therapeutic group, mainly patients with a cluster-B personality disorder as their primary diagnosis; the supportive group, mainly patients with a psychotic disorder; and the third group, mildly intellectually disabled patients.

The social therapeutic group resides in a therapeutic environment in which group functioning plays an important role. The modification of negative and anti-social personality traits are of interest here (Van Nieuwenhuizen et al., 2011).

Patients in the supportive group are vulnerable patients with a psychotic disorder who require more guidance, training, counseling, and structure. However, they often also suffer from co-morbid personality problems, mostly with anti-social features (Van Nieuwenhuizen et al., 2011). If ROM is to benefit treatment, it is important to observe patients' strengths and weaknesses and to control for too high expectations. In patients with schizophrenia, it is important not only to reduce psychotic symptoms but also to improve general functioning (Mulder et al., 2010).

Patients with a Mild Intellectual Disorder (MID) often face limited adaptability and learning capacities. The American Association on Intellectual and Developmental Disabilities (AAIDD, 2010) identifies three components of adaptive behavior: 1. conceptual skills (language, reading, time and number concepts); 2. social skills (interpersonal skills, responsibility, wariness); and 3. practical skills (daily living activities, occupational skills, healthcare use).

With regard to ROM, it is important to mention that self-report instruments are not suitable for all patients. Particularly in the last group, cognitive and communicative limitations make it difficult to question patients about their subjective experiences (De Baaij, Hoekman, Volman, & Zaad, 2006). Socially desirable answers are very likely to be given and multiple choice questions are likely to be misunderstood (De Baaij et al., 2006). Some self-report lists are available for supportive patients with psychotic problems but not recom-

mended for severely confused patients (Nugter & Buwalda, 2012). Questioning by significant, closely related persons can be an alternative, but research has shown that inter-rater reliability between patients and significant persons is low (De Baaij, Hoekman, Volman, & Zaad, 2006) and that not all patients in forensic centers have a social network with whom they maintain a close or therapeutic relationship.

Considering differences in treatment needs and the strengths and weaknesses of these three patient groups, standardized instruments must be used to focus on the most significant factors that give an indication of treatment progress. Therefore, three measurement sets have been composed for the different patient environments. Beside the use of generally accepted instruments proposed by the ROM expert group and adopted by the Quality in Forensic Care working group, other instruments are needed in FPC De Kijvelanden and FPC 2landen to measure treatment progress in forensic psychiatry in general and in the three different patient groups in particular.

The HKT-30 as a general instrument

At least once a year, the HKT-30 (and SVR-20)* is conducted in FPC 2landen and FPC De Kijvelanden and included in ROM assessment; the SVR-20 is conducted once a year for sex offenders only. The predictive validity and reliability of the HKT-30 are reasonable (AUC = .72, ICC = .77), and an ICC > .75 (Blok, De Beurs, Ranitz, & Rinne, 2010) can be seen as good (Hildebrand, Hesper, Spreen, & Nijman, 2005). Because the dynamic items may change over short periods (Goethals & Van Marle, 2012), it is also important to choose a ROM tool that focuses on the risk of recidivism indicator and shows progress over shorter periods. However, ROM is not meant to be a risk assessment, but it could be used to monitor risk factors. In the context of ROM, moreover, three-monthly or half-yearly measurements are the maximum attainable. It remains the responsibility of the treatment team, therefore, to observe, discuss, and treat changes in dynamic factors.

The Health of the Nations Outcome Scales

The Health of the Nations Outcome Scales (HoNOS; Wing et al., 1998) maps the mental health and social functioning of psychiatric patients (Mulder et al., 2004). The HoNOS consists of a psychotic and a neurotic dimension and is independent of language and pathology. Institutions that already use the HoNOS report that it contributes to the evaluation of treatment and the measurement of changes in important life domains (Mulder et al., 2004).

A Dutch study of 559 patients shows a reasonable to good reliability (alpha = .78, ICC = .92). The HoNOS also appears to be sufficiently valid (Mulder et al., 2004). The HoNOS

*Currently the HKT-R, and, when indicated, the Static-99R, Stable-2007, and Acute-2007.

provides an understanding of the seriousness of problems in different areas and at individual and group levels. Scoring is not time-consuming (± 15 minutes), and optimal scoring is obtained by establishing a consensus score between someone who knows the patient well and an independent investigator.

The HoNOS consists of 12 items on a five-point scale, divided into the four subscales of behavior, limitations, symptomatology, and social problems. Three extra items have been added: *maniform disinhibition, treatment motivation,* and *medication adherence*. The HoNOS secure was developed for the benefit of secure settings. The last two items require some explanation prior to the measurement (Dickens et al., 2007). This instrument contains seven additional items that indicate the need for security and risk management, covering physical, relational, and standard management at present and future points in time (Dickens, Sugarman, & Walker, 2007). Research on the HoNOS secure shows reasonable validity (Cronbach's alpha = .73). The inter-rater reliability ranges from moderate to good (ICC = .39 - .88). The first five items have a good inter-rater reliability (ICC> .64), but the last two items show a moderate inter-rater reliability (ICC =. 39 - .53).

Routine Outcome Monitoring for social therapeutic forensic patients

A social therapeutic environment treats patients with anti-social lifestyles and clinical scores on Cluster B personality traits, such as impulsiveness, hostility, substance abuse, and lack of empathy. The Instrument for Forensic Treatment Evaluation (IFTE) and The Manchester Short Assessment of Quality of Life (MANSA) have been selected next to the general use of the HoNOS.

Risk of recidivism: the Instrument for Forensic Treatment Evaluation (IFTE)

In a social therapeutic environment, the Instrument for Forensic Treatment Evaluation (IFTE) can be used to measure treatment changes. This instrument is based on the dynamic factors from the HKT-30 (and later HKT-R) and some items from the ASP NV (Schuringa, 2010).

The IFTE consists of 22 items. Factorial structure research shows three components: *problem behavior, protective behavior,* and *resocialization skills*. The IFTE uses a seventeen-point scale to indicate progress on the items (Schuringa, 2014). The IFTE is independently scored by multiple therapists, which has the advantage that patient functioning is assessed from multidisciplinary points of view and that discrepancies can be discussed (Schuringa, 2010). Besides the IFTE practitioners' report, the instrument also contains a patient self-report version (Schuringa, Bogaerts, & Spreen, 2011). Patients score the same items which are, of course, translated in a patient transcript. Patients can score their own behavior, which can of-

fer very useful information for therapeutic conversations and for the future treatment plan. Figure 1 and figure 2 illustrate a fictional patient report based on the IFTE.

IFTE report

The IFTE is scored by three clinicians. The ROM output is then compiled in a report summarizing the scores, based on the method developed by Schuringa (2011). The report describes the general treatment outcomes at different times on three factors: *protective skills, problem behavior,* and *resocialization skills*. The graph (see Figure 2) further displays the treatment direction on the three factors. The separate items are displayed through bar charts and a narrative explanation of the scores. A higher score represents positive behavior, a lower score negative behavior. For example, the highest score on the item *drug use* stands for "no use of substances in the previous period," and the highest score on the item *coping skills* stands for "adequate coping skills."

The bar chart displays patients' last and previous assessments. The narrative report shows the number of therapists who scored the survey in the "number" column. The "change" column represents changes compared to the last assessment, with the "▲" symbol indicating positive change (progress) and "▼" representing negative change (decline). The "agreement" column specifies the level of agreement between the three therapists for every score (Schuringa, Heininga, & Spreen, 2011), with moderate or low agreement representing disagreement between therapists. The "item" column relates the item of the IFTE, and the "patient" column the actual assessment of a patient's behavior. This report evaluates treatment goals for the upcoming period (six months) and allows us to define patients' risk factors and protective factors. When a risk factor occurs, the item will be displayed in bold. When an item is defined as a goal, the item will be displayed with a grey background (see Figure 1):

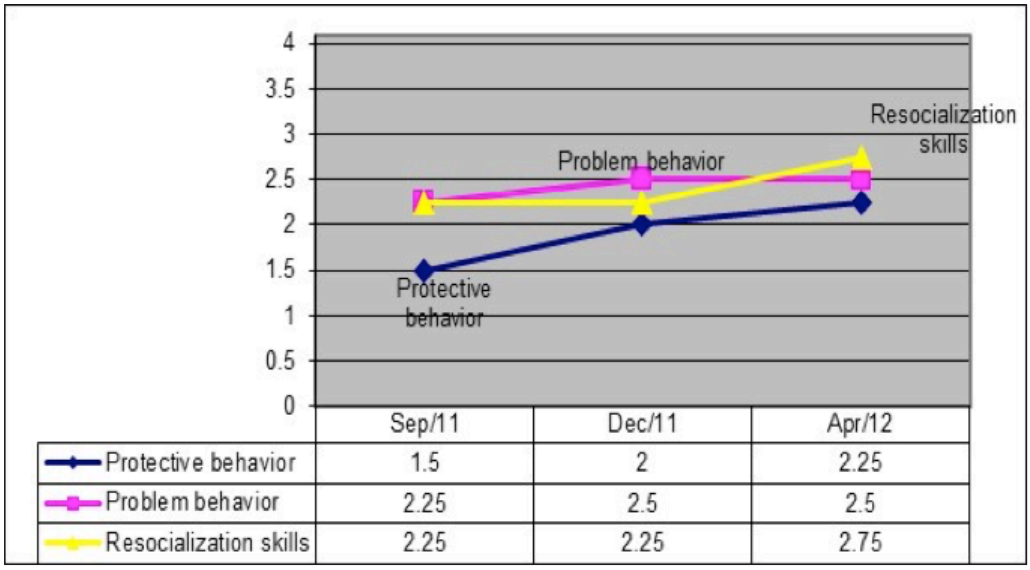
Figure 1

Goal and risk factors

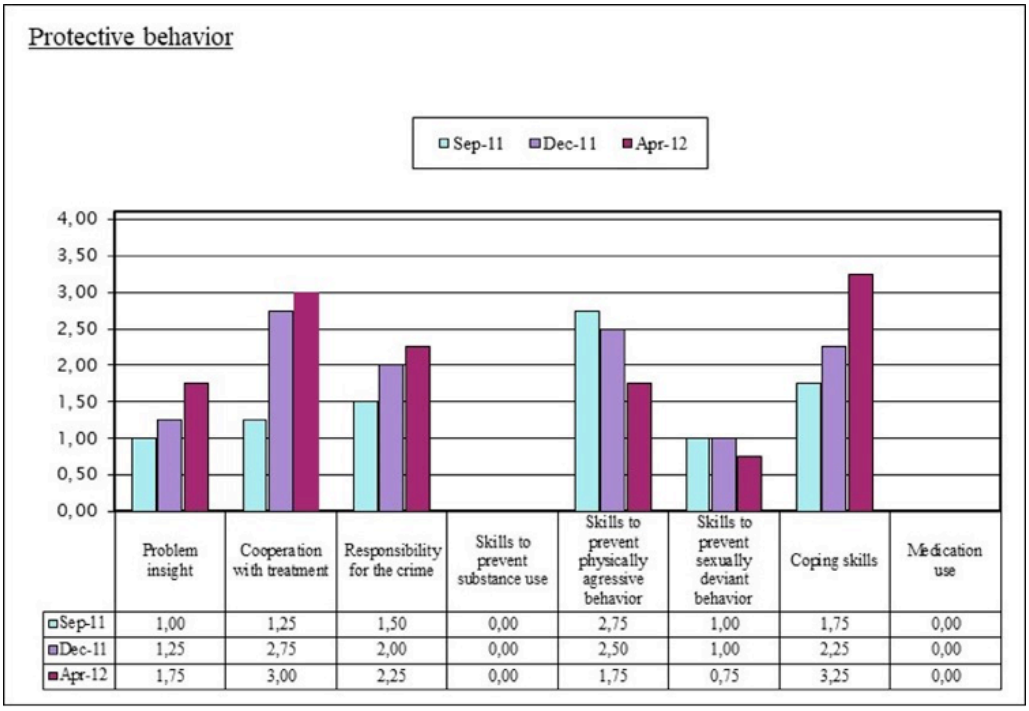
	=Treatment goal
Bold	**= Risk factor**

Figure 2
IFTE report (Schuringa, 2011)

Summary

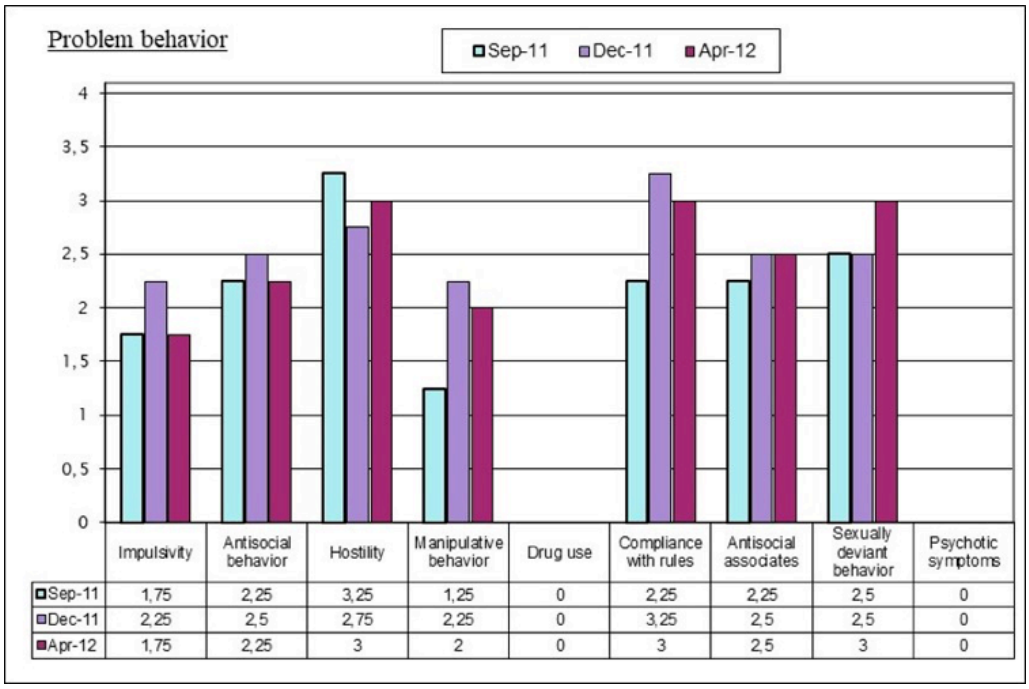


	Sep/11	Dec/11	Apr/12
Protective behavior	1.5	2	2.25
Problem behavior	2.25	2.5	2.5
Resocialization skills	2.25	2.25	2.75



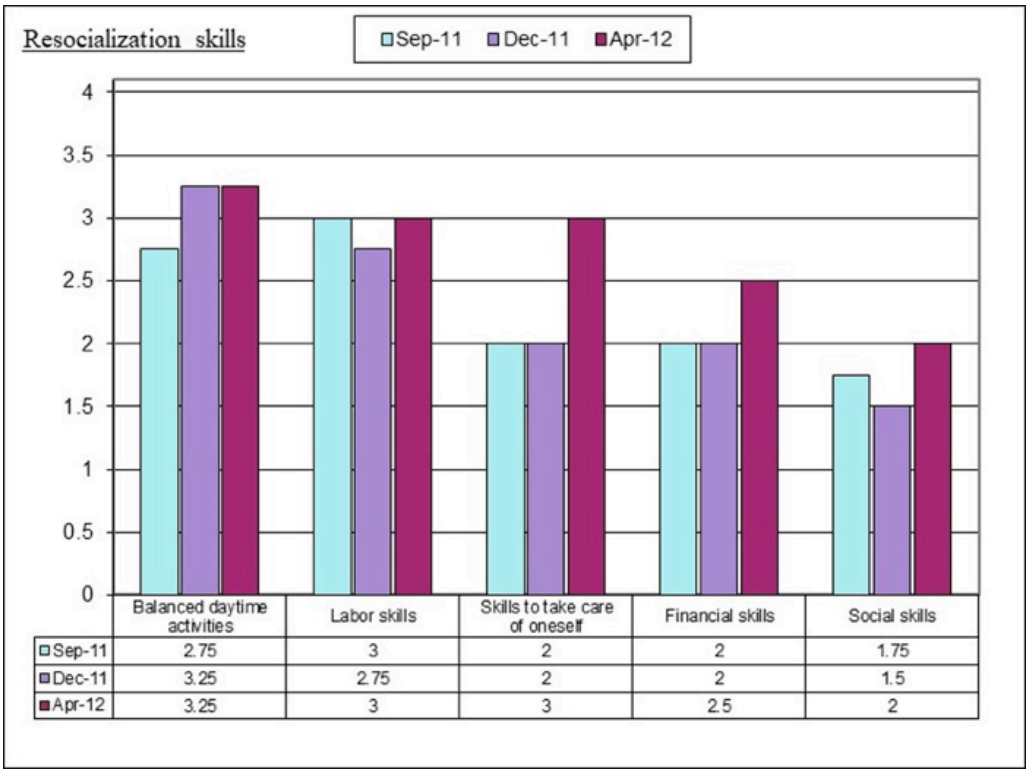
Protective behavior				
Item	**Patiënt**	Agreement	Change	Number
Problem insight	Does have problem awareness, but does not behave accordingly.	high	▲	3
Cooperation with treatment	Cooperates actively, sometimes obstinately or indifferently.	high	-	3
Responsibility for the crime	Partly takes responsibility for the offense, but also partially blames others or circumstances.	high	-	3
Skills to prevent substance use				0
Skills to prevent physically aggressive behavior	Has limited skills to prevent physically aggressive behavior.	**moderate**	▼	2
Skills to prevent sexually deviant behavior	Has limited skills to prevent sexually transgressive behavior.	high	▼	3
Coping skills	Has sufficient coping skills, but stability is doubtful during persisting problems.	high	▲	3
Medication use				0

	=Treatment goal
Bold	**= Risk factor**



	Impulsivity	Antisocial behavior	Hostility	Manipulative behavior	Drug use	Compliance with rules	Antisocial associates	Sexually deviant behavior	Psychotic symptoms
Sep-11	1,75	2,25	3,25	1,25	0	2,25	2,25	2,5	0
Dec-11	2,25	2,5	2,75	2,25	0	3,25	2,5	2,5	0
Apr-12	1,75	2,25	3	2	0	3	2,5	3	0

Problem Behavior				
Item	**Patient**	Agreement	Change	Number
Impulsivity	Shows some impulsivity in behavior, has the ability to control behavior to some extent.	high	▼	3
Antisocial behavior	Does not consider others sometimes, and can cross boundaries or cause conflicts.	high	▼	3
Hostility	Hostile behavior, showing itself in occasional irritation, cursing, anger, and/or making negative statements.	high	-	3
Manipulative behavior	Takes advantage of others in a manipulative way sometimes.	high	-	3
Drug use				0
Compliance with rules	Adheres to all rules and agreements, but makes moderate verbal protest against rules.	**Moderate**	-	3
Antisocial associates	Has some people in his network that approve of violence or a criminal lifestyle, but these do not influence him.	high	-	3
Sexually deviant behavior	Shows inappropriate behavior, such as staring or making inappropriate comments.	**Moderate**	▲	3
Psychotic symptoms				0



Resocialization skills				
Item	**Patiënt**	Agreement	Change	Number
Balanced daytime activities	Often has a balanced daily routine.	high	-	3
Labor skills	Has some problems with working skills or daily routine.	high	-	2
Skills to take care of oneself	Minor deficiencies in self-care skills, does not necessarily need support. **This was**: Show deficiencies in self-care skills; patient needs and accepts support.	high	▲	3
Financial skills	Shows deficiencies in financial skills; patient needs and accepts support.	high	-	2
Social skills	Sometimes gets in trouble due to lack of social skills. **This was:** Regularly gets in trouble due to lack of social skills.	high	▲	3

Figure 2 illustrates a fictional report of a patient who committed a sexual offense. We observe some progress on the factor *protective behavior*, as the item *problem insight* increases. However, problem insight still remains insufficient, which is considered a risk factor. The factor *problem behavior* indicates limited skills to prevent sexually deviant behavior. The patient can be impulsive, shows some verbal hostility, and does not consider others. The factor *resocialization skills* displays limited progress on the item *social skills*. As the Figure shows, future treatment will focus on several factors. The patient completed the Manchester

Assessment of Quality of Life and the IFTE self-report. The Health of the Nation Outcome Scales is rated by the head of "risk management and treatment." These scores are not displayed in this Chapter.

Quality of life: The Manchester Short Assessment of Quality of Life (MANSA)
The MANSA (Priebe, Huxley, Knigth, & Evans, 1999) uses a short interview (± 15 minutes) or a structured questionnaire. The MANSA is a shortened version of the Lancashire Quality of Life Profile (LQLP; Oliver, Huxley, Bridges, & Mohamad, 1996), which has shown to be sufficiently reliable in several countries. However, the interviews take a very long time, some questions are not clearly defined, and questions regarding sexuality are missing. The MANSA has been developed to address these shortcomings (Priebe et al., 1999).

The MANSA has three parts: static personal characteristics, such as age and gender; personal characteristics, such as education; and subjective items, such as psychological health, social and intimate relationships, sexuality, housing, and financial situation. Items are scored on a seven-point scale (Van Nieuwenhuizen, Schene, & Koeter, 2000). Research shows a high correlation of the MANSA with the LQLP. The internal consistency is reasonable (alpha = .74). The instrument, however, is related to everyday life items and not specifically to symptoms and psychopathology, which, of course, can influence the scores (Priebe et al., 1999). In our ROM tool, clinical psychopathology could be abstracted from the electronic patients files, the IFTE (self-report and practitioners' version), and the HoNOS.

Routine Outcome Monitoring for mildly intellectual disabled forensic patients

The dynamic risk outcomes scales (DROS), the Intellectual Disability Quality of Life (ID-QOL) and Social functioning and adaptive behavior: the Social Self-Reliance Scale (SRZ-P) has been selected for the mildy intellectual disabled patient group

Risk of recidivism: the Dynamic Risk Outcome Scales (DROS)
In a mildly intellectually disabled group, the DROS is used to measure risk factors. The DROS measures 15 dynamic risk factors based on the dynamic risk factors of the HKT-30. The factors are divided into 43 items and scored on a five-point scale, except for factor 15 (social network), which uses a combination score. Research (Drieschner, 2010) has shown that the internal consistency of the different items is moderate to good (.69 - .90) with a very good overall internal consistency (.92). In a study with 82 patients, the DROS shows

an effect size, Cohens' d = .60 after 12 months and Cohens's d = .93 after 24 months (N = 34) and appears to be more dynamic than the HKT-30 (Cohen's d =. 33 and .66). This change is considered valid on the basis of the correlations with the K-items on the HKT-30 after 12 and 24 months (r = .74, and r = .69; Drieschner, 2010).

The DROS is conducted biannually, preferably by practitioners who know patients well and have sufficient current observations to evaluate patients' current cognitions, behavior, and skills. Assessors must also possess sufficient clinical knowledge of the population and have to work accurately. In settings in which the DROS is already used, the instrument is judged as providing added value in the treatment and evaluation of intellectually disabled patients (Drieschner, 2010). Figure 3 illustrates a fictional patient report derived from the DROS.

Dynamic Risk Outcome Scales Report

Figure 3 shows 15 risk factors and items, based on the schedule developed in Trajectum Hoeve Boschoord (Hout, 2011, Drieschner, & Hesper, 2011). Table 1 shows codes based on codes that were also developed in Trajectum Hoeve Boschoord (Drieschner, 2010). In FPC 2landen, the DROS is scored in consensus between two therapists. In the "Periodic goals" section (see Figure 3), scores are reported for the two last measurements (Score 1 and Score 2). The column entitled "Achieved" displays differences between the two measurements with: "+" (positive change/progress); "–" (negative change/decrease); or "=" (no change/stabilization) (see Table 1). A progression of 1.15 on a scale score is considered meaningful (Drieschner, 2010). When a factor is scored with a "+" twice (progress of two on a scale), we assume a meaningful progress on that factor.

The "Treatment necessity" section displays which items are considered to be risk factors (R) and other treatment issues (A) (see Table 1). The "UP" column displays the treatment perspective, and the "Treatment" column presents the necessity of treatment progress for the benefit of the treatment perspective. A single exclamation mark ("!") means that progress is preferred and may influence the treatment perspective. Double exclamation marks ("!!") indicate that progress is essentially required (see Table 1) and that, without improvement, serious problems could occur in the current treatment perspective (Drieschner, 2010).

In the "Periodic goals" section, specific goals for the upcoming period can be highlighted. During treatment evaluation, treatment changes between previous and current evaluations can be observed and treatment goals can be involved. In the "Goal" column, the "+" symbol stands for progress and "(+)" for medium progress. The "S" code stands for stabilization of a higher (4 or 5) score, and the "-" code means that no improvement is expected in the next period (see Table 1) (Drieschner, 2010).

Table 1

Codes in DROS schedule (Drieschner & Hesper, 2011)

"Achieved" column	
+	Positive progress
-	Negative progress
=	No change
"Goal" column	
+	Enough progress to attain a higher DROS score
(+)	Progress, though not necessarily enough in order to attain a higher DROS score
S	Stabilization of a higher DROS score
-	No improvement expected in the upcoming period
"Risk factor" column	
R	Risk factor
A	Treatment issue, but not necessarily a risk factor
"Treatment goal" column	
!	Desirable progress
!!	Essentially required progress

Figure 3

DROS schedule

Dynamic Risk Outcome Scales
K. Drieschner, B. Hesper (2011).

Patient X

Department x

			Treatment necessity			Periodic goals			
Risk variable		**Item**	**Risk Factor**	**U P**	**Treat-ment**	**goal**	**Score 1**	**Achieved**	**Score 2**
Problem awareness/	1.1	Responsibility	R		!	(+)	3	+	4
problem insight	1.2	Risk awareness, risk factors, and danger signals					4	=	4
	1.3	Acceptance of one's own limitations/ disabilities	R		!!	+	2	+	3
Attitude towards	2.1	Recognition current necessary prof. care				+	3	+	4
current treatment or supervision	2.2	Attitude towards current professional care	A	R	!		3	+	4
				I					
Ideas and expectations of	3.1	Necessary care	A	B	!	+	3	+	4
post-treatment situation	3.2	Remaining elements	A	W	!	S	4	=	4
Cooperation with treatment/supervision	4.1	Commitment to change				S	4	+	5
	4.2	Medication compliance	R		!!	+	2	=	2
	4.3	Rule conformity	A		!!		3	+	4
	4.4	Transparency/openness	A		!		2	+	3

Criminogenic attitudes	5.1	Attitude towards physical violence	A	!	S	4	=	4
	5.2	Egocentrism	R	!		3	+	4
	5.3	Attitude towards antisocial behavior	R	!!	(+)	3	+	4
	5.4	Attitude towards prosocial conventions				4	-	3
Coping skills	6.1	Coping with conflictual interactions	A	!		3	-	2
	6.2	Coping with risk urges	R	!!	(+)	3	=	3
	6.3	Coping with remaining stressors	R	!!		2	+	3
Hostility	7.1	Hostile perception and attitude	R	!		4	-	3
	7.2	Resentment and revenge				4	+	5
	7.3	Resentments				5	=	5
Sex-related cognitions	8.1	Sexual preoccupation	R	!!	+	2	+	3
and sexually transgressive	8.2	Sexual misinterpretation				5	=	5
Behavior	8.3	Sexually transgressive behavior	R	!!	+	1	+	3
Impulsivity	9.1	Thoughtlessness				5	=	5
	9.2	Sensationalism	R	!	(+)	3	=	3
Maintaining stabilizing	10.1	Maintaining prosocial contacts	A	!		3	=	3
and prosocial structures	10.2	Maintaining meaningful daily activities	A	!		3	+	4
Self-reliance	11.1	Hygiene	A	!	(+)	1	+	2
		Nutrition	A	!		3	=	3
		Circadian rhythm	A	!	(+)	3	-	2
	11.2	Literacy	A	!	(+)	2	+	3
		Count			(+)	3	=	3
		Sense of time/telling time				3	+	4
Social skills	12.1	Basic social skills and behavior			(+)	3	=	3
and behavior	12.2	Sub-assertiveness	A	!		3	=	3
	12.3	Collaboration skills	A			2	+	4
Addiction,	13.1	Yearning for substances/gambling	R	!!	(+)	3	=	3
substance use	13.2	Idealizing substances/gambling				5	=	5
and gambling	13.3	Accountability for substances/ gambling	R	!	(+)	3	+	4
Psychotic symptoms	14.1	Delusions/delusional ideas	A	!		4	=	4
		Hallucinations				5	=	5
	14.2	Psychotic vulnerability	A			3	+	4
Social network	15.1						=	

This description is based on a fictional patient with a mild intellectual disability who has committed a sexual offense. The patient resides permanently in a forensic psychiatric center (high-risk security). The patient received previous treatments in psychiatry and detention. Important risk factors, as presented in Figure 3, are *problem awareness and problem insight; medication compliance; copings skills; sex-related cognitions and sexually transgressive behavior; sensationalism; egocentrism; substance use; attitude towards antisocial behavior etc.* (see Figure 3).

For evaluation purposes, required treatment progression is expressed by several items such as *acceptance of one's own limitations*; *medication compliance*; *attitude towards antisocial behavior*; *sex-related cognitions;* and *sexually transgressive behavior,* etc. The DROS displays progress on several scales, especially on the risk factors of *sex-related cognitions; sexually transgressive behavior;* and *cooperation with treatment/supervision*. The item *medication compliance* remains problematic, and *coping skills* and *self-reliance* are also problematic. The *hostile perception and attitude* item shows an increase of negative behavior in the past period.

In the following evaluation period, the focus could lie on stress handling, self-care skills training, and risk management with a focus on hostility. The patient completed the Intellectual Disability Quality of Life interview, and therapists completed the Social adaptability Scale-P (SRZ-P). The Health of the Nation Outcome Scales and Health of the Nation Outcome Scales Secure were rated by the head of "risk management and treatment." These scores are not displayed in this Chapter.

Quality of Life: the Intellectual Disability Quality of Life (IDQOL)

IDQOL is based on the DUCATQOL, which measures the overall quality of life (QOL) of hospitalized children. The IDQOL requires an affective evaluation of the patients' various aspects of daily life or work and is intended for patients with an IQ higher than about 60. By using simple images that convey satisfied or dissatisfied emotions, patients may give a subjective (affective) judgment about their own life in three domains: social environment, environment in general, and life-related aspects.

The internal consistency of the scales (alphas respectively .77, .73, and .80) and the total instrument (alpha = .87) is satisfactory to high (Douma, Kersten, Koopman, Schuurman, & Hoekman, 2001; Hoekman, Douma, Kersten, Schuurman, & Koopman, 2001). An advantage of the IDQOL-16 is that the questionnaire is relatively short (16 items) and that people with intellectual disabilities have contributed to its construction phase (Douma et al., 2001). For the intellectually disabled group, the IDQOL-16 is used, with the extra subscale of satisfaction with medication, derived from the IDQOL-44. The list is administered by independent researchers who are not involved in treatment. During the interview, it is important to verify patients' understanding of the questions and whether they can connect their own experiences to the questions (Douma et al., 2001).

Social functioning and adaptive behavior: the Social Self-Reliance Scale (SRZ-P)
Patients with an MID experience more difficulties in conceptual, practical, and social domains. The Social Self-Reliance Scale (SRZ-P, sociale redzaamheidsschaal; Kraijer & Kema, 1994) has been developed to measure the independency skills of mildly intellectually disabled patients in four areas: self-help (such as table manners), communication (such as understanding other people), socialization (such as use of public services), and occupation (work and leisure time). The instrument measures independent behavior in the community (NJI, n.d.).

The SRZ-P is rated as "good" by the Dutch Committee on testing (COTAN). Research shows that patients who exhibit problem behavior need more guidance and, therefore, obtain lower scores on self-reliance in the field of living (Mulder et al., 2010). Research also shows that patients with low labor performance have more difficulty with adaptive behavior. (Gravesteijn, 2008). The internal consistency of the SRZ-P is good to excellent (.92 and .94) for the total scale; the subscales vary between .80 and .91. The inter-rater reliability for the total scale was good (r = .93) for the subscales; the correlation ranges between .74 and .93 (NJI, n.d.). The SRZ-P can contribute to evaluations of leave requests, as adaptive behavior is tested and practiced during patients' leaves, and to patients' functioning on living skills in the clinic.

Routine Outcome Monitoring for supportive forensic patients

For the benefit of ROM for patients residing in a supportive treatment milieu, the IFTE, The Positive and Negative Syndrome Scale (PANSS), the MANSA and the SRZ-P have been selected.

Risk of recidivism: the Instrument for Forensic Treatment Evaluation (IFTE)
For the use of the IFTE for patients with a psychotic vulnerability, it is preferred that the IFTE-sr should be scored under supervision. It is important, however, to examine patients' mental state, for when patients exhibit symptoms such as delusions or hallucinations, they are not designated to fill in the IFTE. As mentioned earlier, the IFTE is considered to be an adequate instrument for assessing treatment progress for patients residing in a supportive treatment environment. Psychometric qualities of the IFTE are described in **chapter 4** of this thesis.

The Positive and Negative Syndrome Scale (PANSS)

In a supportive treatment environment, a significant number of patients regularly suffer from psychotic episodes according to the DSM-IV-TR (American Psychiatric Association (APA), 2000). Studies show that a psychotic disorder such as schizophrenia may be changeable (Van Os, & Kahn, 2007). As it is important to reduce psychotic symptoms, it is also imperative to measure changes in these symptoms effectively (Mulder et al., 2010). The PANSS has been recommended in the consultation document "ROM for patients with severe psychiatric illness" (EPA) (ROM expert group adults EPA, 2011).

The PANSS' 30 items give a good picture of positive and negative psychotic symptoms for typological and dimensional assessment and gauge the relationship between positive and negative symptoms and general psychopathology (Mulder et al., 2010). The Dutch national remission working group (Mulder et al., 2007) developed the PANSS remission Tool. The selected PANSS items have been compiled by an international group of researchers (Van Os & Kahn, 2007), who selected eight major symptoms, including three psychotic symptoms, three negative symptoms, and two disorganization symptoms (Mulder et al., 2007). This instrument focuses on patients' personal development and can be conducted in little time. We include this instrument, therefore, in the ROM-tool for treatment evaluation.

Social functioning and adaptive behavior: the Social Self-Reliance Scale (SRZ-P)

We already discussed the SRZ-P in relation to mildly intellectually disabled patients. Though this instrument was originally developed for patients with an MID, it can also be used to measure treatment progress in psychotic patients with or without an MID. Patients with psychotic disorders also show problems in the independency skills domain (Mulder et al., 2010).

Quality of life: The Manchester Short Assessment of Quality of Life (MANSA)

The MANSA can be used to measure quality of life in patients with severe psychiatric disorders such as schizophrenia and schizophrenia-related sub classifications, such as the paranoid, disorganized, catatonic, undifferentiated, and residual types (DSM-IV-TR; Mulder et al., 2010). It is very important to consider whether patients are mentally able to fill in the questionnaire.

Conclusions

The main goals of forensic psychiatric centers are to prevent recidivism, to protect society, and to aid rehabilitation. This can only be realized when multidisciplinary treatment-related sub-goals are achieved, such as psychiatric, psychological, social, nursing, and educational goals. Treatment-related goals should be seen as moderators that contribute to reducing recidivism.

Three instrument sets have been discussed for ROM assessment of social therapeutic, supportive, and mildly intellectually disabled patient groups in Forensic Psychiatric Centers. We advise that three basic instruments could be part of the forensic ROM-tool, namely the HKT-30 and its revised version, the HKT-R, the IFTE, and the DROS.

However, knowledge and recommendations from adjoining clinical domains with much more ROM experience cannot be ignored. At FPCs 2landen and De Kijvelanden, therefore, we decided to involve specific patient-related instruments and selected a set of ROM tools for psychotic, personality-disordered, and mildly intellectually disabled patients.

The primary purpose of ROM is to identify treatment quality and to inform practitioners periodically about the stagnation, deterioration, or improvement of a patient's progress on risk/protective factors in various life domains. ROM offers practitioners tools to gain insights into risk profiles and provides handles to adjust treatment if indicated. ROM is also intended to inform patients, and the involvement of patients in ROM procedures provides information that can be used in therapeutic conversations.

ROM is also used as a benchmark instrument. In recent months, however, there has been a lot of controversy on benchmarking. In our vision, the therapeutic relationship is central, and the patient-practitioner-institution triangle shapes this therapeutic relationship. Benchmarking, therefore, serves goals other than primary treatment purposes.

ROM results also provide research opportunities. As we mentioned earlier, only a limited number of studies into the effectiveness of forensic psychiatric treatment have been conducted, and most studies have shown that the number of observations is often very small. Some forensic psychiatric centers already use ROM outcomes for the purpose of clinical decisions. Recently, De Vries and Spreen (2012) studied the use of the HKT-30 to underpin decision-making in treatment, such as the start of a rehabilitation process. The combination of the factors substance use, impulsivity, and lack of empathy appears to predict violations in the rehabilitation process, which was also found in the study of Bogaerts, Spreen, Horváth, Polak, and Cima (2012).

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders – DSM-IV-TR.* Arlington, VA: American Psychiatric Association.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2011). Does adding the good lives model contribute to effective crime prevention? *Criminal Justice and Behavior, 38,* 735 – 755. doi: 10.1177/0093854811406356

Bickman, L., Kelley, S. D., Breda, C., De Andrade, A. R., & Riemer, M. (2011). Effect of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial. *Psychiatric Services, 62,* 1423-1429). doi: 10.1176/ appi.ps.002052011

Blok, G. T., De Beurs, E., De Ranitz, A. G. S., & Rinne, T. (2010). Psychometrische stand van zaken van risicotaxatie-instrumenten voor volwassenen in Nederland. [The current psychometric state of risk assessment scales for adults in the Netherlands]. *Tijdschrift voor de Psychiatrie,* 5, 331- 341. Retrieved from: http://www. tijdschriftvoorpsychiatrie. nl/assets/articles/TvP10-05_p331-341.pdf

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). Manual for the Sexual Violence Risk – 20: Professional guidelines for assessing risk of sexual violence. Burnaby, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University. Bogaerts, S., Spreen, M., Horváth, Z., Polak, M. S., & Cima, M. *Career criminality, childhood experiences and adult personality among forensic patients in the Netherlands.* In J.B., Helfgott (Ed.). (Forthcoming). Criminal Psychology. 2012; Volumes 1-3. Westport, CT: Praeger Publishers; 2012.

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and treatment* (User Report No. 2007-06). Ottawa, Ontario: Public Safety Canada. Retrieved from: https://cpoc.memberclicks.net/assets/Realignment/risk_need_2007-06_e.pdf

Carlier, I. V. E., Meuldijk, D.,Van Vliet, I. M., Van Fenema, E. M., Van der Wee, N. J. A., & Zitman, F. G. (2012). Empirische evidentie voor de effectiviteit van routine outcome monitoring; een literatuur onderzoek. [Empirical evidence for the effectiveness of Routine Outcome Monitoring. A study of the literature]. *Tijdschrift voor Psychiatrie, 2,* 121-128. Retrieved from: http://www.tijdschriftvoorpsychiatrie. nl/assets/articles/ TvP12-02_p121-128.pdf

De Baaij, E. J., Hoekman, J., Volman, M. J. M., & Zaad, C. (2006). *Kwaliteit van bestaan bij mensen met een complexe meervoudige beperking*. *NTZ, 32.*

De Beurs, E., & Zitman, F. G. (2007). Routine outcome monitoring. Het meten van therapie- effect in de klinische praktijk met webbased software. *MGV, 62,* 13-28.

De Ruiter, C., & Hildebrand, M. (2007). Risk assessment in Dutch forensic psychiatry. *Netherlands Journal of Psychology, 63,* 152-160. doi: 10.1007/BF03061078

De Ruiter, C., & Trestman, R. L. (2007). Prevalence and treatment of personality disorders in dutch forensic mental health services. *The Journal of the American Academy of Psychiatry and Law, 35,* 92-97. Retrieved from: http://www.jaapl.org/content/35/ 1/92.full.pdf+html

De Vries, K., & Spreen, M. (2012). De HKT-30 als instrument voor beslismomenten binnen een tbs behandeling. [Can the clinical items of the risk assessment instrument HKT-30 predict to what extent forensic patients will break agreements during the resocialisation phase of treatment?]. *Tijdschrift voor Psychiatrie, 5,* 429-438. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/TvP12-05- p429-438.pdf

Dickens, G., Sugarman, P., & Walker, L. (2007). HoNOS-secure: A reliable outcome measure for users of secure and forensic mental health services. *The Journal of Forensic Psychiatry and Psychology,* 4, 507-514. doi: 10.1080/147899

Douma, J. C. H., Kersten, M. C. O., Koopman, H. M., Schuurman, M. I. M., & Hoekman, J. (2001) Het 'meten' van kwaliteit van bestaan van mensen met een verstandelijke handicap; een overzicht van de Nederlandse instrumenten. *Nederlands Tijdschrift voor de Zorg aan Verstandelijk Gehandicapten*, 1, 17-36. Retrieved from: http://www.kennispleingehandicaptensector.nl/docs/producten/klapper_kwaliteit_Meten.pdf

Drieschner, K. (2010). De Dynamic Risc Outcome Scales (DROS). Trajectum.

Drieschner, K., Hesper, B., Marrozos, I., & Hout, E. (2011). *Duurzaamheid meerwaarde versus inspanningen*. [Durability, value versus effort]. ROM in de GGZ van meten naar weten. Boom, Amsterdam.

Drieschner, K., & Hesper, B. DROS codeerschema. Congres ROM in de GGZ; 2011 nov 10. Driescher, K. Scoring en interpretatie DROS. (Scoring and interpretation DROS). Trajectum. November 2010 (Unpublished).

Drieschner, K. H., Hesper, B., & Marrozos, I. (2010). Effect onderzoek in de tbs-sector: Theoretische overwegingen en praktische uitvoering bij Trajectum Hoeve Boschoord [Effect research in TBS sector: Theoretical considerations and practical implementation at Trajectum Hoeve Boschoord]. *De Psycholoog, 45,* 38-48.

Ellwood, P. M. (1988). Outcomes management. A technology of patient experience. *New England Journal of Medicine, 318*, 1549-56. doi:10.1056/NEJM198806093182329

Expertgroep forensische psychiatrie (2011). *ROM forensische psychiatrie: Adviesdocument. [ROM forensic psychiatry: consultation document].* GGZ Nederland.

Fitzpatrick, R., Chambers, J., Burns, T., Doll, H., Fazel, S., Jenkinson, C., . . . Yiend, J. (2010). A systematic review of outcome measures used in forensic mental health

research with consensus panel opinion. *Health Technology assessment*, 18, 1 – 91. doi. 10.3310/hta14180

Goethals, K. R., & Van Marle, H. J. C. (2012). Routine outcome monitoring in de forensische psychiatrie: een lang verhaal in het kort. [Routine outcome monitoring in forensic psychiatry: a long story cut short]. *Tijdschrift voor Psychiatrie, 2,* 179-183. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/TvP12-02_p179-183.pdf

Gravesteijn, M. (2008). Ondersteuningsbehoefte op werkgebied bij mensen met een lichte verstandelijke beperking: welke factoren hangen hiermee samen? [Support need for people with mild intellectual disabilities in the work place: association with different factors]. Masterthesis: Universiteit Utrecht.

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. [The value of structured risk assessment and diagnosis of psychopathy. A study into the reliability and predictive validity].* Expertise Centrum Forensische Psychiatrie, Utrecht.

Hoekman, J., Douma, J. C. H., Kersten, M. C. O., Schuurman, M. I. M., & Koopman, H. M. (2001). IDQOL – Intellectual Disability Quality of Life. Gemiva-SVG Groep, Gouda; BBI, Utrecht; LUMC afd. Kindergeneeskunde, Leiden.

Hout E. (2011). Pilotstudie specificatie en evaluatie van behandeldoelen met de Dynamic Risk Outcome Scales. [Pilot study, specification and evaluation of treatment goals with the Dynamic Risk Outcome Scales; internal memo]. 2011. Boschoord: Trajectum, intern rapport.

Kraijer, D., & Kema, G. (1994). Sociale redzaamheidsschaal voor zwakzinnigen van hoger niveau, handleiding. Lisse: Swets & Zeitlinger

Lammers, S. M. M. (2007). Blijven stoute jongens stout? Taxatie van het recidiverisico in de tbs. [Remain bad boys bad? Estimation of the risk of recidivism in TBS]. *De Psycholoog , 4*, 194-201.

Mulder, C. L., Van der Gaag, M., Bruggeman, R., Cahn, W., Delespaul, P. A. E., Dries, P., ... Van Os, J. (2010). Routine outcome monitoring voor patiënten met ernstige psychiatrische aandoeningen: een consensusdocument. [Routine outcome monitoring for patients with severe mental illness: A consensus document.] *Tijdschrift voor Psychiatrie, 3,* 169-179. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/ articles/ articles_2927pdf.pdf

Mulder, C. L., Staring, A. B. P., Loos, J., Buwalda, V. J. A., Kuijpers, D., Sytema, S., & Wierdsma, A. I. (2004). De Health of the Nations Outcome Scales (HoNOS) als instrument voor 'routine outcome assessment'. [The Health of the Nations Outcome

Scales (HoNOS) in Dutch translation as an instrument for Routine Outcome Assessment]. *Tijdschrift voor de psychiatrie,* 5, 273-284. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/articles_1238pdf.pdf

Mulder, N., Van Os, J., Kahn, R., Linszen, D., Van der Gaag, M., Wiersma, D., ... Kempen Nationale remissie werkgroep [National remission working group]. (cited 2007 sep). Available from: http://www.inremissie.nl/remissieproject_patienten.html

Nagtegaal, M. H. (2010). Risicotaxatie- en risicomanagementmethoden. Een inventarisatie in de forenisch psychiatrische centra in Nederland. WODC.

Nederlands Kenniscentrum Jeugd. Sociale Redzaamheid voor Zwakzinnigen. [Social reliance for mentally diabled]. Retrieved from: www.nji.nl

Nijman, H. L. I. (2005). *Kunnen we delinquenten niet beter behandelen?* Inaugurele rede. Nijmegen: Radbout Universiteit Nijmegen.

Nugter, M. A., & Buwalda, V. J. A. (2012). Achtergronden en gebruiksmogelijkheden van ROM in de GGZ. *Tijdschrift voor Psychiatrie, 54*, 111-120. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/TvP12-02_p111-120.pdf

Oliver, J. P. J., Huxley, P. J., Bridges, K., & Mohamad, H. (1996) Quality of life and mental health services. Routledge, London

Priebe, S., Huxley, P., Knight, S., & Evans, S. (1999). Application and results of the Manchester Short Assessment of quality of life (MANSA). *International Journal of Social Psychiatry,* 1, 7-12. doi: 10.1177/002076409904500102

Robertson, P., Barnao, M., & Ward, T. (2011). Rehabilitation frameworks in forensic mental health. *Agression and Violent Behavior, 16,* 472-484. doi: 10.1016/j.avb.2011.03.003

ROM Expertgroep Volwassenen EPA: *ROM ernstig psychiatrische aandoeningen, Adviesdocument [ROM severe psychiatric disorders, consultation document].* GGZ Nederland; 2011.

Schuringa, E. (2010). Routine Outcome Monitoring in het FPC dr. S. van Mesdag. [Routine Outcome Monitoring in the FPC Dr.S. van Mesdag]. *GGZ Wetenschappelijk, 1,* 27- 35. Retrieved from" http://static.mesdag.wdmc.eu/niAe9OzHKU- x2JO-03A57HQ/TW4xvBca2ENSZYBwBXk9Pg/1iYa13a3rUkM2JxvSfP4zg/Routine%20Outcome%20Monitoring%20in%20FPC%20Dr.%20S.%20van%20Mesdag%202010%20-%20E.%20Schuringa.pdf

Schuringa, E. (2011). Een voorbeeld uit het patiënt volg systeem van het FPC Dr. S. van Mesdag [An example from the "patient follow system" in the FPC Dr. S. van Mesdag]. *GGzet Wetenschappelijk, 15,* 55-69.

Schurina, E., Bogaerts, S., & Spreen, M. Instrument voor de Forensische Behandelevaluatie self report [Instrument for Forensic Treatment Evaluation, self-report, internal memo]. 2011.

Schuringa, E., Heininga, V., & Spreen. M. (2011). De N=1 statistiek achter het patiënt volg- systeem in het FPC Dr. S. van Mesdag [The N=1 statistics behind the "patient follow system" in the FPC Dr. S. van Mesdag]. *GGzet Wetenschappelijk, 15,* 70-77. Retrieved from: http://static.mesdag.wdmc.eu/niAe9OzHKU- x2JO03A57HQ/TW4x-vBca2ENSZYBwBXk9Pg/m6iZlbFzTUjpJrd8IZV4KQ/De%20N%201%20statistiek%20achter%20het%20patient%20volg%20systeem%20in% 20FPC%20Dr.%20S.%20van%20Mesdag%20-%20Schuringa%20% 20Heininga%20&%20Spreen%20-%202011.pdf

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, 2,* 127-144. doi. 10.1080/15228932.2014.897536

Spreen, M., Ter Horst, P. R. M., Bogaerts, S., Lammers, S. M. M., Hochstenbach, J., Soe-Agnie, S. E., & Reubsaet, R. J. (2009). Risicotaxatie aan de hand van de HCR-20 en de HKT- 30: Een vergelijking tussen beide instrumenten. [Risk assessment by the HCR-20 and HKT-30: A comparison between both instruments]. In T. I. Oei & M. S. Groenhuijsen (Eds.), *Forensische psychiatrie en haar grensgebieden: Actualiteit, geschiedenis en toekomst* (pp. 459-476). Alphen aan den Rijn: Kluwer.

Stinckens, N., Smits, D., Claes, L., & Soenen, S. (2012). Zinvol 'rommen': zoeken naar een evenwicht tussen gebruiksvriendelijkheid en klinische relevantie. [Meaningful 'Romming' (Routine outcome monitoring): Seeking a balance between user- friendliness and clinical relevance]. *Tijdschrift voor Psychiatrie, 54*, 161-165. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/TvP12- 02_p161-165.pdf

Ter Horst, P., & Zuidam, Y. van. (2011). *Begrippenkader eenduidig referentiekader en consensusscore. [Conceptual framework, unambiguous reference and consensusscore].* ROM in de GGZ van meten naar weten. Boom, Amsterdam.

The AAID Ad Hoc Committee on Terminology and Classification (2010). Intellectual Disability, definition, classification and systems of support, 11th edition. AAID, Washington DC.

Van Hees, S., Van der Vlist, P., & Mulder, N. (2011) *ROM in de GGZ van weten naar meten*. p. 197-205. Boom. Amsterdam.

Van Os, J., & Kahn, R. (2007). Remissiecriteria voor de diagnose schizofrenie. [Remission criteria in schizophrenia]. *Tijdschrift voor Psychiatrie, 49,* 21-26. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/articles_1583pdf.pdf

Van Nieuwenhuijzen, C., Bogaerts, S., De Ruiter, E. A. W., Bongers, I. L., Coppens, M., & Meijers, R. A. A. C. (2011). tbs-behandeling geprofileerd; een gestructureerde casus-

senanalyse. [Profiling TBS-treatment: A structured cases analysis]. Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) Ministerie van Veiligheid en Justitie. Retrieved from: http://www.wodc.nl/onderzoeksdatabase/ voorstudie-onderzoek-tbs.aspx?cp=44&cs=6779

Van Nieuwenhuijzen, C., Schene, A. H., & Koeter, M. W. J. Manchester verkorte kwaliteit van leven meting. [Manchester short assessment of quality of life; Dutch translation]. 2000

Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Task Force in Forensic Psychiatry]. (2002). *Handleiding HKT-30 versie 2. Risicotaxatie in de forensische psychiatrie* [Guide HKT-30 version 2. Risk Assessment in Forensic Psychiatry]. Den Haag, The Netherlands: Ministerie van Justitie, Dienst Justitiële Inrichtingen. Ward, T., Collie, R. M., & Bourke, P. (2009). *Models of offender rehabilitation: The good Lives model and the Risk-Need-Responsivity model*. In Assessment and treatment of sex offender a handbook. Edited by: Beech, R. A., Craig, L. A. & Browne, K. D. (2009). John Wiley & Sons, West Sussex.

Ward, T., & Stewart, C. A. (2003). The treatrment of sex offenders: risk management and good lives. *Professional psychology: Research and Practice, 4,* 353-360. doi: 10.1037/0735-7028.34.4.353

Ward, T., & Gannon, T. A. (2006). Rehabilitation, etiology, and self-regulation: the comprehensive good lives model of treatment for sexual offenders. *Agression and Violent Behavior, 11,* 77-94. doi: 10.1016/j.avb.2005.06.001

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

Wing, J., Beevor, A., Curtis, R., Park, S., Hadden, S., & Burns, A. (1998) Health of the Nation Outcome Scales (HoNOS). *Britisch Journal Psychiatry, 172*: 11-18.

CHAPTER THREE

Patient Profiles in Dutch Forensic Psychiatry Based on Risk Factors, Pathology, and Offense

Frida C. A. van der Veeken
S. Bogaerts
J. Lucieer

This is an adapted version of the article published as: Van der Veeken, F. C. A., Bogaerts, S., & Lucieer, J. (2017). Patient profiles in Dutch forensic psychiatry based on risk factors pathology and offense. *International Journal of Offender Therapy and Comparative Criminology, 61*, 1369-1391. doi: 10.1177/0306624X15619636

Abstract

Forensic psychiatry embodies a highly heterogeneous population differing widely in terms of diagnoses, crimes committed, and risk factors. All of these are vitally important for treatment indications and should be accounted for in research. However, there is limited empirical knowledge of patient profiles. This study constructed *patient profiles* on the basis of the three domains mentioned above. Participants were found guilty of having committed crimes due to psychiatric disorders and were admitted to Forensic Psychiatric Center (FPC) 2landen or FPC De Kijvelanden in the Netherlands. Retrospective data were retrieved from patient files. Diagnoses were assessed according to the Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.; *DSM-IV-TR*) criteria and risk factors according to the Historical Clinical Future–30 (HKT-30) instrument. Latent class analysis was conducted to define typologies; external variables were included for validation. Four different classes or "patient risk profiles" with varying psychopathologies, risk factors, and crimes, were identified. Results were consistent with previous studies, and external validation with the Psychopathy Checklist–Revised (PCL-R) two-factor model and the four facets of the PCL-R agreed with results found. Results display specific risk factors for specific psychopathology/offense combinations.

Introduction

The patient population in forensic psychiatry is very heterogeneous and differs in terms of psychopathology, types of offense, and risk factors (Van Nieuwenhuizen et al., 2011). The Risk-Need-Responsivity (RNR) model of offender rehabilitation is a multidimensional and dynamic theoretical approach to forensic treatment objectives and risk of recidivism. The RNR model comprises three principles. The Risk principle implies that treatment intensity should match the risk level of recidivism: A high-risk offender requires a securer setting and a more intense and longer lasting treatment. The Need principle implies that the treatment approach should focus on a patient's specific dynamic reversible risk factors or criminogenic needs related to the risk of recidivism. The Responsivity principle implies that the intervention should be adapted to the offender's learning style and abilities (Bonta & Andrews, 2007).

To translate scientific research and group therapy modules into the principles of the RNR model, it is helpful to categorize the heterogeneous group of forensic patients into specific clinical patient profiles for the purpose of indicating appropriate treatment (Yiend, Freestone, Vazques-Montes, Holland, & Burns, 2013) and estimating risk of future recidivism.

Several studies have focused on patient profiles. Nanayakkara, O'Driscoll, and Allnutt (2012) studied risk levels of civil and forensic patients in a forensic institution and in the community, using the Historical Clinical Risk–20 items (HCR-20; Webster, Douglas, Eaves, & Hart, 1997). Patients in both groups had been diagnosed with schizophrenia, substance abuse, and other disorders such as bipolar disorder or depressive disorder. All patients had been referred to forensic mental health services. Criminogenic needs were higher for the civil patient group than for the forensic group.

Yiend et al. (2013) developed patient profiles using histrionic and narcissistic personality disorders and the Psychopathy Checklist–Revised (PCL-R; Hare, 1991) Factor one and Factor two scores of patients in high-security prisons or high-security psychiatric hospitals. They composed three profiles: a delinquent profile with high PCL-R factor two scores but few personality disorders, a primary psychopathy profile with high PCL-R factor one scores and a narcissistic personality disorder, and an expressive psychopathic profile with high PCL-R factor one scores and a histrionic personality disorder. According to the HCR-20 and the Violence Risk Scale (VRS; Wong & Gordon, 2006), individuals in the delinquent profile were at the highest risk of recidivism. These results are consistent with the study by Bogaerts, Polak, Spreen, and Zwets (2012), who concluded that secondary psychopaths show more problematic behavior and reactive aggression than primary psychopaths, who show fewer risk factors but committed homicide more often.

Bogaerts and Spreen (2011) developed patient profiles by examining 234 inpatient forensic offenders with a primary psychotic disorder and 348 inpatient forensic offenders

with a primary personality disorder. Patient profiles were based on risk and protective factors extracted from the Dutch risk assessment tool, the Historical Clinical Future–30 items (HKT-30; Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Task Force in Forensic Psychiatry], 2002). Hierarchical cluster analysis and Latent Class Analysis (LCA) were used to examine the profiles.

Within the group of psychotic offenders, Bogaerts and Spreen (2011) identified three profiles. Two of these were characterized by the presence of schizophrenia, the presence of serious problems in the social and interpersonal domain, and the absence of personality disorders, impulsivity, and hostility. An important difference between these two profiles was related to the presence of historical determinants such as past criminal offenses, help-seeking behavior, conduct disorders, and school problems. The third profile was characterized by schizophrenia in combination with persistent dysfunctional problems and antisocial personality factors. Within the personality-disordered group, Bogaerts and Spreen again identified three profiles. The first profile was characterized by high-risk factors such as past treatment, past substance use, psychotic symptoms, empathic skills, social skills, and daily activities. The second profile was similar except that the patients were neither diagnosed with psychotic symptoms nor displayed many psychotic symptoms compared with the first profile. The third profile was based on lower overall risk factors and the absence of a psychotic disorder diagnosis.

Van Nieuwenhuizen et al. (2011) developed forensic patient profiles with LCA based on diagnoses derived from the Diagnostic and Statistical Manual of Mental Disorders (*4th ed., text rev.*; *DSM-IV-TR*; American Psychiatric Association [APA], 2000) Axis I and Axis II diagnoses and type of offense among a representative group of 180 forensic patients from 13 forensic psychiatric centers in the Netherlands. Three profiles showed strong similarities with the profiles found by Bogaerts and Spreen (2011): psychotic patients with multiple problems, typical psychotic patients, and antisocial patients. In addition, they identified a fourth and fifth profile: patients with addiction issues and patients with sexual problems and delinquent behavior.

The described patient profiles were developed on a limited set of factors relating to the domains of psychopathology, and/or type of offense, and/or risk and protective factors. In none of these studies were these three domains combined to develop integrative profiles, which would have a significant value for the diagnostic process, treatment indications, and recidivism risk assessment (Bonta & Andrews, 2007). As the interdependence of psychopathology, type of offense, and risk and protective factors is so common in forensic psychiatry, I will explain the importance of such an integrative model in the following paragraphs.

It has often been shown that specific psychopathology may reinforce specific criminal behavior. A paraphilic disorder such as pedophilia is often related to child sexual abuse (Bo-

gaerts, Vanheule, & DeClercq, 2005; Bogaerts, Vervaeke, & Goethals, 2004; Buschman et al., 2010). Rapists who are more violent in nature tend to have more psychopathic and antisocial features (Firestone, Bradford, Greenber, & Serran, 2000). Violent behavior appears to be a small but significant risk in psychotic disorders such as schizophrenia (Peterson, Skeem, Hart, Vidal, & Keith, 2010). Psychopathy (Hildebrand, Hesper, Spreen, & Nijman, 2005; Leistico, Salekin, DeCoster, & Rogers, 2008) and personality disorders, especially antisocial (Bonta, Law, & Hanson, 1998) and narcissistic disorders are predictors of both general and violent offending behavior (O'Driscoll, Larney, Indig, & Basson, 2012).

Type of offense, second, is also related to future recidivism (Coid, Hickey, Kahtan, Zhang, & Yang, 2007). Various studies have related types of offenses to specific risk factors. Sexual offenders, for example, show a lack of empathic understanding (Hall & Hall, 2007), deviant thoughts, poor affect regulation (Scoones, Willis, & Grace, 2012), and problem-solving deficits (Lockmuller, Beech, & Fischer, 2008). Violent offenders show a history of substance use, hostility, and impulsivity (Craig, Browne, Beech, & Stringer, 2006).

Risk and protective factors, finally, are strongly related to recidivism (Andrews, & Bonta, 2010; De Vogel, De Ruiter, & Bouman, 2007; Palermo, 2009; Rennie & Dolan, 2010; Ward & Gannon, 2006). A widely studied set of eight risk factors lists strong predictors of recidivism (Andrews & Bonta, 2010; Andrews et al., 2012). These include a history of antisocial behavior, an antisocial personality pattern (impulsivity, hostility, and lack of empathy), antisocial cognitions, antisocial associations, a dysfunctional family/marital situation, work and school problems, a lack of daily activities, and substance abuse. Historical risk factors include previous lifestyles or transitions leading to changes in a person's life course (e.g., job loss; Laub, Sampson, & Sweeten, 2006), which have been shown to be predictive of future violence and help to understand problematic behavior (Cernkovich & Giordano, 2001). Dynamic risk factors or criminogenic needs are often the subject of treatment (Bonta & Andrews, 2007) with a view to reducing the risk of recidivism.

The aim of this study was to develop patient profiles based on psychopathology, type of offense, and risk and protective factors (historical and dynamic risk factors), derived from patient files. As comorbidity is high in forensic populations, both Axis I and Axis II disorders of the *DSM-IV-TR* were included (APA, 2000). Differentiating between clinical patient characteristics, we hypothesized we would find patient profiles, possibly similar to the profiles found by Van Nieuwenhuizen et al. (2011) and Bogaerts and Spreen (2011). Because studies have shown that different risk factor levels are related to the PCL-R factor scores (Bogaerts et al., 2012; Yiend et al., 2013), we compared the profiles that emerged on both Factors 1 and 2 of the PCL-R (Hare, 1991) and the four-facet model of the PCL-R (Hare & Neumann, 2005) to examine whether our results showed similarities with previous studies.

Method

Procedure

This study was approved by the scientific research committee of the Forensic Psychiatric Center (FPC) where this study was conducted. Retrospective data were obtained from Electronic Patient Files (EPF) of 328 male patients residing within FPC 2landen or FPC De Kijvelanden in the Netherlands between 2003 and 2011. An annual risk assessment is conducted for all patients, and we selected risk assessments from approximately the first 2 years of treatment. First or second risk assessments were selected on the basis of their time after admission (between 6 and 23 months), and we assessed behavior observed in the preceding year. All data involved primary treatment information and were anonymized for this study.

Participants

All patients were admitted to a forensic psychiatric center by court order. They had all committed an offense with a minimum penalty of 4 years and an Axis I or Axis II *DSM-IV-TR* diagnosis (APA, 2000). They faced detention under a hospital order (TBS), meaning a court-imposed treatment measure. After exclusion of incomplete EPFs or untimely risk assessments, 244 patients were included in this study. Mean age was 38.23 (*SD* = 10.57, range = 20-69) at the moment of risk assessment, conducted approximately 14.70 (*SD* = 3.64, range = 6-23) months after admission. Table 2 displays intelligence quotients, index offenses, and Axis I and Axis II *DSM-IV-TR* diagnoses (APA, 2000).

Table 1

Historical Risk Future–30 items (HKT-30)

Historical	Clinical	Future
H01 Legal history	K01 Problem awareness	F01 agreement on future conditions
H02 Violation of term	K02 Psychotic symptoms	F02 Material indicators
H03 Conduct problems before the age of 12	K03 Current substance use	F03 Daily activities
H04 Victim of violence in youth	K04 Impulsivity	F04 Skills
H05 Past care	K05 Empathy	F05 Social network
H06 History of school and work	K06 Hostility	F06 Stressing circumstances
H07 Past substance use	K07 Social and relational skills	
H08 Psychotic disorders	K08 Self-reliance	
H09 Personality disorders	K09 Acculturation issues	
H10 Psychopathy	K10 Treatment attitude	
H11 Sexual deviance	K11 Crime responsibility	
	K12 Sexual preoccupation	
	K13 Coping skills	

Measurements

Risk factors. The HKT-30 is a risk assessment tool derived from the HCR-20 (Webster et al., 1997) and developed for the Dutch forensic situation (Risk Assessment Task Force in Forensic Psychiatry, 2002; Harte & Breukink, 2010). The HKT-30 assesses both static and dynamic risk factors. For all patients, the HKT-30 tool is completed annually by trained raters with observations, file reports, and/or staff interviews. The HKT-30 total score has a moderate to good predictive value for violent recidivism (De Vries & Spreen, 2012; Hildebrand et al., 2005; Spreen et al., 2009). The HKT-30 instrument consists of 11 static historical items, 13 dynamic clinical items, and six dynamic future items, considering situations after clinical admission all displayed in Table 1. All items were scored on a five-point scale indicating the degree of the item's severity, a score of four indicating severe problematic behavior (e.g., physical aggression or severe loss of impulse control) and a score of zero representing no problematic behavior at all or even well-adjusted behavior (e.g., good relational and social skills and good patient-network support). Internal consistency, assessed with Cronbach's alpha, was rather good for the total scale, α = .79, the clinical scale, α = .76, and the future scale, α = .76, and sufficient for the historical items; α = .68. Interrater reliability was calculated with the Pearson correlation coefficient for a subsample (*N* = 22-36). Overall scores for the clinical and future scales were good (*r* = .66-.87), but the interrater reliability of the items *hostility, treatment attitude,* and *coping skills* was moderate (*r* = .51-.58) and that of the future item *skills* was too low (*r* = .29).

Psychopathology. DSM-IV-TR (American Psychiatric Association (APA), 2000) Axis I and Axis II diagnoses were assessed by trained psychiatrists and/or clinical psychologists. Primary Axis I diagnoses were divided into six groups: psychotic disorders, substance use disorders (SUDs), pedophilia, pervasive developmental disorders, a residual category including less common disorders (such as dysthymic disorder, paraphilia other than pedophilia, or bipolar disorder), and "no diagnosis on Axis I." Personality disorders were classified by the *personality disorder* item of the HKT-30 items. The personality disorder item is comparable with Tyrer and Johnson's (1996) classification system of the severity of personality disorders (Bogaerts, Spreen, Horváth, Polak, & Cima-Knijff., 2013). A score of zero on this item indicates "no signs of a personality disorder"; score of one: "pathological features"; score of two: "one or more personality disorders but no cluster B"; score of three: "one or more personality disorders, with one cluster B." A score of four, finally, indicates "two or more personality disorders with two cluster B or one cluster A and B personality disorder" (Risk Assessment Task Force in Forensic Psychiatry, 2002).

Offenses. Index offenses were gathered from the EPFs and categorized into seven offense types: property offenses, violent property offenses, maltreatment, homicide, arson, sexual

offenses other than child sexual abuse, child sexual abuse, and other. The homicide category includes manslaughter, murder, and attempted murder.

Psychopathy. The PCL-R (Hare, 1991) is a 20-item clinical rating scale for assessing psychopathy. For all patients, an assessment based on historical information was conducted by trained clinicians, and a semi-structured clinical interview was held if possible. Items were scored on a three-point scale: zero meaning that the item did not apply, one that the item applied partly, and two that the item applied fully.

Table 2

Patient Characteristics Based on Primary Axis I and Axis II Diagnoses, Index Crime, and IQ

	%
Primary Axis I diagnosis	
Psychotic disorders	31.2
Substance use disorder	34
Pedophilia	7.8
Pervasive disorder	6.1
Other	15.2
Attention disorder	1.6
Mood disorder	2.8
Sexual abuse of a child	1.8
Physical or sexual abuse of an adult	2
Parafilia (other than pedophilia)	1.6
Other	5.4
None	5.7
Primary Axis II diagnosis	
Cluster A	1.6
Cluster B	44.3
Cluster C	2.5
Not otherwise specified (NOS)	35.2
Other	3.3
None	11.9
Index crime	
Property offenses	1.6
Property offenses with violence	8.6
Maltreatment	19.3
Homicide	32.4
Arson	7.8
Sexual offenses (other than child sexual abuse)	13.9
Child sexual abuse	14.8
Others	1.6
IQ[a]	
Mentally retarded	2.5
Borderline retarded	9.4
Low average	24.2
Average	39.8
High average	8.6
Superior	1.6
Very superior	.4
Unknown	12.6

[a]Assessed with the Wechsler Adult Intelligence Scale, RAVEN progressive matrices, Groninger Intelligence Test, or file reports.

PCL-R interrater reliability was good for the total score (intraclass correlation [ICC] = .85), the factor one score (ICC = .80), the factor two score (ICC = .83), and also for the four-facet scores (ICC = .74-.79; Hildebrand et al., 2005). Classes were compared on the PCL-R total score, both the PCL-R two-factor (Hare et al., 1990) and four-facet scores (Hare & Neumann, 2005). The PCL-R two-factor model distinguishes an "affective and interpersonal factor" (Factor 1) and a behavioral or "socially deviant" factor (Factor 2; Hare & Neumann, 2005; Yiend et al., 2013). The four facets of the two-factor model comprise a more concrete representation of the psychopathy construct with good fit indices (Tucker–Lewis index [TLI] = 94, root mean square error of approximation [RMSEA] = .07, standardized root mean square residual [SRMR] = .05; Hare & Neumann, 2005). The four-facet model describes an antisocial facet, an interpersonal facet, an affective facet, and a lifestyle facet. Internal consistency, assessed with Cronbach's alpha, was good for the PCL-R total scale and two-factor solution (α = .85), Factors one (α = .79) and two (α = .83); it was acceptable for the four-facet solution Factors one (α = .73), two (α = .72), 3 (α = .79), and four (α = .76). Interrater reliability was assessed with the Pearson correlation coefficient on a subsample of the PCL-R. Interrater reliability was good for Factor one; N = 50, r = .90, Factor 2, N = 39, r = .95; the interpersonal facet, N = 52, r = .84; the affective facet, N = 50, r = .90; the lifestyle facet, N = 46, r = .90; and the antisocial facet, N = 46, r = .94. Mean score on the PCL-R was M = 20.59 (SD = 7.89, range = 3-38), and the cutoff value for clinical use was 30.

Statistical Analysis

Data were analyzed using the Statistical Package for the Social Sciences (SPSS) 19 and transferred into Latent Gold 4.5, to conduct latent class analysis (LCA). LCA is a statistical method for defining typologies based on selected features (Vermunt, 2004). LCA produces a probabilistic classification of all cases to identify subgroups and will assign cases to the most likely class with a deterministic classification (Mulder, Vermunt, Brand, Bullens, & Van Marle, 2012). Classes are predicted by indicators and active covariates, and latent gold can estimate several models at the same time, to determine the goodness of fit and select the best fitting model (Vermunt & Magidson, 2005). An advantage of LCA is that it can be used to cluster cases into homogeneous groups (Mulder et al., 2012). Variables that are not used to determine class membership but are used in describing the classes or for making comparisons between classes can be included as inactive covariates. The Bayesian information criterion (BIC) and the Akaike information criterion (AIC) indicate the model fit, with a lower BIC and AIC value indicating a better model fit. The BIC was selected to assess model fit within this study. When the best fit has been selected, the Bootstrap p value

is estimated to provide a more precise estimation and improved power: $p > .05$ indicates a good fit. The Entropy R^2 and Reduction of errors show how well the model predicts class memberships based on observed variables; the closer to a value of one, the better the model predicts class membership (Vermunt & Magidson, 2005[1]). Explorative LCA analyses were conducted to construct the best fitting model with predictive items for the model. The −2 log likelihood (−2LL) tests whether the chosen model provides a significant improvement compared with a model with fewer classes (Vermunt & Magidson, 2005[1]). Latent gold 4.5 data were loaded back into SPSS 19 after analyses to compare the classes we found. An ANOVA was conducted to compare the classes on excluded HKT-30 items, the PCL-R total score, two-factor scores, and four-facet scores, which were included as inactive covariates. To assess whether differences in dynamic risk factor scores were due to the period of assessment after admission, an ANOVA was conducted for the period between admission and date of assessment. Scheffe and Tukey post hoc tests were used to assess which classes differed for all ANOVAs.

Results

Model Fitting

Explorative LCA indicated a final model with the HKT-30 items displayed in Table 4 as active indicators, primary Axis I diagnosis and severity of personality disorder as active covariates, and index offense as inactive covariate. The historical and clinical items with low explained variances (R^2 = .01-.17) were excluded from the final model; these were *substance use in the past year*, *victim of violence before the age of 12*, *conduct problems before the age of 12*, *self-care skills*, *psychotic symptoms in the past year*, *past psychotic symptoms*, *sexual deviation*, and *culture-related issues*. These last three items were also excluded from the HKT–Revised, which was validated during this study (Spreen, Brand, Ter Horst, & Bogaerts, 2014). The future item *agreement on conditions* was also excluded; this concerns conditions after treatment and also showed a lower explained variance (R^2 = .17); for the current treatment, this was accounted for in the item *treatment attitude*. The future items *daily activities* and *material indicators* were also excluded as these concern situations outside the FPCs, as *material indicators* concern housing after treatment.

The duration of a TBS order in the Netherlands was approximately nine years at the time of this study (Van Nieuwenhuizen et al., 2011). Most patients had a prison sentence before they were admitted to the forensic institution. As the assessments were conducted during the start of treatment, proper housing or daily activities outside the FPC were not yet accounted for and would not, therefore, discriminate between classes.

The estimation of classes from a three- to six-class solution yielded the lowest BIC value for the four-class model (see Table 3). Model fit was good after bootstrapping (p = .14; see Table 3). Bootstrap −2LL implied a better fit of the four-class solution compared with the three-class model (p < .00). The separate indicators contribute significantly to class discrimination (p < .05; Vermunt & Magidson, 2005), and the Entropy R^2 and Reduction of errors (Lambda) implied the model predicts class membership well.

Table 3

Latent class analysis fit statistics

No. of classes	BIC (L^2)	AIC (L^2)	Npar	L^2	*df*	p^a	Class error	Entropy	Reduction of errors (λ)
3	8,282.4	8,716.0	120	8,964.0	124	.18	.05	.87	.91
4	**8,243.5**	**8,582.7**	**147**	**8,776.7**	**97**	**.14**	**.06**	**.88**	**.91**
5	8,283.2	8,528.0	174	8,668.0	70	.10	.06	.90	.92
6	8,357.4	8,507.8	201	8,593.8	43	.04	.06	.90	.91

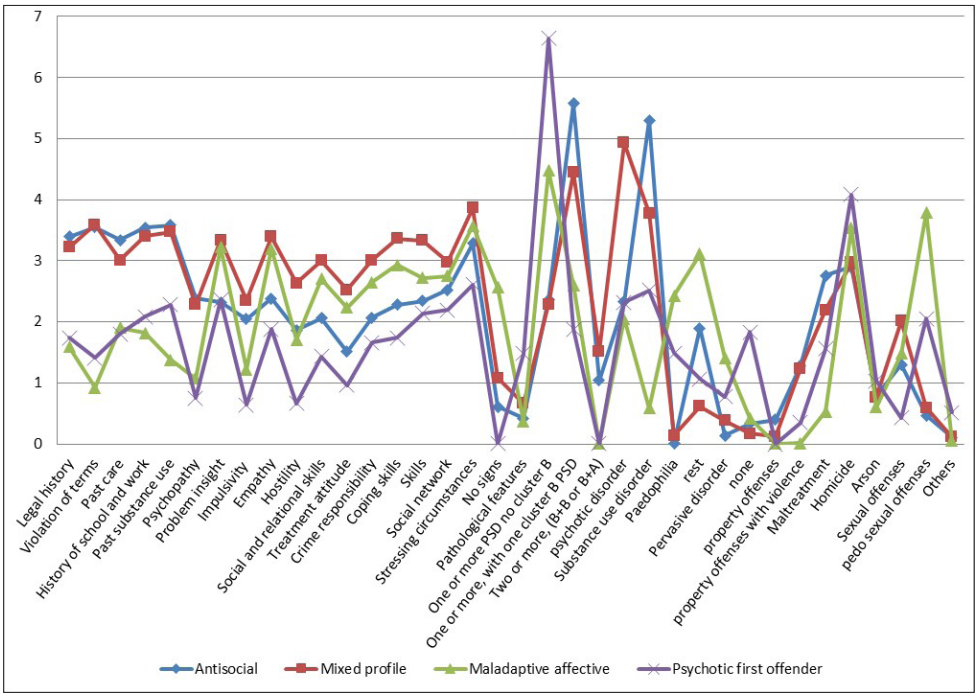
Note. Fit statistics are explained in the "Method" section. BIC = Bayesian information criterion; AIC = Akaike information criterion; Npar = number of parameters. Bold values represent the best fitting model.
[a]*p* after bootstrapping.

The Four-Class Solution to Construct Forensic Patient Profiles

The four-class solution (see Figure 1) displays higher scores on historical items for classes one (antisocial) and two (mixed profile) and lower scores for classes three (maladaptive disordered) and four (psychotic first offender). Class two shows the highest scores on dynamic risk factors, followed by class three. Classes one and four show lower scores on the dynamic risk factors (see Table 4). Patients in class one, or "the antisocial class," were mostly diagnosed with a Cluster B personality disorder (56% one Cluster B, 10% one or more personality disorders Cluster B + B or B + A), 50.6% of whom were diagnosed with an antisocial personality disorder. Fifty-three percent of patients were also diagnosed with an SUD. They had been convicted for the crimes of homicide (29%), maltreatment (28%), sexual offenses (13%), or property offenses with violence (13%). The historical risk factors are especially high (see Table 4), whereas dynamic risk factors are lower (see Table 4). Forty-one percent of the patients showed a secondary Axis I SUD.

Patients in class two, or the "mixed profile with multiple problems," were mostly diagnosed with a Cluster B personality disorder (45% one Cluster B, 15% Cluster B + B or B + A), and a comorbid psychotic disorder (49%) or comorbid SUDs (38% primary, 56% secondary Axis I diagnosis). They had committed several crimes: homicide (30%), maltreatment (22%), sexual offenses (20%), and property offenses with violence (12%). This mixed class shows high historical and dynamic risk factors (see Table 4).

Figure 1
Four-class solution of the latent class analysis



Class three, or "the maladaptive affective disordered class," includes patients with a pervasive disorder (14% primary, 10% secondary diagnosis), a rest disorder (31%; 4% affective disorder, 6% paraphilia [other than pedophilia], 2% dysthymic disorder, 6% sexual abuse of a child or adult, 4% maltreatment of an adult, 2% behavior disorder from childhood, and 6% other), or pedophilia (24% primary, 9% secondary diagnosis), followed by a psychotic disorder (21%). Patients in this class showed the lowest rate of SUDs (6% primary and 25% secondary Axis I diagnosis). They were less often diagnosed with a Cluster B diagnosis, but more often with a Cluster A, C, or a personality disorder not otherwise specified (NOS; 45%). Twenty-six percent did not display any pathological signs of a personality disorder. Most patients had committed child sexual abuse (40%) or homicide (35%). None of the patients with a pervasive developmental disorder was diagnosed with a Cluster B personality disorder. Furthermore, historical risk factors were low, while they displayed high dynamic risk factors; *problem awareness*, *empathy*, *social and relational skills*, *coping skills*, *crime responsibility*, and *stressing circumstances* were especially high (see Table 4).

Class four, or "the psychotic first offender," includes patients with a psychotic disorder (23% primary and 5% secondary Axis I diagnosis), SUD (25% primary and 31% secondary Axis I diagnosis), a rest (11%) diagnosis (3% dysthymic disorder, 3% pathological gambling, 3% maltreatment of an adult, 3% sexual abuse of an adult), and pedophilia (15%). However, there were low rates of Cluster B diagnoses (19% one or more, with one Cluster B). Most patients had a personality disorder NOS or a Cluster A or C diagnosis (66%). Most patients had committed homicide (41%), child sexual abuse (21%), and maltreatment (16%). Historical factors, such as *legal history* and *psychopathy*, were low as were dynamic, clinical, and future risk factors, particularly *hostility* and *impulsivity* (see Table 4).

Table 4
Group HKT-30 means and standard errors for the four profiles

	Class 1 (31%)	Class 2 (33%)	Class 3 (21%)	Class 4 (16%)
Variables	*M* (*SE*)	*M* (*SE*)	*M* (*SE*)	*M* (*SE*)
Legal history	3.39 (.1)	3.22 (.1)	1.59 (.2)	1.74 (.2)
Violation of terms	3.55 (.1)	3.59 (.1)	0.92 (.2)	1.41 (.3)
Past care	3.34 (.1)	3.01 (.1)	1.90 (.2)	1.80 (.3)
School and work history	3.54 (.1)	3.40 (.1)	1.82 (.2)	2.08 (.2)
Past substance use	3.58 (.1)	3.47 (.1)	1.38 (.3)	2.28 (.3)
Psychopathy	2.38 (.2)	2.28 (.2)	1.07 (.2)	0.75 (.2)
Problem awareness	2.32 (.1)	3.34 (.1)	3.22 (.1)	2.36 (.1)
Impulsivity	2.04 (.2)	2.35 (.1)	1.22 (.2)	0.64 (.1)
Empathy	2.38 (.1)	3.39 (.1)	3.20 (.1)	1.88 (.2)
Hostility	1.86 (.1)	2.63 (.1)	1.71 (.2)	0.66 (.1)
Social and relational skills	2.06 (.1)	3.00 (.1)	2.70 (.1)	1.43 (.1)
Treatment attitude	1.51 (.1)	2.52 (.1)	2.23 (.2)	0.96 (.2)
Crime responsibility	2.07 (.1)	3.01 (.1)	2.65 (.2)	1.66 (.2)
Coping skills	2.28 (.1)	3.36 (.1)	2.93 (.1)	1.74 (.2)
Skills	2.34 (.1)	3.33 (.1)	2.72 (.1)	2.14 (.2)
Social network	2.52 (.1)	2.98 (.1)	2.75 (.1)	2.19 (.2)
Stressing circumstances	3.28 (.1)	3.86 (.1)	3.57 (.1)	2.61 (.1)

External Validation

Table five displays ANOVAs with post hoc analysis results. Classes differed significantly on Hare's PCL-R two-factor model; for Factor one, $F(3, 225) = 11.18$, $p = .00$, and for Factor two, $F(3, 189) = 64.89$, $p = .00$. After correction of the uneven item numbers, the Factor two score was significantly higher than the Factor one score for the antisocial profile, $t(66) = -5.05$, $p < .00$. The affective facet was significantly higher for the maladaptive affective disordered profile (maladaptive profile) compared with the other three facets, lifestyle: $t(43) = 11.43$, $p = .00$; antisocial: $t(36) = 9.86$, $p = .00$; interpersonal: $t(45) = 12.55$, $p = .00$.

The variable *months* (moment of risk assessment after admission) did not differ between profiles, $F(3, 240) = 2.15$, $p = .10$. However, *psychotic symptoms in the past year*, $F(3, 239) = 6.97$, $p = .00$, did differ, with the mixed profile scoring significantly higher than the other profiles (Table 5). A higher percentage of patients in the mixed profile showed psychotic symptoms (11% vs. 5% to 8%) or psychosis with a violent and paranoid content (13% vs. 1% to 5%).

Substance use in the past year was higher for the antisocial and mixed profiles, $F(3, 240) = 10.34$, $p = .00$. Twenty-seven percent to 29% of patients in both the mixed and antisocial profiles, respectively, had used a substance in the past year, compared with three percent to 10% in the maladaptive and psychotic first offender profiles. However, about 55% to 56% of patients in the two first profiles had neither used any substances in that period nor were preoccupied with substance use.

Patients in the maladaptive profile scored higher on *sexual preoccupation*, $F(3, 238) = 3.32$, $p = .021$. The mean score implies a suspicion of sexual preoccupation or no sexual preoccupation. However, 13% focused on sexual stimuli, and eight percent of the patients were sexually preoccupied and showed deviant behavior or had violent thoughts.

Table 5

ANOVA External Variables; comparison of risk items and PCL-R two factor and four facet scores

	Class1			Class 2			Class 3			Class4			
Co-variable	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	Tukey test
Sexual deviation	76	0.51	1.10	79	0.65	1.29	49	1.96	1.63	38	0.74	1.37	3 > 1,2,4
Conduct problems < 12	78	1.97	1.35	71	1.73	1.35	44	1.25	1.26	35	0.74	1.09	1 > 3, 4; 2 > 4
Victim of violence	76	2.38	1.52	74	2.01	1.58	48	1.46	1.50	37	1.97	1.61	1 > 3
Psychotic symptoms*	78	0.42	0.96	78	1.13	1.50	49	0.41	0.89	38	0.37	0.97	2 > 1,3,4
Sexual preoccupation*	77	0.27	0.62	79	0.52	0.96	48	0.73	1.11	38	0.32	0.62	3 > 1
Substance use*	78	1.13	1.44	79	1.08	1.39	49	0.08	0.45	38	0.40	0.97	1, 2 > 3, 4
Months	78	14.90	3.99	79	13.92	3.35	49	15.31	3.63	38	15.37	3.44	*Ns*
PCL-R total score*	65	24.19	6.16	50	24.16	7.04	36	15.72	5.91	32	13.19	6.13	1, 2 > 3,4;
Facet 1 (Interpersonal)*	74	3.14	2.20	76	3.38	2.49	46	2.37	2.30	36	1.69	1.90	1, 2 > 4
Facet 2 (Affective)*	74	6.04	1.80	74	6.76	1.35	46	6.33	1.56	34	4.47	1.80	1, 2, 3 > 4
Facet 3 (Lifestyle)*	70	7.00	1.96	70	6.83	2.45	44	2.89	1.99	36	3.53	2.67	1, 2 > 3, 4
Facet 4 (Antisocial)*	71	6.79	2.20	60	6.40	2.31	37	2.11	2.15	34	2.44	1.97	1, 2 > 3, 4
Factor 1 (two-factor model)	74	9.18	3.44	75	10.19	3.38	46	8.70	3.31	34	6.21	3.16	1, 2, 3 > 4
Factor 2 (two-factor model)	67	12.54	2.95	56	11.96	3.82	36	4.86	3.28	34	5.62	3.56	1, 2 > 3, 4

Note. n differs due to missing values on items. *ns* = non-significant; PCL-R = Psychopathy Checklist–Revised.
*$p < .01$, two-tailed

Discussion

The goal of this study was to identify forensic patient classes on relevant characteristics (types of offense, psychopathology, and risk factors) in a highly heterogeneous Dutch forensic population residing in two forensic psychiatric clinics. The different profiles could benefit future studies on the development of more specific group therapies and studies on treatment effectiveness and prognosis. With LCA, we distinguished four classes that differed in risk factors, psychopathologies, and types of offense. The model fit for the four-class solution can be considered good, after bootstrapping. For validation, classes were compared on other clinically relevant factors, such as psychopathy levels and recent substance abuse. The differences found between classes were not affected by the time of assessment.

The first class, "the antisocial class," is characterized by strong personality traits (Cluster B), SUD, different types of offenses, high levels of historical risk factors, and lower levels of dynamic risk factors. The combination of a personality disorder and SUD is related to higher impulsivity, more convictions (Fridell, Hesse, Jaeger, & Kühlhorn, 2008), and more psychiatric symptoms (Trull, Sher, Minks-Brown, Durbin, & Burr, 2000; Zadeh & Damavandi, 2010). Although security measures and regular inspections in forensic institutions limit the use of drugs, the problematic historical behavior could have been worsened by the substance use, explaining the lower dynamic risk factors compared with Clusters two and three. The previous "antisocial lifestyle," including previous offenses, violation of terms, and a problematic school and work history, is in accordance with the higher Factor two PCL-R score, which is related to more reactive aggression (Bogaerts et al., 2012) and the antisocial PCL-R facet score. This also accounts for the diversity in crimes (Davison & Janca, 2012) and Cluster B personality disorder (Hildebrand & De Ruiter, 2012) in relation to the overall PCL-R score.

In sum, this group is characterized by high historical problems, fewer dynamic risk factors, diverse types of offenses, reactive aggression, Cluster B personality disorders, and comorbid SUDs. This could indicate group therapies incorporating comorbidity between Cluster B disorders and SUDs and related risk factors as well as individual schemas relating to reactive aggression.

The second class is referred to as the "mixed profile with multiple problems." Patients are characterized by Cluster B personality disorders, comorbid psychotic disorders, and/or SUD, and display high historical risk factors similar to the antisocial class. In contrast with the antisocial class, however, the dynamic risk factors are also high in this class, which is consistent with the high PCL-R scores on all PCL-R factors. There are similarities with *the mixed cluster* by Bogaerts and Spreen (2011) and *the patient with multiple problems* by

Van Nieuwenhuizen et al. (2011). The co-occurrence of psychotic disorders with Cluster B personality disorders appears to worsen the problematic behavior, in contrast with Class four, which also includes patients with psychotic problems, and Class one, which includes patients with personality disorders.

Results in this class are consistent with previous studies. High levels of substance use (Boutron, Bonnet, & Mak, 1996; Haqqi, 2010), previous hospitalization, and previous convictions are often found to be risk factors in psychotic disorders (Belli & Ural, 2012). The dynamic risk factors of *problem awareness*, *empathy*, *and coping skills* are consistent with problematic behavior found in the comorbid antisocial personality and psychotic disorders (Fullam & Dolan, 2006).

In sum, this group is characterized by high overall risk factors, high comorbidity between Cluster B and psychotic disorders, high PCL-R scores, and a diversity of crimes. Future studies would need to assess the effect of these highly problematic factors on treatment prognosis and effectiveness: Does treatment in this profile take longer than in the other profiles? It is even more important to study treatment responsivity for this group, as studies claim that high psychopathy levels negatively influence treatment responsiveness (Skeem, Poythress, Edens, Lilienfeld, & Cale, 2003).

The third class is the "maladaptive affective disordered class." Patients in this class suffered mostly from pedophilia or pervasive developmental disorders and/or personality disorders NOS. Most offenses concerned homicide or child sexual abuse, and patients displayed low historical risk factors but higher dynamic risk factors: *social skills*, *empathy*, *crime responsibility*, and *problem awareness*. This is consistent with risk factors found in pedophilia (Neutze, Grundmann, Scherner, & Beier, 2012) and pervasive developmental disorders (Murphy, 2010). This class shows similarities with *the patient with sexual problems and sexual crimes* by Van Nieuwenhuizen et al. (2011) and with a cluster found by Woessner (2010), which includes child sexual offenders with paraphilia and fewer personality disorders or SUDs, who show highly adaptive behavior and would require much control and the training of many coping skills. The high affective PCL-R facet score is consistent with the results found and implies poor social and emotional functioning (Allan, Grace, Rutherford, & Hudson, 2007). The same is true for the higher sexual preoccupation found in this profile.

Murphy (2010) describes higher degrees of suppressed anger styles and adaptive behavior in patients with pervasive developmental disorders, which may partly explain the lower historical risk factors. Problem behavior appears to have surfaced after the index crime occurred. The higher dynamic risk factors of *coping skills*, *lack of social support*, and *skills to hold one's own outside the forensic institution* are consistent with the needs addressed in the Circles of Support and Accountability (COSA) project studies, implying that pedophile offend-

ers require a highly supportive network (Hannem, 2013; Höing, Bogaerts, & Vogelvang, 2013). The level of victim violence is remarkably low, although this is an oft-mentioned risk factor in the development of sexual offenders (Lee, Jackson, Pattinson, & Ward, 2002; Mouridsen, 2012).

In sum, this class is characterized by high pedophilia and pervasive developmental disorders, low historical risk factors, higher dynamic risk factors, particularly *coping*, *problem awareness*, *social skills*, and *crime responsibility*, related to the higher affective facet score on the PCL-R. Treatment could focus on the acquisition of social skills and coping skills (Woessner, 2010), and a prognosis could focus on post-treatment external control of the patient group, similar to the COSA project (Hannem, 2013).

The fourth and last class, the "psychotic first offender," is characterized by low overall risk factors compared with the other classes, homicide, and low levels of psychopathy. Patients are more often diagnosed with psychotic disorders, SUDs, and/or a Cluster A, C, or personality disorder NOS. This class shows similarities with the *psychotic cluster* found by Bogaerts and Spreen (2011) and *the typical psychotic patient* found by Van Nieuwenhuizen et al. (2011). Compulsory admissions and risky behavior are more often found in patients suffering from psychotic disorders and Cluster B personality disorders or SUDs than in patients showing no comorbidity (Boutron et al., 1996; Curson, Duke, Harvey, Pantelis, & Barnes, 1999; Fullam & Dolan, 2006; Haqqi, 2010). The prevalence of patients who commit an offense as a direct result of a psychosis is small. Most patients with a mental illness have committed an offense as a result of hostility or emotional reactivity, consistent with risk factors found for patients in the mixed profile, explaining the low number of the *typical psychotic patient* (Peterson et al., 2010). However seldom it occurs, homicide is still considered the most significant complication of a psychosis (Bo, Abu-Akel, Kongerslev, Haahr, & Simonsen, 2011). Untreated psychotic symptoms are one of the most important risk factors for violent behavior in psychotic patients, consistent with the higher homicide rates. Considering the low numbers of historical conduct, this profile appears to include the "first offender." Although lower than the other classes, patients show somewhat limited problem awareness, empathy, and a problematic school and work history. Empathy is one of the previously identified risk factors for violent behavior in psychotic patients (Bo et al., 2011) and could be associated with a deficit in the processing of emotional stimuli, often found in patients with psychotic disorders (Fullam & Dolan, 2006).

In sum, this profile is characterized by the lowest risk factors overall, low comorbidity numbers, psychotic disorders, and low PCL-R scores. Treatment could focus on the reduction of psychotic symptoms and the enhancement of skills that are beneficiary for outflow, such as social skills and self-reliance skills. Further studies would have to show whether the treatment

prognosis for patients with a psychotic disorder in this profile differs from patients with a psychotic disorder in the mixed profile.

Limitations

The sample size in this study is rather small; a number of at least 500 respondents would be preferable for performing LCA (Vermunt, 2004). The Bootstrap function was conducted, therefore, to estimate the proper *p* value. Psychopathology and type of offenses were translated into groups for a proper LCA, whereas some other studies differed within those groups, for example, types of schizophrenia (Belli & Ural, 2012). However, the offense and psychopathology variables would have been too large for analysis if we had not grouped these variables. Moreover, the goal was not to study one group of offenders but to study clinical profiles of Dutch forensic psychiatric patients. Although the results show similarities with the nationwide study conducted by Van Nieuwenhuizen et al. (2011), we will not generalize the results to the entire Dutch TBS population.

The risk assessments we selected were conducted between 2005 and 2012. In these years, risk assessments were widely studied and developed, leading to improvements in scoring methods. These may possibly have influenced assessment considerations made over the years. It will be important, therefore, to reassess the results with newly developed risk assessment tools, such as the HKT–Revised (Spreen et al., 2014) in the future.

The future items *Material Indicators*, *Agreement on Conditions*, and *Daily Activities* were excluded as these consider post-treatment conditions, and forensic psychiatric treatment takes approximately 9 years (Van Nieuwenhuizen et al., 2011). However, forensic psychiatric clinics are now developing new treatment modules and earlier release modules to shorten treatment in the future, possibly leading to an earlier focus on these items in the future.

Historical risk factors for child abusers, furthermore, were rather small. It is possible that the HKT-30 and the following HKT–Revised do not assess the proper historical risk factors for this group, although the sexual deviance item also showed low scores. However, paraphilia itself is considered to be a serious risk factor for recidivism (Allan et al., 2007; Hanson & Harris, 2000).

Conclusions

By identifying four patient profiles, we classified the heterogeneous group of forensic psychiatric patients into recognizable groups, which may be a help in studying optimum treatment modules. Three of the four profiles we discovered are in line with previous studies (Bogaerts & Spreen, 2011; Van Nieuwenhuizen et al., 2011), which confirms our clinically substantiated profiles. These three profiles are *the psychotic patient with multiple problems*,

matching with our mixed profile; *the psychotic patient*, which is in line with our psychotic first offender; and *the patient with sexual problems and sexual crimes*, corresponding with the maladaptive affective disordered profile in this study.

The *antisocial patient* and *the patient suffering from addiction* (Van Nieuwenhuizen et al., 2011) were not found in this study. The *antisocial patient* shows similarities with the antisocial class. However, the *antisocial patient* by Van Nieuwenhuizen et al. (2011) typically committed life-threatening crimes, whereas patients in the antisocial class are possibly generalists, as in *the patient suffering from addiction* by Van Nieuwenhuizen et al. (2011). *The patient suffering from addiction* is more often diagnosed with a personality disorder NOS. However, Van Nieuwenhuizen et al. (2011) did not differentiate between personality disorder NOS with severe Cluster B traits or Cluster A or C traits. It is possible that both profiles found by Van Nieuwenhuizen et al. (2011) were integrated into the antisocial class, whereas we studied severity of personality disorder.

Although the maladaptive affective disordered profile shows similarities with *the patient with sexual problems and sexual crimes*, there are some important differences to note. In the maladaptive affective disordered profile, we find not only patients with sexual disorders and sexual crimes but also a higher number of patients with pervasive developmental disorders and higher homicide numbers. However, consistent with the cluster found by Van Nieuwenhuizen et al. (2011), this class does appear to include more "specialist" offenders than "generalist" offenders, as were found in the antisocial and mixed profiles. Sexual misconduct, moreover, is often prevalent in offenders with pervasive developmental disorders (Kawakami et al., 2012; Mouridsen, 2012).

Clinical implications

These results imply different risk factors for different groups and combinations of diagnoses and offense types. Psychotic patients in Class four, for example, show fewer risk factors than patients with psychotic symptoms with a comorbid personality disorder in Class two. Classes show high levels of comorbidity, but therapeutic guidelines do not provide treatment indications in cases of comorbidity (Dell'Osso & Pini, 2012). In line with the RNR model, treatment should focus on the patients' needs, and many treatment modules involve group therapies.

The findings in this study, the study by Van Nieuwenhuizen et al. (2011), and Bogaerts and Spreen (2011) imply that group treatment modules could enhance treatment guidelines, concerning comorbidity and the risk factor combinations found. Group therapy modules could then be refined, providing a better fit with both the need and the responsiveness principles highlighted by the RNR model (Bonta & Andrews, 2007). This could

have the positive effect of better treatment outcomes in a shorter period. Even more profiles could be used to indicate treatment prognosis. Beekman, Van Os, Van Marle, and Van Harten (2012) claimed that diagnostic tools are often used without knowledge of treatment effectiveness and prognosis. If we include relevant treatment characteristics, treatment effectiveness and prognosis could be estimated at the start and during treatment.

Future Research

Future research should examine whether it is beneficial to refine treatment modules and, if so, how they should be refined, and present these to patient groups similar to the classes found in this study. For example, a treatment module for personality-disordered patients with a history of aggressive behavior could be refined by offering a separate treatment module for patients with schizophrenia and severe personality traits. A nationwide study, reassessing our results in a larger forensic psychiatric population, could help to corroborate the results found in this study. Future studies might examine how specific treatment outcomes can be predicted by the profiles.

The presence of a personality disorder is generally related to a worse treatment prognosis. However, does this account for both the antisocial and the mixed profiles, both of which show severe personality traits but different risk factor levels? Routine Outcome Monitoring (ROM) evaluates treatments as a whole, unlike randomized controlled trials, which only evaluate parts of treatments (Zitman, 2012). ROM gives us the opportunity to study treatment progress and patient functioning in different treatment programs at different moments of treatment (Van der Veeken, Bogaerts, & Lucieer, 2012). With a set of ROM tools measuring relevant forensic clinical characteristics, it is possible to study whether treatment progress differs among the four classes and to discover how these factors develop during treatment and what the influence of different treatment modules is. This could aid the further development or refinement of group therapy modules according to "what works" principles (Van Marle, 2012). Treatment progress information for different patient profiles could help to establish treatment indications and prognoses for individual patients made at the start of their treatment (Zitman, 2012). The development of new treatment modules and a better understanding of treatment prognosis could benefit both clinicians and patients.

References

Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse: A Journal of Research and Treatment, 19*, 347-367. doi:10.1007/s11194-007-9052-5

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

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: Matthew Bender.

Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., & Wormith, J. S. (2012). Are the major risk/need factors predictive of both female and male reoffending? A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology, 56*, 113-133. doi:10.1177/0306624X10395716

Beekman, A. T. F., Van Os, J., Van Marle, H. J. C., & Van Harten, P. N. (2012). Stagering en profilering van psychiatrische stoornissen [Staging and profiling of psychiatric disorders]. *Tijdschrift voor Psychiatrie, 54*, 915-920.

Belli, H., & Ural, C. (2012). The association between schizophrenia and violent or homicidal behavior: The prevention and treatment of violent behavior in these patients. *West Indian Medical Journal, 61*, 538-543.

Bo, S., Abu-Akel, A., Kongerslev, M., Haahr, U. H., & Simonsen, E. (2011). Risk factors for violence among patients with schizophrenia. *Clinical Psychology Review, 31*, 711-726. doi:10.1016/j/cpr.2011.03.002

Bogaerts, S., Polak, M., Spreen, M., & Zwets, A. (2012). High and low aggressive narcissism and anti-social lifestyle in relationship to impulsivity, hostility, and empathy in a group of forensic patients in the Netherlands. *Journal of Forensic Psychology Practice, 12*, 147-162. doi:10.1080/15228932.2012.650144

Bogaerts, S., & Spreen, M. (2011). *Persoonlijkheidspathologie, slachtofferschap vóór het 18e levensjaar, huiselijk geweld en delict informatie in een klinische groep terbeschikkinggestelden* [Personality pathology, victimization before the age of 18, domestic violence and crime information in a clinical group of inpatients]. Den Haag, The Netherlands: Onderzoek in opdracht van het Ministerie van Justitie en Veiligheid, WODC.

Bogaerts, S., Spreen, M., Horváth, Z., Polak, M. S., & Cima-Knijff, M. J. (2013). Career criminality, childhood experiences and adult personality among forensic patients in the Netherlands. In J. B. Helfgott (Ed.), *Criminal psychology* (pp. 381-402). Westport, CT: Praeger.

Bogaerts, S., Vanheule, S., & DeClercq, F. (2005). Recalled parental bonding, adult attachment style, and personality disorders in child molesters: A comparative study. *The Journal of Forensic Psychiatry & Psychology, 16*, 445-458.

Bogaerts, S., Vervaeke, G., & Goethals, J. (2004). A comparison of relational attitude and personality disorders in the explanation of child molestation. *Sexual Abuse: A Journal of Research and Treatment, 16*, 37-47. doi: 10.1177/107906320401600103

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and treatment* (User Report No. 2007-06). Ottawa, Ontario: Public Safety Canada.

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

Boutron, N., Bonnet, K., & Mak, T. (1996). Drug abuse: A significant variable in schizophrenia research. *Biological Psychiatry, 39*, 1053-1054. Retrieved from http://www.ncbi. nlm.nih.gov/pubmed/8780842

Buschman, J., Bogaerts, S., Foulger, S., Wilcox, D., Sosnowski, D., & Cushman, B. (2010). Sexual history disclosure polygraph examinations with cybercrime offences: A first Dutch explorative study. *International Journal of Offender Therapy and Comparative Criminology, 54*, 395-411. doi:10.1177/0306624X09334942

Cernkovich, S., & Giordano, P. (2001). Stability and change in antisocial behavior: The transition from adolescence to early adulthood. *Criminology, 39*, 371-410. doi:10.1111/j.1745-9125.2001.tb00927.x

Coid, J., Hickey, N., Kahtan, N., Zhang, T., & Yang, M. (2007). Patients discharged from medium secure forensic psychiatry services: Reconvictions and risk factors. *British Journal of Psychiatry, 190*, 223-229. doi:10.1192/bjp.bp.105.018788

Craig, L. A., Browne, K. D., Beech, A., & Stringer, I. (2006). Differences in personality and risk characteristics in sex, violent and general offenders. *Criminal Behaviour and Mental Health, 16*, 183-194. doi:10.1002/cbm.618

Curson, D. A., Duke, P. J., Harvey, C. A., Pantelis, C., & Barnes, T. R. (1999). Four behavioural syndromes of schizophrenia: A replication in a second inner-London epidemiological sample. *Schizophrenia Research, 37*, 165-176. doi:10.1016/S0920-9964(98)00151-0 Davison, S., & Janca, A. (2012). Personality disorder and criminal behaviour: What is the nature of the relationship? *Current Opinion in Psychiatry, 25*, 39-45. doi:10.1097/YCO.0b013e32834d18f0

Dell'Osso, L., & Pini, S. (2012). What did we learn from research on comorbidity in psychiatry? Advantages and limitations in the forthcoming DSM-V era. *Clinical Practice & Epidemiology in Mental Health, 8*, 180-184. doi:10.2174/1745017901208 010180

De Vogel, V., De Ruiter, C., & Bouman, Y. (2007). Meting van protectieve factoren en het belang ervan voor behandeling in de forensische setting [Measurement of protective factors and the importance within treatment in the forensic setting]. *Tijdschrift voor Psychotherapie, 33*, 102-119. doi:10.1007/BF03062262

De Vries, K., & Spreen, M. (2012). De HKT-30 als instrument voor beslismomenten binnen een tbs-behandeling [Can the clinical items of the risk assessment instrument HKT-30 predict to what extent forensic patients will break agreements during the resocialisation phase of treatment?]. *Tijdschrift voor Psychiatrie, 54*, 429-438. Retrieved from http://www.tijdschriftvoorpsychiatrie.nl/issues/453/articles/9380

Firestone, P., Bradford, J. M., Greenberg, D. M., & Serran, G. A. (2000). The relationship of deviant sexual arousal and psychopathy in incest offenders, extrafamilial child molesters, and rapists. *Journal of the American Academy of Psychiatry and the Law, 28*, 3 03-308.

Fridell, M., Hesse, M., Jaeger, M. M., & Kühlhorn, E. (2008). Antisocial personality disorder as a predictor of criminal behaviour in a longitudinal study of a cohort of abusers of several classes of drugs: Relation to type of substance and type of crime. *Addictive Behaviors, 33*, 799-811. doi:10.1016/j.addbeh.2008.01.001

Fullam, R., & Dolan, M. (2006). The criminal and personality profile of patients with schizophrenia and comorbid psychopathic traits. *Personality and Individual Differences, 40*, 1591-1602. doi:10.1016/j.paid.2006.01.003

Hall, R. C., & Hall, R. C. (2007). A profile of pedophilia: Definition, characteristics of offenders, recidivism, treatment outcomes, and forensic issues. *Mayo Clinic Proceedings, 82*, 457-471. doi:10.4065/82.4.457

Hannem, S. (2013). Experiences in reconciling risk management and restorative justice: How Circles of Support and Accountability work restoratively in the risk society. *International Journal of Offender Therapy and Comparative Criminology, 57*, 269-288. doi:10.1177/0306624X11432538

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior, 27*, 6-35. doi:10.1177/009 38548 00027001002

Haqqi, S. (2010). Substance abuse in Schizophrenia—A short report. *Procedia—Social and Behavioral Sciences, 5*, 2239-2241. doi:10.1016/j.sbspro.2010.07.442

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

Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D., & Newman, J. P. (1990). The revised Psychopathy Checklist: Reliability and factor structure. *Psychological*

Assessment: A Journal of Consulting and Clinical Psychology, 2, 338-341. doi:10.1037/1040-3590.2.3.338

Hare, R. D., & Neumann, C. S. (2005). Structural models of psychopathy. *Current Psychiatry Reports, 7*, 57-64. doi:10.1007/s11920-005-0026-3

Harte, J., & Breukink, M. (2010). Objectiviteit of schijnzekerheid? Kwaliteit, mogelijkheden en beperkingen van instrumenten voor risicotaxatie. [Objectivity or false security? Quality and possibilities of risk assessment tools]. *Tijdschrift voor criminologie, 52*, 52-72. Retrieved from http://hdl.handle.net/1871/21902

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. doi:10.1016/j.ijlp.2012.04.001

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. [The value of structured risk assessment and the diagnosis of psychopathy. A study for reliability and predictive validity]. Utrecht, The Netherlands: Expertise Centrum Forensische Psychiatrie.

Höing, M. A., Bogaerts, S., & Vogelvang, B. (2013). Circles of support and accountability: How and why they work for sex offenders. *Journal of Forensic Psychology Practice, 13*, 267-295. doi:10.1080/15228932.2013.808526

Kawakami, C., Ohnishi, M., Sugiyama, T., Someki, F., Nakamura, K., & Tsujii, M. (2012). The risk factors for criminal behaviour in high-functioning autism spectrum disorders (HFASDs): A comparison of childhood adversities between individuals with HFASDs who exhibit criminal behaviour and those with HFASD and no criminal histories. *Research in Autism Spectrum Disorders, 6*, 949-957. doi:10.1016/j.rasd. 2011.12.005 Laub, J. H., Sampson, R. J., & Sweeten, G. A. (2006). Assessing Sampson and Laub's life- course theory of crime. In F. T. Cullen, J. P. Wright., & K. R. Blevins (Eds.), *Taking stock: The status of criminological theory—Advances in criminological theory* (Vol. 15) 313 - 334. New Brunswick, NJ: Transaction.

Lee, J. K. P., Jackson, H. J., Pattinson, P., & Ward, T. (2002). Developmental risk factors for sexual offending. *Child Abuse & Neglect, 26*, 73-92. doi:10.1016/S0145-2134(01)00304-0

Leistico, A. M. R., Salekin, R. T., DeCoster, J., & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law and Human Behavior, 32*, 28-45. doi:10.1007/s10979-007-9096-6

Lockmuller, M., Beech, A., & Fischer, D. (2008). Sexual offender with mental health problems: Epidemiology, assessment and treatment. In K. Soothill, P. Rogers, & M. Dolan (Eds.), *Handbook of forensic mental health* (pp. 446-479). Cullompton, UK: William.

Mouridsen, S. E. (2012). Current status of research on autism spectrum disorders and offending. *Research in Autism Spectrum Disorders, 6*, 79-86. doi:10.1016/ j.rasd.2011.09.003

Mulder, E., Vermunt, J., Brand, E., Bullens, R., & Van Marle, H. (2012). Recidivism in subgroups of serious juvenile offenders: Different profiles, different risks? *Criminal Behaviour and Mental Health, 22*, 122-135. doi:10.1002/cbm.1819

Murphy, D. (2010). Understanding offenders with autism-spectrum disorders: What can forensic services do? Commentary on . . . Asperger syndrome and criminal behavior. *Advances in Psychiatric Treatment, 16*, 44-46. doi:10.1192/apt.bp.109.006775 Nanayakkara, V., O'Driscoll, C., & Allnutt, S. (2012). HCR-20 risk profiles of patients referred to forensic mental health services in New South Wales. *Psychiatry, Psychology and Law, 19*, 711-719. doi:10.1080/13218719.2011.615819

Neutze, J., Grundmann, D., Scherner, G., & Beier, K. M. (2012). Undetected and detected child sexual abuse and child pornography offenders. *International Journal of Law and Psychiatry, 35*, 168-175.

O'Driscoll, C., Larney, S., Indig, D., & Basson, J. (2012). The impact of personality disorders, substance use and other mental illness on re-offending. *The Journal of Forensic Psychiatry & Psychology, 23*, 382-391. doi:10.1080/14789949.2012.686623 Palermo, G. B. (2009). Editorial: Delinquency: Risks and protective factors. *International Journal of Offender Therapy and Comparative Criminology, 53*, 247-248. doi:10.1177/0306624X09335886

Peterson, J. L., Skeem, J. L., Hart, E., Vidal, S., & Keith, F. (2010). Analyzing offense patterns as a function of mental illness to test the criminalization hypothesis. *Psychiatric Services, 61*, 1217-1222. doi:10.1176/appi.ps.61.12.1217

Rennie, C. E., & Dolan, M. C. (2010). The significance of protective factors in the assessment of risk. *Criminal Behaviour and Mental Health, 20*, 8-22. doi:10.1002/cbm.750 Scoones, C. D., Willis, G. M., & Grace, R. C. (2012). Beyond static and dynamic risk factors: The incremental validity of release planning for predicting sex offender recidivism. *Journal of Interpersonal Violence, 27*, 222-238. doi:10.1177/0886260511416472 Skeem, J. L., Poythress, N., Edens, J. F., Lilienfeld, S. O., & Cale, E. M. (2003). Psychopathic personality or personalities? Exploring potential variants of psychopathy and their implications for risk assessment. *Aggression and Violent Behavior, 8*, 513- 546. doi:10.1016/S1359-1789(02)00098-8

Spreen, M., Brand, E., Ter Horst, P., & Bogaerts, S. (2014). Handleiding en Methodologische Verantwoording HKT-R, Historisch, Klinische en Toekomstige–Revisie [Manual and methodological justification HKT-R historical, clinical and future–Revised]. Groningen, The Netherlands: Dr. van Mesdag kliniek.

Spreen, M., Ter Horst, P. R. M., Bogaerts, S., Lammers, S. M. M., Hochstenbach, J., Soe- Agnie, S. E., & Reubsaet, R. J. (2009). Risicotaxatie aan de hand van de HCR-20 en de HKT-30: Een vergelijking tussen beide instrumenten.[Risk assessment with the HCR-20 and the HKT-30: A comparison between both instruments]. In T. I. Oei & M. S. Groenhuijsen (Eds.), *Forensische psychiatrie en haar grensgebieden: Actualiteit, geschiedenis en toekomst* [Forensic psychiatry and het boundaries: Actuality, history and future] (pp. 459-476). Alphen aan den Rijn, The Netherlands: Kluwer.

Trull, T. J., Sher, K. J., Minks-Brown, C., Durbin, J., & Burr, R. (2000). Borderline personality disorder and substance use disorders: A review and integration. *Clinical Psychology Review, 20*, 235-253. doi:10.1016/S0272-7358(99)00028-8

Tyrer, P., & Johnson, T. (1996). Establishing the severity of personality disorder. *American Journal of Psychiatry, 153*, 1593-1597. doi:10.1176/ajp.153.12.1593

Van der Veeken, F. C. A., Bogaerts, S., & Lucieer, J. (2012). Routine outcome monitoring in forensic psychiatry: FPC 2landen and FPC De Kijvelanden. In T. I. Oei & M. S. Groenhuijsen (Eds.), *Progression in forensic psychiatry* (pp. 355-375). Alphen aan den Rijn, The Netherlands: Kluwer.

Van Marle, H. J. C. (2012). De stagering met profielen in de forensisch psychiatrische delictanalyse [Staging with profiles in offence analysis: Diagnostic steps in forensic psychiatry]. *Tijdschrift voor Psychiatrie, 54*, 973-978.

Van Nieuwenhuizen, C. H., Bogaerts, S., De Ruiter, E. A. W., Bongers, I. L., Coppens, M., & Meijers, R. A. A. C. tbs-behandeling geprofileerd; een gestructureerde casussenanalyse [*ProfilingTBS-treatment: A structured cases analysis]*. Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) Ministerie van Veiligheid en Justitie. 2011. Retrieved from http://www.wodc.nl/onderzoeks database/voor studie-onderzoek-tbs.aspx

Vermunt, J. K. (2004). Toepassingen van latente klassenanalyse in sociaal wetenschappelijk onderzoek. *Sociale Wetenschappen, 47*(1), 2-14. Retrieved from http://members.home.nl/jeroenvermunt/sw2004.pdf

Vermunt, J. K., & Magidson, J. (2005). *LG 4.0 user's guide. Basic and advanced*. Belmont, MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. (2005[1]). *Technical guide for Latent Gold 4.0: Basic and advanced*. Belmont, MA: Statistical Innovations.

Ward, T., & Gannon, T. A. (2006). Rehabilitation, etiology, and self-regulation: The comprehensive good lives model of treatment for sexual offenders. *Aggression and Violent Behavior, 11*, 77-94. doi:10.1016/l/avb.2005.06.001

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

Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Task Force in Forensic Psychiatry]. (2002). *Handleiding HKT-30 versie 2. Risicotaxatie in de forensische psychiatrie* [Guide HKT-30 version 2. Risk Assessment in Forensic Psychiatry]. Den Haag, The Netherlands: Ministerie van Justitie, Dienst Justitiële Inrichtingen.

Woessner, G. (2010). Classifying sexual offenders: An empirical model for generating type-specific approaches to intervention. *International Journal of Offender Therapy and Comparative Criminology, 54*, 327-345. doi:10.1177/0306624X08331213

Wong, S., & Gordon, A. (2006). The validity and reliability of the violence risk scale: A treatment friendly violence risk assessment tool. *Psychology Public Policy Law, 12*, 279-309. doi:10.1037/1076-8971.12.3.279

Yiend, J., Freestone, M., Vazques-Montes, M., Holland, J., & Burns, T. (2013). The clinical profile of high-risk mentally disordered offenders. *Social Psychiatry and Psychiatric Epidemiology, 48*, 1169-1176. doi:10.1007/s00127-013-0696-9

Zadeh, M. D., & Damavandi, A. J. (2010). The incidence of personality disorders among substance dependents and non-addicted psychiatric clients. *Procedia—Social and Behavioral Sciences, 5*, 781-784. doi:10.1016/j.sbspro.2010.07.184

Zitman, F. G. (2012). Stagering, profilering en routine outcome monitoring [Staging, profiling and routine outcome monitoring]. *Tijdschrift voor Psychiatrie, 54*, 979-984.

CHAPTER FOUR

The Instrument for Forensic Treatment Evaluation: Reliability, Factorial Structure and Sensitivity to Measure Behavioral Changes

Frida C. A. van der Veeken
Stefan Bogaerts
Jacques Lucieer

This is an adapted version of the article published as: 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 Practice.* doi: 10.1080/24732850.2018.1468675

Abstract

Routine outcome monitoring is widely used in general mental healthcare but is still in its infancy in forensic psychiatry. As usable forensic ROM tools are lacking, the Instrument for Forensic Treatment Evaluation has been developed.

The goal of this study is to assess the psychometric values of the IFTE and its ability to assess change. Therefore, multiple IFTE assessments of 218 male forensic inpatients were gathered, after which the interrater reliability, test-retest reliability, and internal consistency were assessed. Principal axis factoring was assessed to examine the clinically used factor structure. Patient scores were divided into low and high IFTE factor scores at T0 to examine changeability between T0 and T3.

The results display moderate to good values for the interrater reliability, test-retest reliability, and internal consistency for most items. Factor analyses partly confirmed the original factor structure of the instrument. However, several items loaded on a fourth factor. The factor scores display a moderate to large change over time.

These results mostly support the clinical use of the IFTE in treatment. However, *sexually deviant behavior* should be studied in a less structured setting. Future studies would have to assess the use of the IFTE in supporting treatment decisions and its use in individual treatment.

Introduction

The clinical importance of Routine Outcome Monitoring (ROM) to evaluate treatment in mental health care is generally known (De Beurs & Zitman, 2007). ROM stands for routine evaluations of treatment goals and supports decision-making to improve the quality of treatment (Ellwood, 1988). In forensic psychiatry, ROM is important to decide whether a patient receives modality leave as part of their rehabilitation. Treatment evaluation in forensic psychiatry is still in its infancy, which can be explained by its twofold objective and sometimes contradictory principles: to protect society (security goal) and to rehabilitate offenders to prevent reoffending (treatment goal and risk management) (Van Nieuwenhuizen et al., 2011).

In previous decades, leading rehabilitation models have been introduced and increasingly used for forensic assessment and treatment (Schuringa, Spreen, & Bogaerts, 2014). One of the most important theoretical frameworks in forensic psychiatric rehabilitation is the Risk-Need-Responsivity (RNR) model (Bonta & Andrews, 2007), which consists of three principles: the Risk, Need, and Responsivity principles. The Risk principle assumes that treatment frequency, intensity, and duration should be adjusted to the level of a patient's risk. The Need principle emphasizes that treatment should concentrate on an individual patient's criminogenic needs, more particularly on the reversible dynamic risk factors. The Responsivity principle, finally, emphasizes the therapeutic working alliance between patient and clinician in which a patient's strengths and weaknesses must be considered in order to improve and maintain an optimum treatment relationship (Bonta & Andrews, 2007). The RNR model has appointed eight central risk factors, which are, directly or indirectly, predictive of future criminal behavior (Andrews & Bonta, 2010). These include: a history of antisocial behavior, antisocial personality pattern (i.e., impulsivity, hostility), antisocial cognitions, antisocial associations, substance abuse, family/marital relationships, school/work, and pro-social recreational activities (Bonta & Andrews, 2007; Andrews et al., 2011).

According to Andrews and Bonta (2010), criminogenic needs are closely related to reoffending and treatment should effectuate the reduction of these criminogenic needs (f.e., impulsivity, hostility, lack of empathy, and psychopathology in general) (Wong, Gordon, & Gu, 2007). Treatment evaluation in forensic psychiatry must involve several aspects, such as the clinical domain, rehabilitation, humanitarian perspectives, and public safety, which are very appropriate in forensic psychiatry (Atkisson et al., 1992). The ultimate goal of forensic treatment is to prevent inpatient aggression and outpatient recidivism after release (Kunst, Bogaerts, & Winkel, 2009). Furthermore, forensic psychiatric treatment and rehabilitation should be seen as a phased process in which the risk of reoffending is derived from the severity of specific dynamic risk factors that are subject to treatment.

In the Netherlands, patients are admitted to a forensic psychiatric center (FPC) by order of the state. Patients with a TBS order (*ter beschikkingstelling*; meaning involuntary admission by order of the state) (De Ruiter & Trestman, 2007) have committed a crime with a minimum penalty of four years, under the influence of their mental state and with a high risk of reoffending. Patients can be considered (partly) unaccountable for the offense (Nederlands Instituut Forensische Psychiatrie en Psychology [Dutch Institute for Forensic Psychiatry and Psychology] (NIFP), 2016). A prison sentence may be imposed before patients are admitted to the forensic hospital (De Ruiter & Trestman, 2007). The goal of the TBS order is to protect society and rehabilitate patients into society. Patients receive treatment and structure throughout the day with guidance, labor, educational and leisure activities, and different leave modules, all adding to a patients' rehabilitation, the reduction of risk factors, and the improvement of protective factors and resocialization skills.

To evaluate treatment, adequate evaluation instruments are necessary showing satisfactory psychometric properties, which are sufficiently specific and sensitive to measure and visualize behavioral changes (Wooditch, Tang, & Taxman, 2014). In recent years, several evaluation instruments have been used in forensic psychiatry to assess patients' treatment progress. The Health of the Nations Outcome Scales (HoNOS), the Beck depression inventory and the HoNOS secure (Dickens, Sugarman, & Walker, 2007; Fitzpatrick et al., 2010; Yiend et al., 2011) assess psychiatric clinical symptoms and the need for security (public safety). Quality of life instruments (i.e., the Manchester assessment of Quality of Life (Priebe, Huxley, Knight, & Evans, 1999; Van Nieuwenhuizen, Schene, & Koeter, 2000) assess patients' quality of life (humanitarian).

In forensic psychiatry, however, there are hardly any assessment tools that provide insight into the changeability of problem behavior, protective behavior, and resocialization skills (Alexander et al., 2010; Fitzpatrick et al., 2010; Goethals & Van Marle, 2012; Yiend et al., 2011), while treatment of forensic patients focuses on reducing problematic behaviors, and improving and continuing protective behavior and resocialization skills, with a minimal risk of reoffending (Schuringa et al., 2014).

Therefore, the IFTE (Schuringa, 2011; Schuringa et al., 2014) has been developed to evaluate problematic, protective, and resocialization factors (Schuringa et al., 2014). The IFTE has been designed as a dynamic forensic evaluation tool with items derived from the Historical Clinical Future – Revised (HKT-R; Spreen, Brand, Ter Horst, & Bogaerts, 2014), and the Atascadero Skills Profile (Schuringa et al., 2014). The HKT-R is a validated nationwide Dutch risk assessment instrument for forensic psychiatric inpatients comparable to the HCR-20^{v3} (Bogaerts, Spreen, Ter Horst, & Gerlsma, 2017; Douglas, Hart, Webster, & Belfrage, 2013). The HKT-R must be completed for all patients residing in

a forensic center at least annually and for every new leave movement. The IFTE assesses behaviors that, according to the RNR model, should be diminished or enhanced during treatment. These concern problem behaviors, such as *impulsivity* and *hostility* and protective behaviors, such as good *coping skills* and *medication use* to prevent future recidivism (Rennie & Dolan, 2010; Ward & Gannon, 2006; Palermo, 2009), and resocialization skills, necessary and predictive for good patient functioning outside the forensic setting (Rennie & Dolan, 2010; Ward & Gannon, 2006; Palermo, 2009; De Vogel, De Ruiter, & Bouman, 2007). All skills that can be improved during forensic treatment that should therefore be monitored during treatment in a forensic setting.

A previous study showed good psychometric qualities of the IFTE: the interrater reliability was high, all items showed ICCs higher than .60, and the ICC was almost perfect for seven of the 22 items. The internal consistency of the three factors was high: for problem behavior (α = .86), protective behavior (α = .90), and resocialization skills (α = .88). Test-retest properties were also sufficient for all items and good for the three factor scores, assessed with Cronbach's alpha varying from .85 to .89 (Schuringa et al., 2014).

In this study, we re-assess the psychometric values of the IFTE in two different FPCs. We expect to find results similar to the previous study by Schuringa et al. (2014). Furthermore, we will assess the usability of the IFTE as a ROM tool to measure behavioral change over time. We expect that patients who show more problem behavior and fewer resocialization and protective skills will gain more treatment progress than patients who display better functioning behavior. Therefore, we divided the patients into a low and high problematic group, a low and high resocialization group, and a low and high protective group. To test the psychometric properties of the IFTE and its utility for ROM, interrater reliability, internal consistency, and test-retest reliability were assessed. Exploratory factor analysis (EFA) was computed in order to re-assess whether we find the same factor structure as found by Schuringa et al. (2014).

Method

Participants

This study was based on a clinical inpatient sample of 218 forensic psychiatric patients with a TBS order residing in two Dutch FPCs between 2011 and 2014. All patients were male. Mean age of the inpatients was *M* = 40.7 (*SD* = 10.06) range = 22 – 73 during the first assessment. Most patients were diagnosed with a primary axis-I DSM-IV-tr (American Psychiatric Association (APA), 2000) substance use disorder (41.2%), followed by a primary axis-I schizophrenia disorder (17.9%), followed by another psychotic disorder (6.5%), pedophilia (8.3%), a pervasive developmental disorder (4.7%), a mood disorder (3.7%), or other (12.8%). About five point five percent of the patients did not have an axis-I disorder.

On axis-II, of the DSM-IV-tr (APA, 2000) most patients were primarily diagnosed with a personality disorder not otherwise specified (NOS) (36.7%), followed by an antisocial personality disorder (32.2%), narcissistic personality disorder (5.5%), a borderline personality disorder (5.5%), or other (5.9%). About 14.2 % did not have an axis-II disorder, or a diagnosis on axis-II was postponed.

Most patients had committed a crime against life, including attempted murder (37.1%), followed by assault (15.8%), sexual crimes other than child sexual abuse (15.4%), child sexual abuse (13.1%), theft with violence (9.5%), arson (6.3%), or other (2.3%). The majority of patients had a Dutch cultural background (65.1%) (they were either born in the Netherlands or their parents were), followed by an Antillean cultural background (5.0%), Surinamese (8.7%), Moroccan (5.0%), Turkish (4.1%), or other (11.5%).

Procedure

All data gathered for this study were primary treatment information and were retrieved from patients' ROM files and electronic patient files (EPF). Results are reported at group level and cannot be traced to an individual patient. This study is part of a ROM study in forensic psychiatry and has been approved by the scientific research committee of the FPC where this study was conducted. Ethical standards were considered during this study, and informed consent was not retained while all data concern primary treatment information as part of clinical ROM.

The ROM measurements including the IFTE started in September 2011 for approximately 40% of all patients in FPC De Kijvelanden, and for 20% of all patients in FPC 2landen. In 2012, measurements were implemented for all patients in both Forensic

Centers. At both centers, the indicators of symptom level, daily functioning, quality of life, and risk were assessed biannually (Van der Veeken, Bogaerts, & Lucieer, 2012). All ROM assessments were conducted before an individual patients' multidisciplinary treatment evaluation in order to evaluate treatment. These evaluations take place every four to six months in order to discuss and possibly adjust the patients' treatment plan.

As the IFTE is a multidisciplinary tool, multiple disciplines are involved in IFTE assessments: a staff member or personal coach (patients receive guidance on their ward by staff members and their personal coaches), psychologists or psychiatrists, art or psychomotor therapists, other therapists, and labor consultants are asked to perform an IFTE in preparation of the multidisciplinary treatment evaluation. IFTEs are performed to evaluate behavior in-between two treatment evaluation meetings. After IFTE assessment, a report is constructed including the composite score. Treatment teams have access to these reports and can use them in preparation of treatment evaluations. All ROM questionnaires were scored in an Excel document pertaining to an individual patient. IFTEs were copied into these Excel documents with shortened instructions for therapists to assess the IFTE. Individual treatment reports are also constructed in these Excel documents. IFTEs assessed in advance of a multidisciplinary treatment evaluation by at least one psychologist/psychiatrist or staff member/coach were selected to be included in the analysis.

The Instrument for Forensic Treatment Evaluation

The IFTE is an observation tool developed to measure treatment progress in forensic psychiatric patients on both risk and protective factors. The IFTE consists of 22 items, including four items related to more specific treatment issues: *drug use*, *skills to prevent substance use, skills to prevent physical aggression*, and *skills to prevent sexually deviant behavior*. Multiple disciplines can score the IFTE individually, based on their own observations of the patients' behavior (Schuringa et al., 2014), resulting in a composite score. Apart from the items based on the ASP and the HKT-R, clinically relevant items were added, namely *balanced daytime activities*, *financial skills, manipulative behaviors, sexually deviant behavior, and medication use* (Schuringa et al., 2014).

The IFTE operates a five-point scale, but as three in-between scoring options are possible between every two scoring options, this creates a 17-point scale (see Figure 1). The in-between scoring options give the IFTE a more sensitive character, while a five-point scale has a more limited sensitivity to observe behavioral change. The in-between scoring options also provide more accurate scoring opportunities (Schuringa et al., 2014). Some studies claim that a 7-point scale should be the maximum; however, Leung (2011), who studied the use of higher Likert scales, suggests that an 11-point Likert scale or higher can easily

be used (Schuringa et al., 2014). Preston and Colman (2000) found similar results when studying the validity of scales from two to 101 answer scales.

A higher score on an IFTE item indicates "more behavior", meaning more problematic behavior on the factor *problem behavior* (i.e., more impulsivity or more psychotic symptoms) or more resocialization skills (i.e., better social skills or better labor skills). Lower scores indicate less problematic behavior or fewer protective skills (i.e., inadequate copings skills) or resocialization skills (inadequate labor skills, social skills). All IFTE items include the box "not enough information" (N.E.I.), taking into consideration that multiple disciplines can score the IFTE, but that not all of these may have information on every aspect of the patients' functioning. Some items are not applicable to all patients; these include the box "not applicable (N.A.)". For example, not all patients have a psychotic disorder or use psychiatric medication. This means that not all items have been scored for all included patients, resulting in a differing N within the analysis.

Figure 1
Example of an IFTE item

3 **Does the patient acknowledge his crime, and does he take responsibility?**

N.E.I.	This item enhances whether the patient accepts and acknowledges responsibility for the committed crime or crimes. Denial or responsibility can be part of a more general denying and trivializing attitude, but can also solely concern the committed crime. Circumstances the patient can plead to avoid responsibility are drugs- or alcohol use "I drank so much I was not aware of my actions".

Denies — Partial — acknowledges

0	•	•	•	**1**	•	•	•	**2**	•	•	•	**3**	•	•	•	**4**

0	**Denies the crime or his part completely**.
1	Displays himself shallow or distant from the crime. Trivializes the consequences and displays a denying attitude.
2	Partially takes responsibility; hides behind co-perpetrators or circumstances.
3	Largely recognizes and takes responsibility for the committed crime.
4	**Acknowledges and accepts crime and responsibility for the crime.**

Statistics

The clinically assessed scores were used in the analysis. The composite scores of multiple disciplines were used for all analyses, apart from the interrater reliability analysis. These are the scores that were used and interpreted in the clinical setting and that should, therefore,

be tested in the analysis. Data were derived from the clinically used Excel documents and loaded into the statistical package of the social sciences 19 (SPSS 19).

Interrater reliability

The interrater reliability was conducted per item and assessed between psychologist/psychiatrist and a professional coach. Intra-class correlation coefficients (ICC) were conducted with a two-way randomized model for consistent measures which reflected the average measures. The ICC assesses the consistency between measures (Field, 2009); while items were rated by multiple raters and the mean score was used within treatment evaluations, average measures were chosen (Hallgren, 2012). Consistent measures were chosen due to the large 17-point scale. In order to achieve the highest number of sets, patients with multiple measures were included multiple times (one to four times). The degree of reliability was considered fair at .21 - .40, moderate at .41 - .60, substantial at .61 - .80, and almost perfect at .81 – 1.00 (Landis & Koch, 1977).

Test-retest reliability

In order to assess test-retest reliability, we selected measurements assessed at the start of the ROM implementation in September 2011 with a period of approximately three months, conducted by a coach and/or psychologist/psychiatrist between t0 and t1. Test-retest reliability is "the ability of a measure to produce consistent results when the same entities are tested at two different points in time" (Field, 2009, p. 795). Three months seems a rather short period for assessing behavioral changes in patients who receive treatments with a duration of approximately nine years (Van Nieuwenhuizen et al., 2011). Even more so, previous analyses showed very marginal dynamic changes in a three-month period (Van der Veeken, 2011). IFTE assessments implemented in September 2011 were implemented for all patients residing in the selected wards, without consideration of the time patients had already been in treatment. The intra-class correlation coefficient was calculated for every item. A two-way random model was selected, while the IFTE does not need to be rated by the same person at all times, hence the multidisciplinary scoring.

Factor analysis

Explorative factor analysis (EFA) was conducted on the 22 IFTE items in order to re-assess the factor structure found by Schuringa et al. (2014). The composited scores of the first patient assessments where used for the analysis. Due to the large number of missing values (see Table 3 for N per item), an EFA was chosen over confirmatory factor analysis (CFA), while it was not possible to assess CFA with all 22 IFTE items. Missing values were not

missing at random, tested with Little's MCAR test, and could, therefore, not be replaced. Parallel analysis (PA) as described by O'Connor (2000) was conducted in order to assess the number of factors.

Internal consistency

Internal consistency was assessed with Cronbach's alpha (Field, 2009), per factor. Again, the composited scores of the first patient assessments were used. Values near .80 were considered good (Field, 2009).

Changeability of behavior

In the clinical assessment of behavioral change, the composited score between multiple raters (assessments with at least one psychologist/psychiatrist or one staff member/coach assessment) was used. While several therapists perform the IFTE, the month of the last IFTE per assessment was selected to calculate the number of months between assessments. Changeability was assessed between t0 (a patients' first assessment) and t1, t2, and t3 (patients' first through fourth assessments were gathered for the analyses). Analyses were conducted for the three factors, namely *problem behavior, protective behavior* and *resocialization skills*. Per factor, two groups were classified, based on their level of functioning: low and high protective behavior, low and high problem behavior, and low and high resocialization skills. As the IFTE has no normative data, we used the descriptive mean-score per factor, to split the patient group into the earlier mentioned level of functioning groups. Effect size *d* was conducted from dependent sample t-tests to assess the effect size of change over time for both groups per factor, with the formula tc described in Dunlop, Cortina, Vaslow and Burke (1996) calculated with the computation tool (psychometrica.de). An effect size *d* of .10 to .30 is considered a small effect, .30 to .50 a medium effect, and .50 or larger a large effect (Field, 2009).

Results

Interrater reliability

The numbers of scored items differ between items due to the N.E.I. and N.A. scoring options. *N* varies from 105 to 449. *Z*-scores for Skewness and Kurtosis varied widely (*Z* Skewness = -13.21 to -0.64 and .26 to 15.50, *Z* Kurtosis = -.35 to -4.3 and .31 to 11.38), with extreme values for the items *sexual deviant behavior* and *psychotic symptoms* (*Z* Skewness = 15.50 to 20.98, *Z* Kurtosis = 23.75 to 29.25). As the assumption of normality was

violated, we added Spearman's correlation coefficient r_s to complement the results. Apart from the items *antisocial associates* and *skills to prevent physically aggressive behavior,* all items showed at least a substantial agreement (see Table 1), and the items *balanced daytime activities, labor skills,* and *drug use* even showed a very good agreement (see Table 1) (Landis & Koch, 1977).

Table 1
Interrater reliability and confidence interval of the IFTE items and N per item

Item	*N*	ICC	95% confidence interval	r_s
Problem insight	440	.722*	.665 - .769	.562*
Cooperation with treatment	436	.810*	.770 - .842	.654*
Responsibility for the crime	396	.715*	.653 - .766	.515*
Coping skills	449	.722*	.665 - .769	.566*
Balanced daytime activities	420	.868*	.840 - .891	.702*
Labor skills	248	.894*	.864 - .918	.648*
Social skills	443	.674*	.607 - .729	.505*
Skills to take care of oneself	431	.785*	.741 - .822	.555*
Financial skills	225	.753*	.679 - .810	.573*
Impulsivity	430	.698*	.635 - .750	.521*
Antisocial behavior	422	.744*	.690 - .788	.608*
Hostility	437	.708*	.647 - .758	.519*
Sexually deviant behavior	425	.635*	.559 - .698	.462*
Manipulative behavior	406	.751*	.698 - .795	.611*
Compliance with rules	432	.649*	.576 - .709	.582*
Antisocial associates	358	.501*	.386 - .595	.227*
Medication use	194	.767*	.691 - .825	.502*
Psychotic symptoms	172	.820*	.756 - .866	.623*
Skills to prevent substance use	245	.760*	.691 - .813	.580*
Drug use	301	.893*	.865 – .914	.766*
Skills physically aggressive beh**	235	.585*	.463 - .679	.443*
Skills sexually deviant beh***	105	.684*	.535 - .785	.533*

ICC = Intra-class correlation coefficient for two random model on average measures, *sig, p<.01, r_s = Spearman's correlation coefficient, **Skills to prevent physically aggressive behavior, ***Skills to prevent sexually deviant behavior

Test-retest reliability

Table 2 shows test-retest reliability assessed with the intra-class correlation coefficient on average measures. Ninety-six assessments on two follow-up measurements (t0 and t1) were selected to conduct the test-retest reliability. For this smaller group of patients, the mean age of selected patients was *M* = 40.16 (*SD* = 9.95), range = 22 – 69. Mean treatment period at the first IFTE assessment for patients was 46.83 months (*SD* = 27.72, range = 3 - 165). Patients were diagnosed with a substance use disorder (51%), schizophrenia (2.1%), another psychotic disorder (7.2%), pedophilia (9.3%), pervasive developmental disorder (9.3%), or other (14.3%). Six point three percent did not have an axis one diagnoses. On axis two, patients were diagnosed with an antisocial personality disorder (30.1%), a borderline personality disorder (8.4%), or a narcissistic personality disorder (4.2%). Forty-six point nine percent was diagnosed with a personality disorder not otherwise specified or other (4%); six point two percent did not have an axis two diagnosis, or the diagnosis was postponed. Patients committed homicide (37.5%), assault (13.5%), property offenses with violence (10.4%), sexual offenses other than child sexual abuse (14.6%), child sexual abuse (16.7%), arson (6.3%), or other (1.0%). Again, the majority of patients had a Dutch cultural background (74.%) (either they were born in the Netherlands or their parents were), followed by Antillean (6.3%), Moroccan (3.1%), Turkish (3.1%), or other (12.5%). Table 2 shows test-retest results per item.

All items show an ICC > .55. The sample size (N) varies between items due to the N.E.I. and N.A. scoring options (see Table 2). The *Z*-scores for Skewness and Kurtosis varied again (*Z* Skewness = -.14 to -9.18 and .12 to 7.69, *Z* Kurtosis = -.07 to -2.34 and .06 to 11.51), with extreme values for the item *psychotic symptoms* (*Z* Skewness = 11.11 and 12.30, *Z* Kurtosis = 19.46 and 23.32). As the assumption of normality was violated, Spearman's correlation coefficient r_s was also calculated.

Table 2

Test-retest reliability of the IFTE items, confidence interval, and N per item

Item	*N*	ICC	95% Confidence interval	r_s
Problem insight	95	.861*	.791 - .907	.767*
Cooperation with treatment	95	.850*	.774 - .900	.725*
Responsibility for the crime	90	.834*	.747 - .890	.673*
Coping skills	95	.823*	.735 - .882	.663*
Balanced daytime activities	95	.852*	.778 - .902	.655*
Labor skills	84	.906*	.855 - .939	.777*
Social skills	95	.838*	.757 - .892	.740*
Skills to take care of oneself	95	.892*	.838 - .928	.675*
Financial skills	91	.845*	.766 - .898	.701*
Impulsivity	95	.907*	.860 - .938	.817*
Antisocial behavior	90	.917*	.873 - .945	.852*
Hostility	95	.862*	.793 - .908	.750*
Sexually deviant behavior	94	.805*	.706 - .870	.734*
Manipulative behavior	94	.826*	.738 - .884	.701*
Compliance with rules	95	.791*	.687- .861	.687*
Antisocial associates	95	.824*	.736 - .883	.677*
Medication use	52	.792*	.637 - .880	.528*
Psychotic symptoms	66	.571*	.299 - .737	.269*
Skills to prevent substance use	65	.776*	.632 - .863	.679*
Drug use	75	.863*	.784 - .914	.723*
Skills physically aggressive beh**	72	.655*	.449 - .784	.526*
Skills sexually deviant beh***	29	.840*	.660 - .925	.707*

Intra-class correlation coefficient for two way random model on average measures *p<.01, r_s = Spearman's correlation coefficient
**Skills to prevent physically aggressive behavior, ***Skills to prevent sexually deviant behavior

Factor analysis

N varied between 69 to 218, *Z*-scores of Skewness and Kurtosis also varied between items (*Z* Skewness = -.55 to -6.96 and .12 to 10.13, *Z* Kurtosis = -.16 to -3.21 and .7 to 7.64), with extreme values for *sexually deviant behavior* and *psychotic symptoms* (*Z* Skewness = 14.31 and 15.58, *Z* Kurtosis = 26.97 and 20.86). EFA based on eigenvalues recommends a factor analysis with five factors. PA for principal factor analysis with permutations as described by O'Connor (2000) was conducted. PA also suggested a model with five factors. However, to assess PA, the item *skills to prevent sexually deviant behavior* could not be included. PA with the item *skills to prevent sexually deviant behavior* was not possible due

to the large number of missing values. PA with all IFTE items apart from *skills to prevent sexually deviant behavior* resulted in a PA conducted for 67 cases. While we considered 67 cases to be a small number of participants, PA was also conducted without the NA items *medication use, skills to prevent substance use, drug use*, and *skills to prevent physically aggressive behavior*. This analysis suggested a three-factor model for 120 cases.

A first principal axis factoring was conducted with a direct oblimin rotation for five factors. Direct oblimin rotation was selected while factors could correlate. The five factor solution explained 60.3% of the variance. The structure matrix displayed that the fifth factor held only one item, namely the item *sexually deviant behavior* with a loading of .36. Therefore we decided to run a principal axis factoring with four factors. Kaiser-Meyer-Olkin measure (KMO) indicated a good fit for the analysis, KMO = .80 (Field, 2009). Bartlett's test of sphericity $\chi 231)\ ^{2})$ = 504.882, p <.01 indicated that correlations were sufficiently large between items. The KMO values for the individual items all load above the limit of .50 (Field, 2009). The four-factor model explained 56% of the variance. Factor one displayed an explained variance of 35.8%, factor two displayed an explained variance of 10.06%, factor three an explained variance of 5.18% and factor four displayed an explained variance of 4.93%. Table 3 shows factor loadings on both the pattern and the structure matrix. Both the pattern and the structure matrix have been displayed, while it is possible that factors relate to each other, which could lead to suppressed values in the pattern matrix (Field, 2009). The structure matrix takes the relationship between factors into account, but it is advised to report both the pattern and structure matrix when the structure matrix is interpreted (Field, 2009).

As cross loadings are shown for several items, EFA was also conducted in Mplus to assess whether a two, three or four factor model would display a better fit. As the item *sexual deviant behavior* did not load on any of the factors and the item *skills to prevent sexual transgressive behavior* had a very low N, EFA was conducted without these items. EFA with Mplus shows that a two factor model displays less cross loadings, however the fit indices do not display a good fit (CFI = .78, TLI = .72, SRMR = .08, χ^2 = 658.02 p < .01) according to the eigenvalues, EFA with Mplus also indicated a four factor model as the best fit, however model fit is limited (CFI = .92, TLI = .87, SRMR = .05, χ^2 = 289.87 p < .01).

Internal consistency

The internal consistency for the separate clinically used factor distribution displayed good results: *problem behavior* α = .81, N = 133; *protective behavior* α = .83, N = 24, and *resocialization skills* α = .84, N = 163.

Table 3
Principal axis factoring of the Instrument for Forensic Treatment Evaluation factor loadings

	Factor	*N*	Pattern matrix				Structure matrix			
			Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3	Factor 4
Problem insight	Prot	215	**.83**	-.01	-.07	-.04	**.86**	-.24	-.48	.18
Cooperation with treatment	Prot	215	**.61**	-.15	-.08	.20	**.74**	-.42	-.45	.42
Responsibility for the crime	Prot	205	**.85**	.02	.03	-.12	**.80**	-.15	-.36	.07
Coping skills	Prot	217	**.43**	-.23	-.36	.11	**.70**	-.50	-.66	.38
Balanced daytime activities	Reso	217	.14	.10	**-.65**	.29	.50	-.25	**-.75**	.40
Labor skills	Reso	176	.22	.09	**-.61**	.27	.56	-.26	**-.75**	.40
Social skills	Reso	217	.51	-.21	**-.28**	-.01	.70	-.43	**-.59**	.25
Skills to take care of oneself	Reso	218	.15	.06	**-.62**	-.12	.40	-.10	**-.65**	.00
Financial skills	Reso	191	-.12	-.14	**-.69**	.06	.28	-.33	**-.69**	.22
Impulsivity	Prob	216	.15	**.80**	.26	.06	-.18	**.81**	.40	-.29
Antisocial behavior	Prob	217	-.21	**.80**	-.03	-.04	-.43	**.87**	.31	-.43
Hostility	Prob	217	-.15	**.67**	.04	-.10	-.38	**.76**	.32	-.43
Sexually deviant behavior	Prob	217	-.03	**.15**	.21	.01	-.17	.21	.26	-.10
Manipulative behavior	Probl	211	-.02	**.68**	-.13	-.06	-.16	**.67**	.08	-.33
Compliance to rules	Probl	217	.24	**-.29**	-.05	.49	.46	**-.58**	-.34	.68
Antisocial associates	Probl	202	.01	.23	-.14	**-.41**	-.08	.36	-.00	**-.48**
Medication compliance	Prot	156	**.53**	-.02	.03	.00	**.52**	-.16	-.24	.13
Psychotic symptoms	Probl	176	-.01	-.02	.65	.10	-.30	.13	.63	-.02
Skills to prevent substance use	Prot	144	.26	.10	-.05	**.77**	.43	-.31	-.29	**.79**
Drug use	Probl	174	.25	.04	.08	**-.83**	-.00	.34	.12	**-.80**
Skills physically aggressive beh**	Prot	162	.17	-.24	-.34	**.28**	.47	-.50	-.54	.49
Skills sexually deviant beh***	Prot	69	**.68**	.02	-.07	.08	**.73**	-.22	-.41	.25

*Prot = Protective skills, Prob = problem behavior, Reso = Resocialization skills, **Skills to prevent physically aggressive behavior, ***Skills to prevent sexually deviant behavior

Changeability of the IFTE over time

The factor *problem behavior* (*M* = 38.83, *SD* = 16.27, *N* = 197) was computed without the N.A. items *drug use* and *psychotic symptoms.* The factor *protective behavior* (*M* = 50.32, *SD* = 14.77, *N* = 146) was computed without the items *skills to prevent substance use*, *skills to prevent physically aggressive behavior,* and *skills to prevent sexually deviant behavior*, because the excluded items were not applicable for all patients, and inclusion would lead to a very small sample size. All items could be included in the factor *resocialization skills* (*M* = 57.51, *SD* = 15.52, *N* = 165). However, due to the "not enough information" option, the number of participants may vary between analyses, leading to differing means at T0 for the different analyses. *Z*-scores for Skewness and Kurtosis calculated for the score differences between factors, display a normal distribution for most score differences (A *Z*-score for Skewness and Kurtosis -2 < 0 > 2), apart from the score difference for high protective behavior between T0-T1 (*Z* Skewness = -2.97, *Z* Kurtosis = 3.24), and for the score difference for low problem behavior between T0-T1 (*Z* Skewness = 4.78, *Z* Kurtosis = 8.90) and T0-T2 (*Z* Skewness = 3.47, *Z* Kurtosis = 1.96), and for the score differences for high resocialization skills between T0-T1 (*Z* Skewness = -5.05, *Z* Kurtosis = 4.79), between T0-T2 (*Z* Skewness = -7, *Z* Kurtosis = 12.03), and between T0-T3 (*Z* Skewness = -4.96, *Z* Kurtosis = 6.61).

Table 4 shows means, dependent t-test statistics, and the effect sizes over four evaluation moments, which refers to a treatment time of approximately 5 to 16.72 months, differing between groups. The factor high problem behavior shows a small effect over time and low resocialization skills shows a medium effect over time. A lower score on *problem behavior* indicates less problematic behavior, whereas a higher score indicates more problematic behavior. Therefore, a decrease is desirable here while, on the other two factors, an increase in scores is desirable. Low protective behavior shows a medium effect for an increase in skills between t0 and t1 and t0 and t2, whereas the effect is large between t0 and t3. Low problem behavior and high resocialization skills both show a medium to large effect over time, and high protective behavior shows a medium effect, implying a decrease in protective behavior and resocialisation skills and an increase in problematic behavior.

Table 4

Changeability over three measurements for high and low scores; number of participants, means, mean period between assessments, and effect size

Factor (high or low)	Low protective behavior	High protective behavior	Low problem behavior	High problem behavior	Low resocialization skills	High resocialization skills
t0-t1						
N	55	56	82	74	43	78
M (SD) - M (SD)*	38.29 (8.41) - 41.37 (13.01)**	62.19 (8.01) - 59.48 (11.43)**	26.17 (7.88) - 28.66 (11.41)**	52.63 (11.17) - 49.70 (12.75)**	41.83 (12.23) - 44.46 (15.58)**	68.90 (5.46) - 67.19 (9.14)
M months	5.62 (1.95)	5.20 (2.19)	5 (2.33)	5.12 (1.82)	5.19 (2.09)	4.32 (1.66)
t, r	-2.21, .608	2.051, .529	-225, 509	2.35, .606	-2.14, .859	1.81, .437
d	-0.26	0.27	-.25	0.24	-0.17	0.22
t0-t2						
N	36	39	66	60	35	64
M (SD) - M (SD)*	39.22 (8.04) - 44.23 (12.15)***	62.41 (7.74) - 59.93 (9.36)	26.91 (7.62) - 31.32 (12.66)***	52.31 (10.87) - 48.64 (15.07)**	43.32 (10.50) - 48.06 (16.69)**	69.04 (5.74) - 65.15 (11.41)***
M months	11.00 (2.69)	10.64 (3.08)	10.39 (3.13)	11.22 (2.62)	10.68 (2.80)	9.84 (2.67)
t, r	-2.76, .481	1.53, .310	-3.10, .441	2.47, .649	-2.36, .707	2.78, .291
d	-0.47	0.29	-.40	0.27	-0.31	0.41
t0-t3						
N	21	25	42	39	19	48
M (SD) - M (SD)*	38.06 (8.64) - 47.14 (13.89)***	61.71 (6.44) - 59.46 (9.54)	27.53 (6.99) - 32.60 (10.68)***	51.68 (11.47) - 47.77 (17.50)	43.08 (10.51) - 48.20 (17.56)	68.89 (5.41) - 63.84 (11.33)***
M months	16.52 (3.01)	16.04 (3.76)	15.31 (3.01)	16.72 (2.97)	17.21 (3.36)	15.50 (2.87)
t, r	-3.66, .53	0.98, -.049	-2.86, .209	1.67, .557	-2.04, .811	3.11, .252
d	-0.74	0.33	-.56	0.25	-0.29	0.55

*Means of included participants are represented in the analyses. ** sig, $p<.05$ *** sig. $p < .01$

Discussion

In this study, the interrater reliability, test-retest reliability, internal consistency, factor distribution, and the ability to measure changes over time were examined for the IFTE. Most results of the study are promising. With the exception of the item *antisocial associates,* and *skills to prevent physically aggressive behavior* the interrater reliability is moderate to good (Landis & Koch, 1977); especially for the items *drug use*, *balanced daytime activities,* and *labor skills*, the interrater reliability is good. If we consider that the IFTE is a dynamic tool, the test-retest reliability is high, with the exception of the items *psychotic symptoms* and *skills to prevent physically aggressive behavior*. The item *medication use* showed a low non-parametrical test-retest reliability. The internal consistency of the three factors shows good results (Field, 2009). With respect to the instrument's sensitivity for measuring behavioral change over time, patients were divided into high and low factor scores, which gives a medium to large effect of treatment progress across time (Field, 2009).

As could be expected, the intra-class correlation values in this study were lower than the values Schuringa et al. (2014) found, while observers (i.e., psychologist, personal coach) did not observe the patients' behavior in the same environment at all times, whereas the study by Schuringa et al. (2014) studied interrater reliability between therapists of the same discipline. Considering the values found and the fact that different disciplines assessed the IFTE and that the variance is rather high, we can consider the interrater reliability coefficients found in this study to be moderate to good, apart from the item *antisocial associates* and *skills to prevent physically aggressive behavior*. However, this also shows the importance of multiple disciplines assessing the items, as they show lower ICCs than those found by Schuringa et al. (2014). Multiple raters could possibly complement each other's scores. However, as the item *antisocial associates* shows a noticeably low interrater reliability, the item is possibly not fully observed by all disciplines in a forensic setting, or interpreted differently, leading to a lower agreement between raters. Apart from the items *psychotic symptoms* and *skills to prevent physically aggressive behavior,* the test-retest values are good. Psychotic symptoms are dynamic (Van Os et al., 2006), and medication or refusal to take medication could be of great influence on the remission or presence of florid symptoms. This could partly explain the low test-retest value of the item *psychotic symptoms.* Psychotic symptoms should be monitored at all times, which occurs in a forensic psychiatric institution as patients are guided throughout the day. However, the IFTE aims to evaluate the patients' functioning over a particular period. A low score on the *psychotic symptoms* item would imply a longer period of remission, and a higher score would then imply a period of florid symptoms, meaning the patients' symptoms have not been in remission for a longer

period. To consider remission of psychotic symptoms, they should be absent for six months (Van Os et al., 2006). However, this assumption should be studied in a larger sample with patients coping with psychotic symptoms.

The interrater reliability and test-retest reliability results indicate that the item *skills to prevent physically aggressive behavior* may need a clearer explanation, while patients cannot always display all required skills inside the FPC. This may complicate the scoring of the item, as when patients show proper behavior inside but did not show their skills in high-risk situations (a maximum score). The lower values could also indicate a possible bias within this item, which could counteract the reliability.

While the factor analysis needs to be interpreted with caution, factor loadings largely show similarities with the results found by Schuringa et al. (2014). However the items *antisocial associates, skills to prevent substance use,* and *drug use* seem to load on the extra fourth factor. Perhaps these items display a different construct than merely problem behavior. The item *sexual deviant behavior* does not seem to load on any of the four factors. Within the analyses the more clinically usable factor solution was chosen. However within this sample the patients' scores do not seem to fully correspond with this solution. Psychotic symptoms are related to violence (Coid et al., 2013; Bogaerts, Vanheule, & DeClercq, 2005) and are therefore important to monitor in relation to problematic behavior. However, the diagnosis itself is not an established risk factor per se (Fazel & Yu, 2011), and a remission in psychotic symptoms can be considered as being contributive to a patient's resocialization, hence the higher factor score on the resocialization factor, similar as to results found by Schuringa et al. (2014). *Skills to prevent physically aggressive behavior* also loaded on the other factors, similar to the factor analysis by Schuringa et al. (2014). The rationale for putting *psychotic symptoms* in the problematic factor is that a higher level of psychotic symptoms could lead to problematic behavior (Schuringa et al., 2014). *Skills to prevent substance use* and *skills to prevent physically aggressive behavior* are considered protective factors during treatment and are, therefore, placed in the protective factor (Schuringa et al., 2014). Even though the item *social skills* did load on the factor *resocialization skills*, a better value was found on the factor *protective behavior*. Social skills can be beneficiary for the resocialization of a patient, but also for treatment as a whole. The *sexually deviant behavior* item did not load on any of the factors, which is possibly due to the low level of sexually deviant behavior that was observed or recorded in our population and, therefore, the generally low score on this item.

Sexually deviant behavior is known to be a highly problematic factor. However, the absence of sexually deviant behavior can be seen as a positive skill. The results indicate that severe sexually deviant behavior is possibly hardly performed in the FPCs at all, which may imply that this item is not compatible with the highly structured clinical setting. However,

the possibility that the item does not describe sexual deviant behavior as it is perceived within a clinical setting should also be considered. It is possible that therapists get more accustomed to certain remarks made within a forensic setting. Verbally sexual transgressive behavior is described by a score of five on the seventeen point scale and most often a score of one is assessed, meaning no sexual transgressive behavior. However the question is, if verbally sexual remarks are made so few. Furthermore, patients mostly learn skills to prevent risky behavior in line with the RNR model, measured with the item *skills to prevent sexually deviant behavior*. The absence of sexually deviant behavior is required outside the FPC, but it is important to know whether a patient does or does not show problematic behavior in treatment (Buschman et al., 2010; Surjadi, Van Horn, Bogaerts, & Bullens, 2010). Therefore, the item fits in the problematic factor and forensic treatment evaluation clinically. For the item *sexually deviant behavior,* it should be tested if the item is more compatible with a less secure and less structured setting.

As PA is a validated and often recommended procedure to assess the number of factors (O'Connor, 2000), PA was conducted in order to assess the correct number of factors for EFA. Although PA indicated a five or three-factor model it was not possible to assess PA with all the IFTE items for a large group of participants due to the large number of missing values. We chose the option "exclude cases pairwise" due to the large number of missing values. However, the missing values were not missing at random according to Little MCAR's test, and these results, therefore, should be interpreted with extreme caution. The number of missing items was very high for a substantial part of the items. A preferable sample would include a minimum of 200 participants, and a higher sample would yield better results. It would be even better to assess factor analysis for a large group of participants, with a minimum of 500 or possibly 1,000 participants. It would also be possible to assess different models with items that are applicable to a group of patients, leading to models that are specific for different patient groups, such as patients with a past of sexually deviant behavior. Normally, these items could be removed from the analysis (Hooper, Coughlan, & Mullen, 2008), but the items do measure important clinical forensic behaviors and are of important clinical value. In line with the RNR model, it is highly important to assess patient-specific characteristics and skills. However clinically useful, therefore, the optional items may affect the statistical results.

It should also be noted that several items display high crossloadings. Looking at the factorloadings a two factor solution could also possibly fit the data, however EFA with Mplus did not display a good fit for the two factor solution. A larger sample with different groups of patients could give answer to the question wat the best factor solution is, considering both clinically and statistically logical factors.

In addition, the analyses cautiously show that the IFTE can assess behavioral changes over time, which is a crucial aspect of a ROM tool (Stinckens, Smits, Claes, & Soenen, 2012). Based on the mean, the scores were divided into a high and low group for the three factors. The problematic group was expected to show more progress, simply because they had more to achieve, while the better functioning patients were already showing better scores. As treatment in a forensic institution can take a long time (approximately 8.4 years (Taskforce behandelduur tbs [Taskforce treatment duration TBS], 2014)) and patients have complex diagnoses and co-morbidity, a relapse is not unusual in forensic psychiatry, which explains the moderate decrease in the better functioning group. The complexity of diagnoses and a long treatment period imply that patients do not show progress on all items in the same period, which explains the low change in the patient group as a whole and the high variance in scores for both groups. Nonetheless, the results show a moderate effect over time and even a large effect for the *protective behavior* factor. The IFTE itself is an evaluation tool for individual change, and feedback reports enhance individual progress.

In conclusion, this study implies that the IFTE is useful in a forensic psychiatric clinical setting to evaluate patients' treatment with multiple raters on a half-yearly basis. The effect size *d*, controlling for correlation effects was computed to assess the effect size (Dunlap, Cortina, Vaslow, & Burke, 1996). Cronbach's alpha assesses whether items measure the same construct (Bland & Altman, 2002) and is a useful coefficient assessing internal consistency: the higher the items correlate, the higher the variance will be, leading to a higher Cronbach's alpha (Bland & Altman, 1997). However, as the assumptions of normality where violated, results should be interpreted with care. Pearson correlation is a widely used method for assessing test-retest reliability. However, the Pearson correlation coefficient can be influenced if multiple raters are involved, and the coefficient is sensitive to differences in association, not in agreement. This means that the Pearson r coefficient would not notify a systematically higher or lower t1 score (Streiner, 2013; Weir, 2005). Therefore, the ICC was used, which would take account of any bias, if present (Streiner, 2013). The ICC also calculates average measures to assess whether the IFTE can be assessed by multiple raters, as the IFTE is a multidisciplinary tool. However as assumptions were violated, Spearman's correlation coefficient was added to complement the results. These results also display a substantial agreement for both the interrater reliability and test-retest analysis.

Although the internal consistency is good for the three factors, the factor analysis itself should be reassessed in a larger group as the results in this study limitedly support the factorial structure. The interrater reliability is moderate to good, and the composite score can be used in treatment evaluation. However, the items *antisocial associates* and *skills to prevent physically aggressive behavior*, scoring lower, should be interpreted with caution and the item

sexually deviant behavior could be considered in accordance with the item *skills to prevent sexually transgressive behavior*.

A composite score is considered to be more reliable than an individual score, meaning that an evaluation with multiple raters is more accurate (Wu, Whiteside, & Neighbors, 2007). A patient's IFTE report indicates whether the raters agree or not on every single item, indicating which items should be discussed by raters to gain consensus, which would give an even better evaluation of the patients' functioning (Schuringa et al., 2014). This has the advantage that all items can be checked for agreement, including the items that showed a lower value in this study. The IFTE, therefore, can give us an accurate measurement of patients' functioning if assessed in the proper way.

Even though the test-retest value over three months is considerably good, this does not mean that the IFTE cannot measure individual change over three months. Individual measurements and the use of the N=1 analysis would display whether change can be measured on several items over a period shorter than six months.

Limitations

Not all composite scores in IFTEs were achieved by more than one rater. Some IFTEs were scored by one psychiatrist/psychologist or staff member/personal coach. Wu, Whiteside and Neighbors (2007) describe that a composite score is more accurate than an individual score, and that a composite score is at its best at a level of five raters. The limited number of raters in this study may possibly have influenced the results. It is important, therefore, to ensure that multiple raters, preferably five, per patient score the IFTE to obtain a more reliable score. The number of participants included in factor analysis is also limited. For a factor analysis, a minimum number of 200 participants is considered best. However, with the optional items, we have not assessed the structure for a total of 218 participants on all items. A larger study with as many completely scored assessments as possible would be preferable to re-assess the factorial structure. Even more, assumptions were violated and the model fit was limited, indicating that results should be interpreted with caution.

As the patient characteristics display, the diversity in diagnoses is high and so are the co-morbidity numbers. In this study, all patients were grouped together. However, it must be considered whether the same change in the same time period can be assessed for this heterogeneous group. Future research should emphasize whether patients differ in treatment progress during treatment. Patient risk profiles could contribute to this evaluation and assess whether a difference in time is required for change, for different items or factors, and whether the same problematic scores arise at the start and during treatment (Van der Veeken, Bogaerts, & Lucieer, 2017).

Several items displayed a non-normal distribution. Even though the analysis included a proportional number of participants, the scoring distribution on the item *sexually deviant behavior* is very low. The same goes for the items *psychotic symptoms, skills to take care of oneself* and *antisocial associates* in the analyses. Especially for these items, the results should be interpreted with caution. The test retest analyses were conducted in social therapeutic environments, including mostly patients with a personality disorder. The items *skills to take care of oneself* and *medication use* are possibly more applicable in supportive therapeutic environments, including patients who are more often diagnosed with a psychotic disorder and patients who need more guidance. Even though the *N* was proportional in most analyses, the analyses to assess change included smaller groups, these results should be reassessed in larger groups to gain more robust results and current results should be interpreted with care.

Implications

This study assesses the psychometric properties of the IFTE to support its clinical use. The IFTE assesses functioning, problematic, and protective factors, and could be used to fill the gap that meta-analyses have pointed out in forensic ROM (Fitzpatrick et al., 2010 Yiend et al., 2011; Alexanders et al., 2010; Goethals & Van Marle, 2012). ROM is performed in the first place to evaluate the patients' treatment, to engage patients in their own treatment, to motivate them by showing the progress they have already made and what is to be gained, and to discuss the behavior they show.

Individual treatment evaluation reports have been developed in an FPC in the Netherlands to display individual progress over time. An individual report shows the agreement between raters, statistical change, clinical change, and the patient's functioning over time, which provides an opportunity to discuss agreements within teams and to assess change, relapse, or stagnation in time with a valid tool and to adjust treatment plans in time. In the FPCs where this study was conducted, patients can also evaluate their own functioning with an IFTE self-report, providing the treatment team with information on the patients' own view in addition to the treatment teams' view. This allows the team to discuss the reports with patients in order to gain team-patient consensus and to improve transparency, which could then help to motivate patients by engaging them in their own treatment evaluation (Carlier et al., 2012). This is especially important in a forensic setting, where patients have been admitted involuntarily and often cope with limited problem insight. The IFTE offers the opportunity to assess and discuss treatment evaluations with forensic psychiatric patients in line with the renowned forensic rehabilitation model: the RNR model. This means that the IFTE is a ROM tool that enhances items that are particularly applicable to the forensic setting and supports the goal of reducing the risk of recidivism. In this way,

treatments designed according to the RNR principles can be evaluated and adjusted if necessary. As it evaluates items related to the criminogenic needs in relation to treatment, the IFTE also supports the responsivity principle: when the expected progress has not been made, one can reevaluate the treatment method in relation to the patients' individual responsivity factors.

Current studies have focused on forensic psychiatric inpatient settings. It would be interesting, however, to assess whether the IFTE can be assessed in outpatient settings or other forensic settings were rehabilitation is an important priority.

Future qualitative and quantitative research would have to assess whether the goal of ROM can be attained with the IFTE and its feedback report in relation to patient motivation, communication, and decision-making in a forensic setting. In order to increase its forensic clinical value, its predictive validity for positive outcomes (i.e., leave) and negative outcomes (i.e., recidivism, aggression) should be assessed. In this way, the IFTE could contribute to decision-making in forensic treatment.

Not all items score above the threshold for the interrater and test-retest reliability analyses. Normally, this would indicate their exclusion but, as mentioned before, the items are of great clinical importance and require assessment. It is important to assess whether the questions need additional explanation. As the lower values could be the effect of possible internal forensic treatment influences, as for the low level of sexually deviant behavior displayed within the institutions, this also means that the items should be interpreted with caution. Future research involving a higher number of patients could reassess the values for the NA items *psychotic symptoms, medication use, drug use, skills to prevent substance use, skills to prevent future physically aggressive behavior*, and *skills to prevent sexually deviant behavior*. Especially the item *skills to prevent future physically aggressive behavior* needs to be reassessed, as it shows both low interrater reliability and test retest values. The item *sexually deviant behavior,* now assessed for all patients, could then also be reassessed for patients who did show sexually deviant behavior in the past.

References

Alexanders, R. T., Green, F. N., O'Mahony, B., Gunaratna, I. J., Gangadharan, S. K., & Hoare, S. (2010). Personality disorders in offenders with intellectual disability: a comparison of clinical, forensic and outcome variables and implications for service provision. *Journal of Intellectual Disability Research, 7,* 650-658. doi. 10.1111/j.1365- 2788 .2010.01248.x

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders – DSM-IV-TR.* Arlington, VA: American Psychiatric Association.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct, fifth edition*. New Providence, NJ: LexisNexis Matthew Bender

Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., & Wormith, S. (2011). Are the major risk/need factors predictive of both female and male reoffending? : A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology, 56,* 113 - 133. doi: 10.1177/0306624X10395716

Atkisson, C., Cook, J., Karno, M., ... *Hohman, A.* (1992) Clinical services research. *Schizophrenia Bulletin*, 18, 561-626.

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *Bmj*, *314*(7080), 572. doi: http://dx.doi.org/10.1136/bmj.314.7080.572

Bland, J. M., & Altman, D. G. (2002). Validating scales and indexes. *Bmj*, *324*(7337), 606-607. doi: http://dx.doi.org/10.1136/bmj.324.7337.606

Bogaerts, S., Spreen, M., Ter Horst, P., & Gerlsma, C. (2017). Predictive validity of the HKT- R Risk Assessment: Two and 5-year violent recidivism in a nationwide sample of Dutch forensic psychiatric patients. *International Journal of offender Therapy and Comparative Criminology,* 1-12. doi: 10.1177/0306624X17717128

Bogaerts, S., Vanheule, S., & DeClercq, F. (2005). Recalled parental bonding, adult attachment style, and personality disorders in child molesters: a comparative study. *The Journal of Forensic Psychiatry & Psychology*, 16, 445-458. doi: 10.1080/ 14789940500 094524

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and treatment* (User Report No. 2007-06). Ottawa, Ontario: Public Safety Canada. Retrieved from: https://cpoc.memberclicks.netassets/Realignment/risk_need_2007-06_e.pdf

Buschman, J., Bogaerts, S., Foulger, S., Wilcox, D., Sosnowski, D., & Cushman, B. (2010). Sexual history disclosure polygraph examinations with cybercrime offences: A first

Dutch explorative study. *International Journal of Offender Therapy and Comparative Criminology*, 54, 395-411. doi: 10.1177/0306624X09334942

Carlier, I. V. E., Meuldijk, D., Van Vliet, I. M., Van Fennema, E. M., Van der Wee, N. J. A., & Zitman, F. G. (2012). Empirische evidence voor de effectiviteit van routine outcome monitoring; Een literatuur onderzoek [Empirical evidence for the effectiveness of routine outcome montoring. A study of the literature]. *Tijdschrift voor Psychiatrie, 2,* 121- 128. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl /issues/450/articles/9326

Coid, J. W., Ullrich, S., Kallis, C., Keers, R., Barker, D., Cowden, F., & Stamps, R. (2013). The relationship between delusions and violence: findings from the East London First Episode Psychosis Study. *JAMA psychiatry*, *70*(5), 465-471. doi:10.1001/jamapsychiatry.2013.12

De Beurs, E., & Zitman, F. G. (2007). Routine outcome monitoring. Het meten van therapie- effect in de klinische praktijk met webbased software. [Routine outcome monitoring. Measuring therapy effects in clinical practic with webbased software]. *Maandblad geestelijke volksgezondheid, 62,* 13-28. Retrieved from: http://www .zorgnetvlaanderen.be/publicatie/Documents/Andere/Routine%20outcome% 20monitoring%20de%20Beurs%20Zitman%20mgv2007-1-13.pdf

De Ruiter, C., & Trestman, R. L. (2007). Prevalence and treatment of personality disorders in dutch forensic mental health services. *The Journal of the American Academy of Psychiatry and Law, 35,* 92-97. Retrieved from: http://www.jaapl.org/content/35/ 1/92.full.pdf+html

De Vogel, V. De., Ruiter, C. De., & Bouman, Y. (2007). Meting van protectieve factoren en het belang ervan voor behandeling in de forensische setting. [Measuring protective factors and the importance within treatment in a forensic setting] *Tijdschrift voor Psychotherapie,* 33, 102-119. doi:10.1007/BF03062262

Dickens, G., Sugarman, P., & Walker, L. (2007). HoNOS-secure: A reliable outcome measure for users of secure and forensic mental health services. *The Journal of Forensic Psychiatry and Psychology,* 4, 507-514. doi: 10.1080/14789940701492279

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). HCR-20V3 Historical, Clinical, Risk Management (Version 3): Professional guidelines for evaluating risk of violence. Vancouver, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170-177. doi: 10.1037/1082-989x.1.2.170

Ellwood, P. M. (1988). Outcomes management. A technology of patient experience. *The New England Journal of Medicine, 23,* 1549 - 1556. doi:10.1056/NEJM198806093182329

Fazel, S., & Yu, R. (2011). Psychotic disorders and repeat offending: Systematic review and meta-analysis. *Schizophrenia Bulletin, 37,* 800-810. doi: 10.1093/schbul/sbp135.

Field, A. (2009). *Discovering statistics using SPSS (and sex and drugs and rock 'n' roll*). Los Angeles, CA: Sage.

Fitzpatrick, R., Chambers, J., Burns, T., Doll, H., Fazel, S., Jenkinson, C., . . . Yiend, J. (2010). A systematic review of outcome measures used in forensic mental health research with consensus panel opinion. *Health Technology assessment*, 18, 1 – 91. doi. 10.3310/hta14180

Goethals, K., & Van Marle, H. J. C. (2012). Routine outcome monitoring in de forensische psychiatrie: Een lang verhaal in het kort. [Routine outcome monitoring in forensic psychiatry: A long story cut short] *Tijdschrift voor Psychiatrie, 2*, 179-183. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/issues/450/articles/9335

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quant Methods Psychol., 8,* 23-34. Retreived from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402032/

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). "Structural equation modeling: guidelines for determining model fit". *The Electronic Journal of Business Research Methods, 6,* 53 - 60, retrieved from: www.ejbrm.com

Kunst, M., Bogaerts, S., & Winkel, F. (2009). Peer and inmate aggression, type D-personality and posttraumatic stress among Dutch prison workers. Stress and Health: Journal of the International Society for the Investigation of Stress, 25(5), 387-395. doi. 10.1002/smi.1247View

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *International Biometric Society, 33,* 159 - 174. doi: 10.2307/2529310

Leung, S-O. (2011). A Comparison of psychometric properties and normality in 4-, 5-, 6-, and 1-point likert scales. *Journal of Social Service Research*, 37, 412-421. doi. 10.1080/01488376.2011.580697

Nederlands Instituut voor Forensische Psychiatrie en Psychologie. (2016). *Jaarbericht 2016.* Den Haag: NIFP

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32,* 396-402. Retrieved from: https://people.ok.ubc.ca/brioconn/nfactors/nfactors.html

Palermo, G. B. (2009). [Editorial]. Delinquency: Risks and protective factors. *Journal of Offender Therapy and Comparative Criminology, 53*, 247-248. doi:10.1177/ 0306624X0 9335886

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104,* 1-15. doi: 10.1016/S0001-6918(99)00050-5.

Rennie, C. E., & Dolan, M. C. (2010). The significance of protective factors in the assessment of risk. *Criminal Behavior and Mental Health, 20*, 8-22. doi:10.1002/cbm.750 Schuringa, E. (2011). Een voorbeeld uit het patiënt volg systeem van het FPC Dr. S. van Mesdag. *GGzet Wetenschappelijk, 2*, 70-77. [An example of the "patient following" system of the FPC Dr. S. Van Mesdag]. Retrieved from: http://static.mesdag .wdmc. eu/niAe9OzHKU -x2JO03A57HQ/TW4xvBca2ENSZYBwBXk9Pg/WG -qCbng-KklSiZk6ShJEGw/Een%20voorbeeldcasus%20uit%20het%20Patient%20Volg %20 Systeem%20-%202011.pdf

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, 2,* 127-144. doi. 10.1080/1522893 2.2014.897536

Spreen, M., Brand, E., Ter Horst, P., & Bogaerts, S. (2014). *Handleiding, Historische Klinische Toekomst-Revisie versie 2014 [Manual, Historical, Clinical, Future, Revised version 2014].* The Netherlands, Groningen: Stichting FPC Dr. S. van Mesdag.

Stinckens, N., Smits, D., Claes, L., & Soenen, S. (2012). Zinvol 'rommen': Zoeken naar een evenwicht tussen gebruiksvriendelijkheid en klinische relevantie. [Meaningful 'romming' (routine outcome monitoring): seeking a balance between userfriendliness and clinical relevance.]. *Tijdschrift voor Psychiatrie, 2,* 161-165. Retrieved from: http:// www.tijdschriftvoorpsychiatrie.nl/assets/articles/TvP12-02_p161-165.pdf

Streiner, D. L. (2013), "A checklist for evaluating the usefulness of rating scales", In Streiner, D. L. (ed.), *A Guide for the Statistically Perplexed: Selected Readings for Clinical Researchers* (pp. 267 - 288). Toronto, Canada: University of Toronto press.

Surjadi, B., Van Horn, J., Bogaerts, S., & Bullens, R. (2010). Internet offending: Sexual and non-sexual functions within a Dutch sample. *Journal of Sexual Aggression*, 16(1), 47- 58. doi: 10.1080/13552600903470054

Taskforce behandelduur tbs. (2014). Meerjarenafspraken forensische zorg 2013-2018. Bevindingen en aanbevelingen. [Long term agreements forensic care 2013-18. Findings and recommendations]. Den Haag; Taskforce behandelduur tbs

Van der Veeken, F. C. A. (2011). *Routine outcome monitoring in the forensic psychiatric centre "The Kijvelanden"* (Master's thesis). Retrieved from http://arno.uvt.nl /show.cgi?fid=122766

Van der Veeken, F. C. A., Bogaerts, S., & Lucieer, J. (2012). Routine outcome monitoring in forensic psychiatry: FPC 2landen and FPC De Kijvelanden. In T. I. Oei & M. S. Groenhuijsen (Eds.), *Progression in forensic psychiatry* (pp. 355-375). Alphen aan den Rijn, The Netherlands: Kluwer.

Van der Veeken, F. C. A., Bogaerts, S., & Lucieer, J. (2017). Patient profiles in Dutch forensic psychiatry, based on risk factors, pathology and offense. *International Journal of Offender Therapy and Comparative Criminology, 61*, 1369-1391. doi: 10.1177 /0306624X15619636

Van Nieuwenhuizen, C. H., Bogaerts, S., Ruijter, de. E. A. W., Bongers, I. L., Coppens, M., & Meijers, R. A. A. C. tbs-behandeling geprofileerd; een gestructureerde casussenanalyse [Profiling TBS-treatment: A structured cases analysis]. Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) Ministerie van Veiligheid en Justitie. 2011. Retrieved from http://www.wodc.nl/onderzoeks database/voor studie-onderzoek-tbs.aspx

Van Nieuwenhuizen, C., Schene, A. H., & Koeter, M. W. J. (2000). Manchester - verkorte kwaliteit van leven meting [The Manchester Short Assessment of Quality of Life]. Eindhoven, Netherlands, Institute of Mental Health Care Eindhoven.

Van Os, J., Burns, T., Cavallaro, R., Leucht, S., Peuskens, J. Heilldin, L., Bernardo, M., Arango, C., Fleischhacker, W., Lachaux, B., & Kane, J.M. (2006). Standardized remission criteria in schizophrenia. *Acta Psychiatrica Scandinavica, 113*, 91-95. doi: 10.1111/j.1600-0447.2005.00688.x

Ward, T., & Gannon, T. A. (2006). Rehabilitation, etiology, and self-regulation: the comprehensive good lives model of treatment for sexual offenders. *Aggression and Violent Behavior, 11,* 77-94. doi:10.1016/l/avb.2005.06.001

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 1*, 231-240.

Wong, S. C. P., Gordon, A., & Gu, D. Q. (2007). Assessment and treatment of violence-prone forensic clients: an integrated approach. *British Journal of Psychiatry*, 190, S66-S74, supplement: 49. doi: 10.1192/bjp.190.5.s66

Wooditch, A., Tang, L. L., & Taxman, F. S. (2014). Which Criminogentic Need Changes are most important in promoting desistance from crime and substance use? *Criminal Justice and Behavior*, 41(3), 276-299. doi: 10.1177/0093854813503543

Wu, S., Whiteside, U., & Neighbors, C. (2007). Differences in inter-rater reliability and accuracy for a treatment adherence scale. *Cognitive Behaviour Therapy*, 4, 230-239. doi. 10.1080/16506070701584367

Yiend, J., Chambers, J. C., Burns, T., Doll, H., Fazel, S., Kaur, ... Fitzpatrick, R. (2011). Outcome measurement in forensic mental health research: An evaluation. *Psychology, Crime and Law*, 3, 277-292. doi. 10.1080/10683160903203938

CHAPTER FIVE

Predictive Validity of the Instrument for Forensic Treatment Evaluation Routine Outcome Monitoring and Clinical Decision-Making in Forensic Psychiatry Based on the Instrument for Forensic Treatment Evaluation

Frida C. A. van der Veeken
Jacques Lucieer
Stefan Bogaerts

This is an adapted version of the article published as: 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, 12.* doi:10.1371/journal.pone.0160787

Abstract

Rehabilitation in forensic psychiatry is achieved gradually with different leave modules, in line with the Risk-Need-Responsivity model. A forensic routine outcome monitoring tool should measure treatment progress based on the rehabilitation theory, and it should be predictive of important treatment outcomes in order to be usable in decision-making. Therefore, this study assesses the predictive validity for both positive (i.e., leave) and negative (i.e., inpatient incidents) treatment outcomes with the Instrument for Forensic Treatment Evaluation (IFTE).

Two-hundred and twenty-four patients were included in this study. ROC analyses were conducted with the IFTE factors and items for three leave modules: guided, unguided and transmural leave for the whole group of patients. Predictive validity of the IFTE for aggression in general, physical aggression specifically, and urine drug screening (UDS) violations was assessed for patients with the main diagnoses in Dutch forensic psychiatry, patients with personality disorders and the most frequently occurring co-morbid disorders; those with combined personality and substance use disorders.

Results tentatively imply that the IFTE has a reasonable to good predictive validity for inpatient aggression and a marginal to reasonable predictive value for leave approvals and UDS violations. The IFTE can be used for information purposes in treatment decision-making, but reports should be interpreted with care and acknowledge patients' personal risk factors, strengths and other information sources.

Introduction

In the Netherlands, offenders who have committed a crime under the influence of a mental illness with a minimum penalty of four years can be admitted to a forensic psychiatric center by order of the state. This order is called *Ter Beschikking Stelling* (TBS). Patients who reside in forensic psychiatric centers are held to be (diminished) non-responsible for their criminal behavior and receive a security measure (Van Nieuwenhuizen et al., 2011). The primary goal of Dutch forensic psychiatric treatment is the prevention of future crimes. This objective can be achieved, step-by-step, through a process involving treatment, rehabilitation and reintegration (Van Nieuwenhuizen et al., 2011). Patients receive care and treatment and follow a structured daily programme, including study, leisure and work. Their re-entry into society takes place gradually, with different leave modalities involving increasing levels of autonomy, on the condition that the treatment cycle proceeds positively (that is, shows a decrease of risk factors and an increase of protective and reintegration factors), and that no inpatient and outpatient violations of rules are committed (Van Nieuwenhuizen et al., 2011).

Leave modalities are necessary milestones in a patient's rehabilitation process (Jeffery & Woolpert, 1974; LeClair & Guarino-Ghezzi, 1991). The Dutch forensic psychiatric system consists of six leave modalities (Rijksoverheid, n.d.) 1; guided leave: patients can leave the institution for a short period of time accompanied by a rehabilitation team, 2; unguided leave: patients are allowed to go outside the clinic without guidance, 3; transmural leave: patients can live outside the institution with other patients under the supervision and responsibility of the institute, 4; probationary leave: the forensic institution is still responsible for patients, and patients are guided by a probation officer outside the clinic, 5; conditional release: patients can live alone or in a group, provided they comply with rules and agreements imposed by the court, such as no alcohol or drugs and mandatory treatment, and 6; patients can achieve unconditional release on the court's decision, which means that rules and agreements are no longer imposed, and the patient is a free man like everyone else (Rijksoverheid, n.d.[1]). Unconditional release is always preceded by conditional release, unless contra-indicated, from May 2013 on, but not necessarily by guided, unguided, transmural or probationary leave. Over the past few decades, rehabilitation has been described in several ways with regard to its goals (Ogloff, 2011). In this study, rehabilitation is a phased process depending on the presence of dynamic risk factors, such as impulsivity or self-control, and the severity of these factors (criminogenic needs) as related to criminal behavior (Andrews & Bonta, 2010). An important theoretical framework of rehabilitation is the well-known Risk-Need-Responsivity (RNR) model. The RNR model is the premier model

for indicating offenders' risk assessment and treatment (Bonta & Andrews, 2007; Andrews & Bonta, 2010; Andrews, Bonta, & Wormith, 2011). Risk assessment instruments are necessary to assess the nature and severity of specific risk factors and, in general, the risk of recidivism. Risk assessment instruments can also be used to specify treatment directions (Van der Veeken, Bogaerts, & Lucieer, 2017) and to determine leave modalities corresponding to a person's level of risk at a particular moment (De Vries & Spreen, 2012).

Over the past two decades, the development of risk assessment instruments has made huge steps forward (Bonta & Andrews, 2007). A first important step was the changing focus of static historical risk factors to reversible dynamic risk and protective factors in several assessment tools, such as the Historical Clinical Risk 20 items (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) and the Historical Clinical Future-30 (HKT-30; Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Task Force in Forensic Psychiatry], 2002). These tools also provide information on criminogenic needs that could be addressed in treatment (Campbell et al., 2009). More recently, fourth-generation risk assessment instruments have been developed, such as the Historical Clinical Risk Management-20, Version 3 (HCR-20^{V3}) (Douglas, Hart, Webster, & Belfrage, 2013) and the Historical Clinical Future-Revised (HKT-R) (Spreen, Brand, Ter Horst, & Bogaerts, 2014). Fourth-generation instruments can be integrated into risk management, aid the selection of treatments and interventions and help to assess the rehabilitation process (Campbell, French, & Gendreau, 2009). Both the revised HKT-R and HCR-20^{V3} have the objectives of assessing risk of recidivism, use in treatment and assessment of treatment goals, which refers to Dutch forensic psychiatry policy.

In the Netherlands, as imposed by the Ministry of Justice and Security (MJS), each inpatient forensic psychiatric center is obliged to establish an annual measure of future risk for patients who have committed a violent and/or sexual offense. The MJS has made mandatory annual assessment by two risk assessment tools (DJI, 2014): the HCR-20^{V3} (Douglas et al., 2013) and the HKT-R (Spreen et al., 2014). To measure changes in risk behavior during inpatient treatment over time (yearly), institutions may only use the 14 Clinical items of the HKT-R (Spreen et al., 2014). While routine treatment evaluations are beneficiary for treatment outcome and provide important treatment information (Lambert et al., 2003; De Beurs et al., 2011), it is doubtful and barely studied whether both instruments are also suitable in the context of routine treatment evaluations (Goethals & Van Marle, 2012).

We must raise the question, therefore, whether a risk assessment tool meant to value future violent behavior can also be used to assess treatment progress routinely at the same time. Secondly, the limited response categories of the HCR-20^{V3} (3-point scale) and the HKT-R (five-point scale) can be problematic in observing short-term changes in behavior.

Schuringa, Spreen and Bogaerts (2014), for example, showed that limited anchor points are not always accurate representations of a patient's behavior because a patient's observed behavior may fall between two anchor points (see next paragraph). This problem is very often the case when people must choose from a limited number of options (Gunderman & Chan, 2013).

Monitoring treatment and assessing inpatient behavior

Monitoring treatment progress involves an integrated approach from the start until the end of treatment (De Beurs et al., 2011). Forensic treatment monitoring aims to understand the decrease, stagnation or increase of the severity of crime-related risk factors and personal, psychological and social factors, in line with the theoretical considerations of fourth-generation risk assessment instruments. The measurement of inpatient risk factors, such as impulsivity, hostility, treatment and coping skills requires validated measurements that are sufficiently specific and sensitive to measure changes over time. Such measurements should have satisfactory/good predictive power for clinical practitioners to gain insight into the likelihood of future rule violation and violent behavior and to aid to decision making. Before one of the above-mentioned leave modalities can be assigned to a patient, for example, behavioral factors such as problematic behavior, protective and resocialization skills must be monitored periodically. These behavioral factors must be predictive of relevant inpatient outcome measures, such as rule violation and aggressive inpatient behavior, as problematic inpatient behavior is a strong predictor of problematic outpatient behavior. Spreen et al. (2014), for example, found predictive values for historical risk factors and clinical risk items assessed over a period of twelve months with the HKT-R in a nationwide representative cohort of 347 forensic psychiatric patients. Routine Outcome Monitoring (ROM) is necessary to make clinically based decisions at the start, during and at the end of treatment (Ellwood, 1988). It is important to evaluate individual treatment, psychological and social functioning, rule violation and aggressive behavior throughout the whole treatment process. This must be done for various purposes, such as the adjustment or continuation of current treatments or the granting of leave modalities. Despite the positive impact of ROM (Carlier et al., 2010) in general psychiatry and the use of ROM in decision making (Larson, 2013), we note that ROM in Dutch forensic psychiatry has only recently been introduced and that only a handful of empirical studies have been conducted in this field (Van den Brink, Hooijschuur, Van Os, Savenije, & Wiersma, 2010; Goethals & Van Marle, 2012; Schuringa et al., 2014).

In consultation with Dutch clinicians (psychologists, psychiatrists and social workers), Schuringa et al. (2014) have recently developed the Instrument for Forensic Treatment

Evaluation (IFTE) to investigate changes in inpatient behavior. This instrument provides solutions for the aforementioned limitations of the HCR-20^{v3} and HKT-R. Schuringa et al. (2014) opted for a 17-point scale to measure forensic psychiatric behavior over time. This 17-point scale contains five anchor points and gives professionals the ability to score between anchor points. As mentioned earlier, the IFTE consists of 14 dynamic items that have been derived from the HKT-R (Schuringa et al., 2014), such as *impulsivity* and *problem insight.* Three items were derived from the Dutch version of the Atascadero Skills Profile (ASP; Vess, 2001), a behavioral observation instrument, namely *skills to prevent substance use, skills to prevent physically aggressive behavior* and *skills to prevent sexually deviant behavior* (Schuringa et al., 2014). Finally, the items *manipulative behaviors, balanced daytime activities, financial skills, sexually deviant behavior* and *medication use* were added as these were valued as 'very useful for treatment evaluation' by clinicians (Schuringa et al., 2014). The 22 items can be clustered into three factors, namely *problem behavior* (*impulsivity*, *drug use* and *hostility*), p*rotective behavior* (*problem insight* and *coping skills)* and *resocialization skills* (*balanced daytime activities* and *social skills)*.

The psychometric qualities of the IFTE were examined in 232 forensic psychiatric patients. Factor analysis confirmed the factor structure with very good internal consistency for the three factors (problem behavior, α = .86, protective behavior, α = .90 and resocialization skills, α = .88). Test-retest reliability for the three factors was very good (problem behavior, α = .85, protective behavior, α = .87, and resocialization skills, α = .89) (Schuringa et al., 2014). The IFTE was evaluated to be a reliable ROM instrument for supporting and indicating inpatient forensic psychiatric treatment (Schuringa et al., 2014).

As mentioned, a tool to measure treatment evaluation should consist of relevant factors that correlate with significant outcome indicators, such as leave modalities and violent inpatient behavior. This means that a treatment evaluation instrument should have sufficient predictive power to predict important future treatment factors. Schuringa, Heininga, Spreen, & Bogaerts (2018), find discriminatory values for inpatient violence and drug use of the three IFTE factors; Problem behavior d = -1.07 for violence, d = 1.47 for drug use; Protective behavior d = .57 for violence, d = -.57 for drug use; Resocialization skills d = .40 for violence.

In this study, we re-examine whether the IFTE can be used to support decision-making in forensic treatment. The goal of this study is to assess the predictive validity for positive treatment outcomes (leave) as well as negative treatment outcomes (inpatient incidents). As it is considered a step forward in treatment, leave is considered a positive treatment outcome (Bernstein, 2012). A granted leave request is re-evaluated every year and, hence, reconsidered every year. We will examine the predictive validity of the IFTE for leave mo-

dalities granted to patients for the first time, for guided, unguided and transmural leave modules for the whole group of patients.

We do know, however, that different diagnoses or combinations of co-morbid diagnoses are related to different risk factors (Van der Veeken et al., 2017). While diagnoses are diverse, most patients in Dutch forensic psychiatry are diagnosed with a psychotic disorder or substance use disorder (SUD) on axis I and a cluster B personality disorder or personality disorder not otherwise specified (NOS) on axis II (Van Nieuwenhuizen et al., 2011). A cluster B personality disorder and SUD co-occur most often (Van Nieuwenhuizen et al., 2011). Patients with a personality disorder may show multiple risk factors, possibly differing when a co-morbid SUD is present. Therefore, the predictive validity for inpatient incidents was studied for the main diagnostic group, that is, patients diagnosed with one of the main diagnoses recognized in forensic psychiatry: a psychotic disorder, a SUD or a personality disorder (NOS). In addition, we examined predictive validity for patients with a main personality disorder (PSD) and for patients with a personality disorder with a co-morbid SUD (PSDS).

Inpatient incidents are defined as inpatient aggression, namely verbal, material, and physical aggression. These three forms of aggression are included in this study. When verbal aggression occurs, staff will intervene in order to prevent any escalation. As patients are guided throughout the day and staff are well prepared for possible incidents, physical aggression might possibly occur less within the institution than in an uncontrolled setting (outside). However, physical aggression is considered to be more severe, and the predictive value for physical aggression, therefore, will also be specifically assessed in this study.

In addition, though not considered a form of aggression, the violation of urine drug screenings (UDS) is classed as an inpatient incident as use of drugs or alcohol is considered a serious violation of rules. An unreliable or refused UDS limits the FPCs ability to ensure internal safety, and therefore, patients will then receive supplementary guidance. We hypothesize that higher problem behavior scores, indicating more problem behavior, are predictive of inpatient incidents, general and physical aggression, and the violation of UDS procedures. Low problem behavior scores are hypothesized to be predictive of all three leave modalities. Higher levels of resocialization skills and protective behavior, indicating developed skills and protective behavior, are hypothesized to be predictive of all three leave modalities, whereas lower levels of resocialization skills and protective skills are hypothesized to be predictive of inpatient incidents. Predictive values will be assessed at item and factor level.

Materials and Methods

Participants

Two-hundred and twenty-four male patients were included in this study. All patients resided in two Dutch forensic psychiatric centers (FPCs). For all patients, the court imposed detention under a hospital order (TBS order). All committed a crime that was related to their mental health status with a minimum penalty of four years, and all received intramural treatment. Participants' mean age at the time of their first ROM assessment was 40 years (*SD* = 10.15, range = 22 – 72). Table 1 shows their primary diagnosis on Axis I or Axis II of the DSM-IV-TR (30) (American Psychiatric Association (APA), 2000) as assessed by clinicians, type of offense and ethnicity. As shown in Table 1, 122 patients were primarily diagnosed with a personality disorder. Ninety-one patients were diagnosed with a personality disorder in combination with an SUD. All gathered information is primary treatment information and was retrieved from individual patient files whose information was anonymized prior to the analysis and not traceable to an individual. Data was analysed in line with the standards of the APA guidelines and Helsinki declaration. Informed consent was not required while all data was primary treatment information and part of clinical routine outcome monitoring. This study is part of a ROM study in forensic psychiatry and has been approved by the scientific research committee of FPC De Kijvelanden.

Table 1

Patient characteristics, primary diagnoses, index crime and ethnicity

		N	Percent
Primary diagnosis			
	Schizophrenia	41	18.3
	Delusional	4	1.7
	Other psychotic disorder	8	3.5
	Pervasive developmental disorder	11	4.8
	Paedophilia	12	5.3
	Substance use disorder	15	6.8
	Cluster B PSD*		
	Antisocial	49	21.8
	Borderline	11	4.9
	Narcissism	9	4.0
	Personality disorder not otherwise specified	53	23.6
	Other	11	5.2
Index crime	Property offenses with violence	22	9.8
	Maltreatment	35	15.6
	Homicide	83	37.1
	Arson	14	6.3
	Sexual offenses	35	15.6
	Child sexual abuse	32	14.3
	Other	3	1.3
Ethnicity	Dutch	145	64.7
	Turkish	9	4.0
	Moroccan	12	5.4
	Antillean	11	4.9
	Surinam	20	8.9
	Other	26	11.3
	Unknown	1	.4

*PSD=personality disorder

Procedure

The IFTE is part of the ROM procedure in two Dutch forensic psychiatric centers, for part of the patient group with an intelligence quotient above 80 since September 2011, and for the whole group of patients with an intelligent quotient above 80 since mid-2012. ROM was implemented for all patients, irrespective of the period of treatment they had already received. The IFTE is scored approximately every four to six months by one to four therapists: a coach (i.e., a staff member who guides the patient), a psychologist or psychiatrist, a second coach and an art or psycho-motor therapist or a labour consultant who have worked with the patient. Scoring takes place just before the routine patient meetings in which treatment

and progress are discussed. The goal of these patient meetings is to evaluate treatment, to assess individual behavior changes and to evaluate a patient's functioning and previously set treatment goals. All ROM questionnaires were scored in an Excel document appointed to an individual patient. IFTEs were copied in these Excel documents with shortened instructions wherein therapists could assess the IFTE. Individual treatment reports are all constructed in these Excel documents.

Though the IFTE is assessed by several therapists, the date of the last conducted assessment or production of the IFTE report was selected as the date of assessment as this is the point of the IFTEs' clinical use. The aim of this study is to assess the suitability of the IFTE in clinical treatment. Therefore, leave approvals and incidents were collected in between two routinely IFTE assessments. Thus, the predictive validity of the clinically used IFTE assessments in the period subsequent to the clinical assessment could be studied. Dates of internal approval of leave requests and of leaves granted by the MJS were collected from the patients' electronic patient file (EPF) from September 2011 to July 2014. Inpatient incidents, including positive UDS, were collected in the same period, from the EPF. Additionally, incidents reported in the safe incident reporting programme ([veilig incidenten melden] VIM) could be collected from the 2012 and 2013 reports, while the reports of these two years were available to the researchers. VIM is a programme for therapists to report verbal, material and physical aggression as well as any other incident that might have jeopardized internal security. However within this research we have only selected aggressive and UDS incidents. Inpatient incidents, gathered from the EPF and available VIM information, leave requests and post-IFTE assessment approvals were selected, and we studied the short-term predictive validity of the IFTE for the selected outcomes.

Table 2

The instrument for forensic treatment evaluation and internal consistency

Protective behavior Alpha = .84	Problem behavior Alpha = .80	Resocialization skills Alpha = .86
Problem insight	Impulsivity	Balanced daytime activities
Cooperation with treatment	Antisocial behavior	Labor skills
Responsibility for the crime	Hostility	Social skills
Coping skills	Sexually transgressive behavior	Skills to take care of oneself
Medication use	Manipulative behavior	Financial skills
Skills to prevent substance use	Compliance with rules	
Skills to prevent physically aggressive behavior	Antisocial associates	
Skills to prevent sexually deviant behavior	Psychotic symptoms	
	Drug use	

Measurements

Instruments

The IFTE has been designed to assess patients' problem behavior, resocialization skills and protective behavior on a routine basis (Schuringa et al., 2014). Table 2 shows the IFTE items on factor level, together with their internal consistency. The IFTE contains 22 dynamic items assessing three factors: *problem behavior* (*impulsivity, manipulative behavior, drug* use); *protective behavior* (*responsibility for the crime , problem insight*); and *resocialization skills* (*balanced daytime activities, social skills*) (Table 2). The IFTE is assessed in a multidisciplinary fashion, that is, by different disciplines, producing a composite score on a 17-point scale with five anchor points and in-between options. A score of zero indicates that a patient did not show the behavior or skill indicated in the item, and a score of seventeen indicates that a patient frequently displayed the behavior or skill. For every item, therapists can choose to tick the box 'not enough information (N.E.I.)' when they do not have enough information to score the item. For some items, they can tick 'non applicable (N.A.)' when an item does not apply to a patient (Schuringa et al., 2014). This may lead to unevenly scored items in the analyses.

Outcome variables

Leave modalities must be approved by the MJS. All FPCs in the Netherlands must request permission for a patient's leave module and its extension. Before leave can be approved by the MJS, an FPC internal committee must approve leave requests. Dates of first MJS and FPC approvals following IFTE assessment were selected. Unapproved leave requests or withdrawn leave approvals were considered as leave request not granted. Most patients started with guided leave, followed by unguided leave and transmural leave; these three leave modules are considered in the analyses. Predictive validity was assessed for all patients for whom a leave module was granted for the first time, for guided, unguided, and transmural leave modules.

If present, the first reported incident, gathered from the EPF and available VIM information, after an IFTE assessment was taken as outcome measure. Incidents were divided into general aggression (including threats, verbal aggression, material aggression and physical aggression), specific physical aggression and serious violation of UDS (refusal of UDS, unreliable UDS, positive UDS or confession of drug use). While patients differ in diagnoses, and diagnostic combinations can be related to different risk factors, the predictive value for inpatient incidents was first studied for the main diagnostic group in both FPCs, with inclusion of primary personality disorders, psychotic disorders and SUDs, and exclusion of patients with mainly a pervasive developmental disorder, paedophilia or other.

After that, patients with a personality disorder as main diagnosis were selected to study the predictive value of incidents in this specific group. Predictive validity for inpatient incidents was also studied for patients with a personality disorder and co-morbid SUD. All diagnoses were derived from the EPFs and were assessed by clinicians according to the diagnostic and statistical manual of mental disorders fourth edition (DSM-IV-TR; APA, 2000).

Statistics

Data were loaded into the Statistical Package for the Social Sciences (SPSS) 19. The composite scores of multiple raters (one to four raters; assessed by at least one coach/staff member and/or psychologist/psychiatrist) were used in the analyses. The receiver operating characteristics (ROC) analysis gives the area under the curve (AUC). The AUC value is a measure for predictive values. A value of .50 means the predictive value is equal to coincidence; a value of one would represent a perfect predictive value (Spreen et al., 2014). AUCs of .60 are considered to be marginal; AUCs in the range of .70 - .80 are considered to be reasonable; AUCs in the range of .80 - .90 and above are considered good; and an AUC of .90 or higher is considered high (Tengström et al., 2006; Spreen et al., 2014). Ninety-five percent confidence intervals were selected; confidence intervals should remain above .50 in order to predict above chance (Tengström et al., 2006). For the total group of patients, ROC analyses were conducted for guided leave approvals, unguided leave approvals and transmural leave approvals. IFTE scores indicating no problematic behavior and developed skills were calculated to be predictive of the outcomes. The ROC analyses for guided leave approvals included patients with a first guided leave approval (yes = 1) and patients with no leave approval (no = 0); patients who had already had a guided, unguided or transmural leave approval were excluded. The analysis for unguided leave approvals included patients with a first unguided leave approval (yes = 1) and patients with no unguided leave approvals or who had already had a guided leave approval (no = 0); patients who had already received an unguided or transmural leave approval were excluded. The analysis for transmural leave approvals included patients with a first transmural leave approval (yes = 1) and patients with no transmural leave approval, a guided or unguided leave approval (no = 0); patients who had already had a transmural leave approval were excluded.

ROC analyses were then conducted for the incidents of general aggression, physical aggression and serious violation of UDS for the three groups: patients with main diagnostic disorders, the PSD group and the PSDS group. Incidents were coded into 1 = yes and 0 = no. IFTE scores indicating problematic behavior or deviant skills were calculated to be predictive of incidents.

Results

For the whole group of patients, 867 IFTEs were assessed between September 2011 and June 2014. AUCs are displayed in Tables 3 through 6 on item- and factor level for leave requests and incidents of the main diagnostic group, the PSD group and the PSDS group.

Leave approval

Table 3 shows the AUCs for guided, unguided and transmural leave approvals. For the patients who had not received guided leave approval, the mean *protective behavior* scores (t (211) = -2.7, p = .01) and mean *resocialization skills* scores (t (63.42) = -5.09, p = .00) on the IFTE were significantly lower (*MProtective behavior* = 43.60, *SD* = 15.19, *N* = 183; *MResocialization skills* = 49.88, *SD* = 17.91, *N* = 202) than those of patients who had received guided leave approval (*MProtective behavior* = 51.56, *SD* = 13.27, *N* = 30; *MResocialization skills* = 61.59, *SD* = 11.06, *N* = 33). Problem behavior scores did not differ significantly (t(284) = 1.36, p = .18).

Mean factor scores differed significantly for patients who had and patients who had not received unguided leave approval on *protective behavior* (t(428) = -3.13 , p = .00), *problem behavior* (t (45.11) = 4.07, p = .00) and *resocialization skills* (t(40.02) = -5.50, p = .00). Mean factor scores for patients who had not received unguided leave approval was *MProtective behavior* = 48.41 (*SD* = 15.15, *N* = 407), *MProblem behavior* = 41.83 (*SD* = 16.32, *N* = 535) and *MResocialization skills* = 54.05 (*SD* = 17.41 *N* = 439). Mean factor scores for the patient group who had received unguided leave approval were *MProtective behavior* = 58.47 (*SD* = 11.87, *N* = 23), *Mproblem behavior* = 33.95 (*SD* = 10.68, *N* = 35) and *MResocialization skills* = 65.32 (*SD* = 10.08, *N* = 29).

Table 3
Granted leave requests

Item	guided leave	Pos-neg***	unguided leave	Pos-neg***	transmural leave	Pos-neg ***
Problem insight	.611 (.527 - .694)*	43 - 277	.658 (.563 - .753)**	36 - 575	.584 (.478 - .691)	35 - 686
Cooperation with treatment	**.713 (.634 - .792)****	43 - 280	**.703 (.631 - .775)****	36 - 579	.665 (.575 - .755)**	35 - 690
Responsibility for the crime	.588 (.504 - .672)	42 - 270	.606 (.514 - .698)*	36 - 561	.616 (.523 - .710)*	34 - 672
Coping skills	.602 (.526 - .678)*	43 - 278	.693 (.613 - .773)**	36 - 577	.661 (.572 - .750)**	35 - 688
Balanced daytime activities	.633 (.555 - .712)**	43 - 280	.688 (.620 - .757)**	36 - 578	.605 (.507 - .703)*	35 - 688
Labor skills	.634 (.547 - .721)*	35 - 227	**.715 (.637 - .792)****	29 - 482	.607 (.511 - .702)*	31 - 580
Social skills	.549 (.462 - .636)	43 – 281	.624 (.534 - .714)*	36 - 579	.639 (.541 - .736)**	35 - 690
Skills to take care of oneself	.543 (.458 - .628)	43 - 280	.595 (.505 - .686)	36 - 578	.566 (.458 - .674)	35 - 689
Financial skills	.582 (.496 - .668)	40 - 242	.640 (.558 - .723)**	35 - 514	.434 (.333 - .535)	32 - 619
Impulsivity	.519 (.434 - .603)	43 - 280	.594 (.513 - .675)	36 - 577	.567 (.474 - .660)	34 - 689
Antisocial behavior	.576 (.490 - .661)	43 - 281	.612 (.532 - .692)*	36 - 577	.665 (.581 - .749)**	34 - 688
Hostility	.585 (.501 - .670)	43 - 280	.574 (.490 - .659)	36 - 576	.652 (.569 - .735)**	34 - 688
Sexually deviant behavior	.484 (.396 - .573)	43 - 280	.549 (.449 - .649)	36 - 577	.588 (.490 - .686)	34 - 688
Manipulative behavior	.516 (.430 - .602)	43 - 276	.558 (.474 - .641)	36 - 571	.614 (.526 - .702)*	34 - 681
Compliance with rules	.674 (.596 - .752)**	43 - 280	**.723 (.647 - .798)****	36 - 576	.643 (.559 - .727)**	34 - 688
Antisocial associates	.467 (.381 - .554)	42 - 251	.566 (.467 - .666)	35 - 546	.577 (.490 - .663)	34 - 656
Medication use	.563 (.461 - .664)	31 - 195	.633 (.541 - .725)*	23 - 424	.596 (.487 - .705)	25 - 491
Psychotic symptoms	.477 (.369 - .586)	27 - 189	.551 (.452 - .650)	25 - 408	.484 (.366 - .601)	25 - 477
Skills to prevent substance use	.650 (.560 - .740)*	28 - 201	**.702 (.619 - .785)****	23 - 419	.672 (.567 - .777)**	27 - 483
Drug use	.519 (.424 - .614)	35 - 238	.627 (.536 - .718)*	26 - 482	.576 (.481 - .672)	30 - 553
Skills to prevent physically aggressive behavior	.536 (.456 - .617)	36 - 227	.645 (.555 - .735)**	29 - 458	.671 (.594 - .749)**	27 - 548
Skills to prevent sexually deviant behavior	.641 (.518 - .765)	17 - 114	.630 (.487 - .774)	18 - 238	.689 (.581 - .797)	9 - 292
Protective behavior	.652 (.548 - .756)**	30 - 183	.686 (.587 - .785)**	23 - 407	.633 (.520 - .747)*	24 - 474
Problem behavior	.564 (.477 - .651)	42 - 244	.645 (.573 - .717)**	35 - 535	.667 (.585 - .749)**	34 - 641
Resocialization skills	.694 (.609 - .779)**	33 - 202	.691 (.606 - .776)**	29 - 439	.592 (.477 - .707)	28 - 535

*p<.05, ** p <.01, *** positive-negative outcomes

The patient group who had not received transmural leave approval also differed significantly from patients who had received transmural leave approval, on *protective behavior* ($t(496) = -2.20$, $p = .03$) and *problem behavior*, $t\ (39.06) = 3.91$, $p = .00$). Mean factor scores for patients who had not received transmural leave approval were *MProtective behavior* = 49.99 (*SD* = 15.01, *N* = 474) and *MProblem behavior* = 40.87 (*SD* = 15.86, *N* = 641). Mean factor scores for the patient group who had received transmural leave approval were *MProtective behavior* = 56.90 (*SD* = 14.70, *N* = 24) and *MProblem behavior* = 32.27 (*SD* = 12.30, *N* = 34).

General and physical aggression Main group

Table 4 displays the AUC values for the main diagnostic group, including 189 patients. Forty incidents of physical aggression were reported approximately 10.52 weeks after assessment (*SD* = 11.01, range = 0 - 54). *Problem behavior* and *resocialization skills* were most predictive of general and specific physical aggression. One-hundred and fifty-eight general aggressive incidents were reported approximately 10.51weeks after assessment (*SD* = 9.87, range = 0 – 54). Two-hundred and twenty-six UDS violations were reported approximately 8.96 weeks after assessment (*SD* = 10.32, range = 0 - 58).

Personality-disordered group

Table 5 displays AUC values for the PSD group, including 122 patients. Twenty-nine physical aggression incidents were reported approximately 10.97 weeks after assessment (*SD* = 11.39, range = 0 - 54). One-hundred and three general aggression incidents were reported approximately 9.95 weeks after assessments (*SD* = 9.87, range = 0 - 54), and 164 UDS violations were reported approximately 9.76 weeks after assessments (*SD* = 10.93, range = 0 – 58).

Table 4

AUCs for the main diagnostic group

Main diagnostic group	General aggression	Pos-neg***	Physical aggression	Pos-neg***	Urine drug screening violation	Pos-neg ***
Problem insight	.641 (.593 - .689)**	157 - 567	.588 (.510 - .666)	40 - 684	.570 (.527 - .613)**	224 - 500
Cooperation with treatment	.695 (.651 - .738)**	158 - 570	.671 (.599 - .742)**	40 - 688	.660 (.619 - .702)**	226 - 502
Responsibility for the crime	.614 (.564 - .664)**	150 - 550	.540 (.455 - .626)	39 - 661	.538 (.493 - .583)	218 - 482
Coping skills	**.714 (.669 - .759)****	157 - 567	**.764 (.699 - .829)****	40 - 684	.631 (.588 - .673)**	224 - 500
Balanced daytime activities	**.706 (.661 - .751)****	156 - 569	**.714 (.649 - .779)****	39 - 686	.673 (.631 - .716)**	225 - 500
Labor skills	**.710 (.661 - .759)****	131 - 497	**.776 (.714 - .838)****	35 - 593	.666 (.619 - .712)**	192 - 436
Social skills	**.715 (.668 - .762)****	157 - 569	**.723 (.649 - .797)****	40 - 686	.594 (.550 - .637)**	226 - 500
Skills to take care of oneself	.656 (.606 - .706)**	157 - 570	.583 (.496 - .670)	39 - 688	.555 (.510 - .601)*	226 - 501
Financial skills	.672 (.618 - .726)**	140 - 517	.661 (.563 - .759)**	35 - 622	.600 (.554 - .646)**	215 - 442
Impulsivity	**.726 (.681 - .771)****	156 - 568	**.800 (.739 - .860)****	40 - 684	.619 (.574 - .663)**	225 - 499
Antisocial behavior	**.745 (.702 - .788)****	156 - 564	**.797 (.744 - .850)****	40 - 680	.657 (.615 - .700)**	225 - 495
Hostility	**.725 (.680 - .771)****	157 - 567	**.749 (.679 - .820)****	40 - 684	.620 (.576 - .664)**	226 - 498
Sexually deviant behavior	.614 (.562 - .666)**	157 - 566	.590 (.499 - .682)	40 - 683	.586 (.541 - .631)**	225 - 498
Manipulative behavior	.640 (.590 - .690)**	155 - 563	**.717 (.638 - .796)****	39 - 679	.627 (.584 - .671)**	225 - 493
Compliance with rules	**.726 (.682 - .770)****	158 - 566	**.715 (.641 - .789)****	40 - 684	**.709 (.668 - .751)****	226 - 498
Antisocial associates	.614 (.560 - .667)**	147 - 542	.609 (.509 - .710)*	37 - 652	.636 (.591 - .681)**	215 - 474
Medication use	.659 (.604 - .715)**	118 - 374	.575 (.479 - .672)	30 - 462	.542 (.489 - .594)	163 - 329
Psychotic symptoms	.593 (.533 - .654)**	122 - 390	.558 (.452 - .664)	31 - 481	.461 (.408 - .514)	157 - 355
Skills to prevent substance use	**.701 (.646 - .755)****	122 - 420	**.697 (.610 - .784)****	33 - 509	**.740 (.696 - .783)****	194 - 348
Drug use	.642 (.587 - .696)**	135 - 470	.638 (.552 - .724)**	37 - 568	**.797 (.756 - .837)****	209 - 396
Skills to prevent physically aggressive behavior	**.717 (.668 - .766)****	143 - 453	**.706 (.622 - .789)****	37 - 559	.593 (.544 - .641)**	198 - 398
Skills to prevent sexually deviant behavior	**.721 (.648 - .794)****	48 - 196	**.724 (.586 - .861)***	10 - 234	.632 (.548 - .717)**	58 - 186
Protective behavior	.686 (.631 - .741)**	115 - 357	.650 (.557 - .742)**	30 - 442	.599 (.547 - .650)**	156 - 316
Problem behavior	**.769 (.723 - .814)****	143 - 529	**.807 (.740 - .874)****	36 - 636	.682 (.639 - .725)**	213 - 459
Resocialization skills	**.748 (.701 - .795)****	117 - 464	**.759 (.693 - .825)****	32 - 549	.672 (.626 - .719)**	186 - 395

* p <.05, ** p <.01, *** positive-negative outcomes

Table 5

AUCs for the personality disordered group

Personality disorders	General aggression	Pos-neg***	Physical aggression	Pos-neg***	Urine drug screening violation	Pos-neg ***
Problem insight	.611 (.551 - .670)**	102 - 412	.557 (.469 - .646)	29 - 485	.579 (.526 - .632)**	163 - 351
Cooperation with treatment	.680 (.624 - .736)**	103 - 412	.665 (.577 - .753)**	29 - 486	.690 (.641 - .739)**	164 - 351
Responsibility for the crime	.596 (.533 - .659)**	98 - 399	.526 (.428 - .624)	28 - 469	.561 (.506 - .615)*	160 - 337
Coping skills	**.705 (.650 - .760)****	103 - 410	**.759 (.679 - .839)****	29- 484	.650 (.599 - .700)**	163 - 350
Balanced daytime activities	.697 (.642 - .753)**	102 - 412	**.730 (.654 - .806)****	29 - 485	**.708 (.659 - .758)****	163 - 351
Labor skills	**.715 (.656 - .773)****	91 - 365	**.797 (.723 - .872)****	25 - 431	**.720 (.670 - .771)****	143 - 313
Social skills	**.713 (.657 - .769)****	103 - 412	.685 (.596 - .773)**	29 - 486	.640 (.589 - .691)**	164 - 351
Skills to take care of oneself	.632 (.569 - .696)**	102 - 413	.593 (.492 - .694)	28 - 487	.557 (.502 - .611)*	164 - 351
Financial skills	.635 (.567 - .704)**	90 - 377	.628 (.510 - .747)*	25 - 442	.640 (.587 - .694)**	156 - 311
Impulsivity	**.728 (.674 - .783)****	101 - 413	**.807 (.744 - .870)****	29 - 485	.650 (.598 - .702)**	163 - 351
Antisocial behavior	**.746 (.695 - .797)****	102 - 411	**.751 (.680 - .822)****	29 - 484	.692 (.643 - .741)**	164 - 349
Hostility	**.735 (.681 - .790)****	102 - 413	**.736 (.651 - .820)****	29 - 486	.645 (.594 - .697)**	164 - 351
Sexually deviant behavior	.600 (.537 - .663)**	102 - 412	.562 (.456 - .669)	29 - 485	.603 (.549 - .656)**	163 - 351
Manipulative behavior	.660 (.602 - .718)**	100 - 409	.694 (.606 - .782)**	28 - 481	.626 (.575 - .677)**	163 - 346
Compliance with rules	**.728 (.675 - .780)****	103 - 411	**.716 (.623 - .809)****	29 - 485	**.746 (.699 - .793)****	164 - 350
Antisocial associates	.650 (.589 - .711)**	98 - 392	.618 (.501 - .735)*	28 - 462	.617 (.564 - .670)**	156 - 334
Medication use	.638 (.567 - .710)**	69 - 240	.578 (.458 - .699)	20 - 289	.565 (.498 - .631)	105 - 204
Psychotic symptoms	.554 (.477 - .631)	75 - 249	.561 (.428 - .695)	21 - 303	.488 (.420 - .555)	102 - 222
Skills to prevent substance use	**.725 (.662 - .787)****	81 - 295	**.722 (.626 - .818)****	24 - 352	**.755 (.704 - .806)****	141 - 235
Drug use	.669 (.603 - .734)**	88 - 324	.662 (.567 - .758)**	27 - 385	**.796 (.749 - .844)****	149 - 263
Skills to prevent physically aggressive behavior	**.729 (.672 - .786)****	94 - 336	**.719 (.625 - .814)****	27 - 403	.634 (.579 - .690)**	145 - 285
Skills to prevent sexually deviant behavior	**.714 (.629- .799)****	29 - 152	.650 (.470 - .831)	6 - 175	.662 (.563 - .760)**	43 - 138
Protective behavior	.677 (.608 - .747)**	69 - 232	.640 (.530 - .750)*	20 - 281	.641 (.576 - .707)**	103 - 198
Problem behavior	**.776 (.722 - .829)****	95 - 383	**.790 (.710 - .870)****	27 - 451	**.703 (.654 - .752)****	155 - 323
Resocialization	**.756 (.700 - .812)****	82 - 342	**.769 (.689 - .848)****	23 - 401	**.723 (.672 - .775)****	137 - 287

* p <.05, ** p <.01, *** positive-negative outcomes

Personality disordered group with co-morbid substance use disorders

Table 6 shows AUC values for the PSDS group. For the PSDS group, including 91 patients, 22 physical aggression incidents were reported approximately 8.32 weeks after assessment (*SD* = 7.57, range = 0 - 26), and 70 general aggression incidents approximately 8.26 weeks after assessment (*SD* = 8.95, range = 0 – 37). One hundred and thirty-six UDS violations were reported approximately 9.88 weeks after assessment (*SD* = 10.12, range = 0 – 58).

Table 6

AUCs for the co-morbid personality and substance use disorder group

PSDS****	General aggression	Pos-neg***	Physical aggression	Pos-neg***	Urine drug screening violation	Pos-neg ***
Problem insight	.645 (.574 - .715)**	70 - 308	.622 (.526 - .718)	22 - 356	.598 (.539 - .657)**	136 - 242
Cooperation with treatment	**.712 (.650 - .773)****	70 - 308	**.723 (.639 - .806)****	22 - 356	.682 (.626 - .737)**	136 - 242
Responsibility for the crime	.630 (.557 - .702)**	66 - 299	.573 (.479 - .666)	21 - 344	.549 (.487 - .611)	133 - 232
Coping skills	**.751 (.685 - .817)****	70 - 308	**.828 (.754 - .903)****	22 - 356	.670 (.614 - .727)**	135 - 243
Balanced daytime activities	**.735 (.670 - .799)****	70 - 308	**.776 (.695 - .856)****	22 - 356	.692 (.635 - .750)**	136 - 242
Labor skills	**.739 (.670 - .808)****	63 - 275	**.849 (.778 - .920)****	19 - 319	**.718 (.660 - .777)****	122 - 216
Social skills	**.755 (.690 - .820)****	70 - 309	**.733 (.632 - .834)****	22 - 357	.657 (.600 - .714)**	136 - 243
Skills to take care of oneself	.646 (.569 - .723)**	69 - 309	.655 (.537 - .772)*	21 - 357	.582 (.520 - .644)**	136 - 242
Financial skills	.634 (.552 - .716)**	63 - 285	**.701 (.593 - .810)****	19 - 329	.637 (.576 - .698)**	132 - 216
Impulsivity	**.727 (.657 - .798)****	68 - 308	**.831 (.761 - .902)****	22 - 354	.651 (.592 - .710)**	135 - 241
Antisocial behavior	**.736 (.674 - .799)****	69 - 304	**.781 (.708 - .854)****	22 - 351	.684 (.628 - .739)**	135 - 238
Hostility	**.726 (.658 - .794)****	69 - 309	**.779 (.683 - .876)****	22 - 356	.635 (.577 - .694)**	136 - 242
Sexually deviant behavior	.573 (.496 - .651)	69 - 307	.576 (.449 - .703)	22 - 354	.596 (.535 - .656)**	135 - 241
Manipulative behavior	.675 (.604 - .745)**	67 - 303	**.733 (.639 - .827)****	21 - 349	.630 (.571 - .688)**	135 - 235
Compliance with rules	**.745 (.684 - .806)****	70 - 307	**.756 (.661 - .852)****	22 - 355	**.734 (.680 - .789)****	136 - 241
Antisocial associates	.662 (.586 - .738)**	67 - 293	.656 (.524 - .788)*	21 - 339	.606 (.545 - .666)**	130 - 230
Medication use	.614 (.523 - .706)*	46 - 170	.570 (.419 - .722)	15 - 201	.517 (.438 - .596)	83 - 133
Psychotic symptoms	.584 (.490 - .678)	50 - 195	.608 (.446 - .769)	15 - 230	.491 (.415 - .566)	85 - 160
Skills to prevent substance use	**.728 (.652 - .804)****	62 - 275	**.770 (.668 - .872)****	20 - 317	**.756 (.703 - .810)****	122 - 215
Drug use	.680 (.604 - .755)**	65 - 288	**.704 (.606 - .802)****	21 - 332	**.784 (.731 - .837)****	126 - 227
Skills to prevent physically aggressive behavior	**.747 (.680 - .815)****	65 - 254	**.757 (.654 - .860)****	20 - 299	.614 (.549 - .678)**	121 - 198
Skills to prevent sexually deviant behavior	**.711 (.594 - .828)****	18 - 106	**.733 (.625 - .842)**	4 - 120	**.725 (.622 - .829)****	35 - 89
Protective behavior	.697 (.614 - .779)**	46 - 164	**.704 (.598 - .810)****	15 - 195	.631 (.554 - .708)**	81 - 129
Problem behavior	**.785 (.719 - .850)****	64 - 283	**.847 (.771 - .923)****	20 - 327	.697 (.640 - .753)**	128 - 219
Resocialization skills	**.787 (.722 - .852)****	58 - 259	**.831 (.751 - .912)****	18 - 299	**.720 (.660 - .781)****	118 - 199

* p <.05, ** p <.01, *** positive-negative outcomes

Discussion

The aim of this study was to assess the predictive validity of the IFTE for both positive treatment outcomes (i.e., leave) and negative treatment outcomes (i.e., inpatient incidents), in order to examine whether the IFTE can be used in clinical decision-making. ROC analyses were conducted for three types of leave modalities: guided leave, unguided leave and transmural leave for the whole group of patients; and for three types of incidents: general aggression, physical aggression and serious violation of UDS for patients with main diagnoses, personality disorders and personality disorders with SUDs. Leave modalities are one of the most important interventions in rehabilitation treatment (Lyall & Bartlett, 2010), and incidents may have serious implications for care and treatment plans and risk management strategies (Vojt, Thomson, & Marshall, 2013). Though patients with and without granted leave requests differed significantly on factor scores, predictive validity for leave requests and UDS violations was marginal. Predictive validity for aggression and physical aggression in particular showed better predictive values.

Results imply a marginal predictive validity for all factor scores for all leave approvals, except *problem behavior* for guided leave approvals and *resocialization skills* for transmural leave approvals. All resocialization items, apart from *skills to take care of oneself,* showed a significant predictive validity for unguided leave. *Cooperation with treatment, labor skills, compliance with rules* and *skills to prevent substance use* were most predictive of unguided leave. *Cooperation with treatment* was most predictive of guided leave. The protective skills items *cooperation with treatment*, and *skills to prevent substance use* were significantly predictive of all leave modalities. *Antisocial behavior, hostility, manipulative behavior,* and *compliance to rules* were all marginally predictive of transmural leave.

The factors *protective behavior* and *resocialization skills* were significantly higher for patients with a granted guided leave request and unguided leave request. The *problem behavior* factor was significantly lower for patients with a granted unguided leave approval and a transmural approval, and the factor *protective behavior* was also significantly higher for patients with a granted transmural request. This means that the IFTE shows more skills for patients with granted leave requests and less problematic behavior for patients with unguided and transmural leave requests.

These results, together with the marginal predictive validity, cautiously indicate that these factors could be taken into consideration in decision-making. However, even though these items are significantly predictive, the values are not high. An AUC value of .90 or higher would be most preferable, followed by a value in the area of .80 - .90 (Spreen et al., 2014). The results could possibly be influenced by the moment when leave requests are currently made.

One of the aims of routine outcome monitoring is to shorten treatment, and the current leave approvals possibly do not occur at the most optimum time in treatment.

Previous studies have studied factors predicting discharge or length of stay (Ross, Querengässer, Fontao, & Hoffman, 2012; Andreasson et al., 2014). These studies have found a relationship of mostly historical or diagnostic factors with discharge, such as mood disorder, psychotic disorder, history of substance use and absconding. Absconding and current conviction for violent crime were related to longer hospital stay, and mood disorder was related to shorter hospital stay (Andreasson et al., 2014). Ross et al. (2012) found that mostly historical factors, such as type of offense and psychiatric disorder, were related to discharge. While these factors provide important information at the start of treatment, they supply fewer monitoring opportunities.

In order for ROM to aid decision-making, they should assess changeable factors. No previous studies are known to the authors to have studied the predictive validity of dynamic risk and protective factors for leave approvals or positive treatment outcomes, other than discharge. Previous studies have focused mainly on violations during leave or unauthorized leave (O'Shea & Dickens, 2014). De Vries and Spreen (2012) mentioned that 'the factors on which therapists base their decisions are now barely studied.' When they studied decision-making with the risk assessment tool HKT-30, they found a higher value of *social skills, self-reliance, hostility, impulsivity* and *coping skills* in patients who violated rules during leave, and a predictive value of (AUC = .71) for the combination of *substance use, impulsivity* and *a lack of empathy* for violation of rules during leave (De Vries & Spreen, 2012). Similar to results in this study, the results from De Vries and Spreen (2012) imply that these factors should be taken into consideration in leave-related decision-making.

For inpatient incidents, the *problem behavior* and *resocialization skills* factor scores showed a reasonable predictive validity for general and physical aggression in the three groups. The factors *resocialization skills* even showed a good predictive validity for physical aggression in the PSDS patient group and the factor *problem behavior* showed a good predictive validity for physical aggression in the PSDS groups and main diagnostic group. The *protective factor* showed a reasonable predictive value for physical aggression in the PSDS group and a marginally predictive value for physical aggression in the other groups, and for general aggression in all three groups.

Most items showed a significant moderate predictive value for general aggression. *Coping skills, balanced daytime activities, labor skills, social skills, impulsivity, antisocial behavior, hostility* and *compliance with rules* showed the best predictive values for both general and physical aggression in all groups. *Labor skills*, *coping skills,* and *impulsivity* even showed a good predictive value for physical aggression in the PSDS group, and *impulsivity* showed a

good predictive value for physical aggression in the PSD group. Slightly higher predictive items were found in the PSDS group compared to the PSD and main diagnostic groups. However, *medication use, sexually transgressive behavior, financial skills* and *psychotic symptoms* showed low predictive values in the PSDS group, as did *problem insight, responsibility for the crime, sexually transgressive behavior* and *psychotic symptoms* in the PSD group for both general and physical aggression. *Antisocial associates* was marginal for physical and general aggression in all groups. The predictive validity of *responsibility for the crime* was low for physical aggression in all groups. Finally, *problem insight, responsibility for the crime, skills to take care of oneself, sexually transgressive behavior, medication use* and *psychotic symptoms* were all low predictors for physical aggression in the main diagnostic group and PSD group.

We would have expected higher predictive validity for the item *skills to prevent physically aggressive behavior*. However, this item is scored on the basis of particular skills necessary for an individual patient to prevent future violent recidivism, and these skills are different for different patients. As violence can be explained by different factors, as we see in these results, this may possibly influence the predictive validity of this specific item in a group assessment. Grevatt, Thomas-Peter, and Hughes (2004) even found violence throughout the lifespan to be a protective factor for institutional violence, possibly due to proper identification and management by the treatment teams. This could also be the case in our population.

Even though one would expect a higher predictive validity for the item *drug use,* as it is often marked as a risk factor and even considered a factor that complicates resocialization (De Vries, & Spreen, 2012), this is not a surprisingly low value. As we mentioned before, patients suspected of having used a substance receive extra guidance and are not allowed to go on leave. They are often guided more closely throughout the day, giving patients less opportunity to cause incidents. This could possibly influence the predictive validity of this item.

The generally moderate predictive validity for short-term incidents in forensic psychiatry is similar to results found in previous studies, in which clinical factors show a better predictive value than static historical risk factors (Grevatt et al., 2004). However, these results were assessed for incidents at the start of treatment. Vojt et al. (2013) did not find significant predictive values of clinical HCR-20 items for short-term incidents, whereas Wilson, Desmarais, Nicholls, Hart, and Brink (2013) found moderate to good predictive validity of the short-term assessment of risk and treatability (START; Webster, Martin, Brink, Nichols, & Desmarais, 2009), and the clinical HCR-20 items for institutional violence.

The predictive validity for UDS violations were considerably low. Even though most items did show a significant predictive value, most AUC values were lower than .65. The factor *problem behavior* showed somewhat higher AUCs in all groups, and the items *drug use* and *skills to prevent substance use* were reasonably predictive. This is somewhat similar to the

values found in the previous study by Schuringa et al (2018). Schuringa et al. (2018) found a discriminatory value of two IFTE items *skills to prevent substance use,* and *drug use* for actual drug use in the assessment period and future drug use. The *resocialization* factor was reasonably predictive in the PSD and PSDS groups. The definition of UDS violations may have been too broad in this study as refusal and unreliable screenings were also included. However, we cannot know for certain whether a patient who has refused a UDS has used a substance. Patients may have different reasons for refusing UDS: they may find the procedure too invasive or they may refuse out of a general refusal to cooperate with internal procedures.

Limitations

A limitation of this study is the fact that all IFTEs were assessed in preparation of the bi-annual patient meetings. The IFTEs were assessed by the patients' treatment team, and scores were available to the team. This may have influenced treatment decisions even though the IFTE is not yet used as an indicator for leave modules. Treatment and treatment plans are evaluated (with or without the use of IFTE assessment) and possibly adjusted in this meeting. This might affect the IFTEs predictive value while treatment had possibly already been adjusted on the basis of signs observed by the treatment team.

Treatment teams in forensic psychiatric settings are trained to observe possibly alarming signs. However, we know that, in risk assessment, actuarial and structured professional judgments are more reliable than clinical judgments (Bonta & Andrews, 2007). Whipple and Lambert (Whipple & Lambert, 2011), moreover, doubt the ability of clinicians to properly recognize treatment response. Monitoring of signs or progress in treatment could be more reliably assessed with the help of the IFTE, even more so if the IFTE is assessed by multiple members of a treatment team with a view to obtaining a composite score.

Another limitation in this study of predictive validity for ROM assessment is the fact that not all items were related to patients' aggression. ROM is conducted to evaluate individual treatment. It is essential that the reports are read by the treatment team, who know which items are important in considering a specific patient. The IFTE reports also provide the opportunity to mark relevant treatment factors for individual patients. This could possibly improve the considerations made with the help of individual IFTE treatment evaluations.

Even though we attempted to assemble multiple raters, some items were assessed by a single rater. The inclusion of IFTEs rated by at least three treatment team members could possibly lead to more reliable scores, which, of course, would produce better results. The IFTEs predictive validity for withdrawals during leave was not studied due to the low number of withdrawals during the study period. In the future, however, it would be interesting to study if the IFTE can also predict this outcome.

It was intended that IFTEs were assessed every four to six months in preparation of individual treatment evaluations. However due to the dynamic setting in which this study was conducted, treatment evaluations were postponed, brought forward, or IFTEs were not assessed by at least one coach/staff member or psychologist/psychiatrist, leading to a more diverse period between assessments. This could also influence the period between an IFTE assessment and leave approvals or incidents. All data in this study was primary treatment evaluation and IFTEs are primarily used as additive information concerning treatment progress at the moment of an individual treatment evaluation. It was therefore not possible to control the IFTE assessment frequency as would be the case in a study, which is merely used for scientific research. More routinely assessments could possibly benefit results.

Even though we have assessed the predictive validity for granted leave requests for the whole group of patients, we cannot conclude that predictive values would not differ for different patient groups. However, we did not study the predictive values for the different groups, while we expected that similar factors would be considered in the approval of leave, and breaking the group down would lead to very small numbers of granted leaves. Future study would have to look into a possible difference. Also, we did not assess predictive values for incidents for smaller diagnostic groups. We cannot make conclusions for the predictive validity for these patient groups.

Finally, even though we used multiple sources (VIM, reports and official measures) to detect different forms of aggression, it is likely that not all aggressive incidents were reported in these documents. It is possible that aggression, especially verbal aggression, occurs more often than we report in clinical institutions.

Conclusions and clinical use

An advantage of the IFTE is that its items are based on the clinical dynamic risk factors of the HKT-R which are predictive of future recidivism (Spreen et al., 2014). Whereas the HKT-R's five-point scale gives us less opportunity to assess change on a six-monthly basis, the IFTE allows us to assess clinical dynamic risk factors and relevant skills on a routine basis. This gives us the possibility to assess predictive values for treatment outcomes in a shorter period, as changes in both inpatient risk and progress can be monitored earlier.

Our results tentatively imply that the IFTE can be used in treatment and can support treatment decision-making. The predictive values are moderate and stress the importance of considerations based on plural information sources in decision-making. However, scores on the IFTE could imply the consideration of changes in treatment plans, whether these be more intensive treatment due to a higher risk of deviant behavior or the next step in treatment in the case of reduced problem behavior and improved rehabilitation and protective

skills. This may concern not only leave but also other forms of raised autonomy, such as the increase of balanced daytime activities.

The inclusion of the IFTE in treatment considerations could facilitate treatment duration. Over the past few years, forensic treatment in Dutch FPCs has risen to approximately nine years (Van Nieuwenhuizen et al., 2011). Earlier leave or other forms of raised autonomy might benefit the treatment period. The study by Spreen et al. (2014) shows that patients who have experienced all leave modalities in their treatment show less recidivism than patients who skipped a leave module. This implies that gradual rehabilitation is important in all leave modules; the period between admission and first leave approval, however, has been extended (Ter Horst, Jessen, Bogaerts, & Spreen, 2015), which is not beneficiary for treatment outflow. Start of leave at the appropriate moment in treatment and requested with the proper considerations, therefore, could benefit treatment.

This study shows the importance of a patient's strengths in treatment considerations. Resocialization skills are not inferior to problem behavior in this study. This is similar to results found by Wilson et al. (2013), who found predictive values of protective items for institutional violence, and to results by Wilson, Desmarais, Nicholls, and Brink (2010), who found that patients who did not show inpatient violence had higher patient strength scores than patients who did show inpatient violence. This underscores the importance of treating risk factors and developing personal strengths and skills, as claimed by the well-known rehabilitation models (Bonta, & Andrews, 2007; Ward, Mann, & Gannon, 2005).

Future directions

This study has assessed the predictive validity for inpatient outcomes in large identifiable diagnostic patient groups. Other diagnostic combinations, even if smaller in number, do occur in the forensic psychiatric population (Van der Veeken et al., 2017). We have not studied predictive values for these smaller identifiable groups in forensic psychiatry in this study. However, it is important to assess which items in forensic ROM are predictive of future incidents for different patient groups. Future research might study whether the IFTE is also predictive of inpatient incidents for these groups of patients, preferably in a larger group of patients. A larger dataset could also give us the opportunity to study which items of the IFTE are important in considering leave requests for different patient groups. Even more, with a larger dataset and ROM assessments throughout the entire treatment, it would be possible to assess which factors contribute to a successful treatment outcome, in the meaning of unconditional leave without recidivism. In this way we could study successfully proceeded leave modules, and which IFTE factors and diagnostic factors contribute to a successful treatment and can be used in decision making for the differing groups.

Future research should examine if use of the IFTE in treatment considerations truly affects treatment in a positive way. The first results of the use of ROM in treatment and treatment decisions are promising (Carlier et al., 2012). In addition, ROM also gives us the opportunity to discuss treatment progress, stagnation or decline with patients and to set treatment goals in consultation. Results in regular mental healthcare show that these feedback discussions have a positive effect on treatment cooperation (Van der Feltz-Cornelis et al., 2014). Research could examine whether forensic psychiatry patients would also benefit from discussing outcome measurements.

To support decision-making in the matter of leave modalities, the short-term predictive validity of the IFTE for violations during leave, especially absconding, could also be studied. The IFTE could also be used to monitor patients' functioning over time in relation to the moment of leave requests, which could cause risk assessment tools to be used in assessing absconding risks annually and more routinely. A larger study, involving multiple settings, would be advised as absconding is relatively infrequent.

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders – DSM-IV-TR.* Arlington, VA: American Psychiatric Association.

Andreasson, H., Nyman, M., Krona, H., Meyer, L., Anckarsäter, H., Nilsson, T., & Hofvander, B. (2014). Predictors of length of stay in forensic psychiatry: The influence of perceived risk of violence. *International Journal of Law and Psychiatry, 37,* 635 – 642. doi: 10.1016/j/ijlp/2014.02.038

Andrews, D. A., & Bonta, J. (2010). The psychology of criminal conduct, fifth edition. Matthew Bender and Company, Providence.

Andrews, D. A., Bonta, J. L., & Wormith, J. S. (2011). The risk-need-responsivity (RNR) model. Does adding the good lives model contribute to effective crime prevention? *Criminal Justice and Behavior,38,* 735-755. doi: 10.1177/0093854811406356

Bernstein, D. P. (2012). "*Big boys don't cry!" Or do they? Can forensic patients change?* (Inaugural lecture as 'professor of forensic psychotherapy) Maastricht University, Maastricht, The Netherlands.

Bonta, J., & Andrews, D. A. (2007). Risk need responsivity model for offender assessment and rehabilitation, 2007-06. Ottawa; Public Safety Canada.

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders a meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, *36*, 567-590. doi: 10.1177/ 0093854809333610

Carlier, I. V. E., Meuldijk, D., Van Vliet, I. M., Van Fenema, E., Van der Wee, N. J. A., & Zitman, F. G. (2010). Routine outcome monitoring and feedback on physical mental health status; evidence and theory. Journal of Evaluation in Clinical Practice, 18, 104-110. doi: 10.1111/j.1365-2753.2010.01543.x

Carlier, I. V. E., Meuldijk, D., Van Vliet, I., Van Fenema, E. M., Van der Wee, N. J. A., & Zitman, F. G. (2012). Empirische evidentie voor de effectiviteit van routine outcome monitoring; een literatuuronderzoek. [Empirical evidence for the effectiveness of Routine Outcome Monitoring. A study of the literature] *Tijdschrift voor Psychiatry*, *54*, 121-8. Retrieved from: http://www.tijdschriftvoorpsychiatrie. nl/assets/articles/TvP12-02_p121-128.pdf

De Beurs, E., Den Hollander-Gijsman, M. E., Van Rood, Y. R., Van der Wee, N. J. A., Giltay, E. J., Van Noorden, M. S., ... Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands; practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology and Psychotherapy, 18,* 1-12. doi: 10.1002/cpp.696

De Vries, K., & Spreen, M. (2012). De HKT-30 als instrument voor beslismomenten binnen een tbs-behandeling. [Can the clinica items of the risk assessment instrument HKT-30 predict to what extent forensic patients will break agreements during the resocialization phase of treatment?] *Tijdschrift voor psychiatrie*, *54*, 429 – 438 Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/TvP12-05- p429-438.pdf

Dienst Justitiële Inrichtingen (2014). Risicotaxatie-instrument HKT-R digitaal beschikbaar [Risk assessment-tool HKT-R digitally available]. Retrieved from https:// www.forensischezorg.nl/kennis-delen/actueel/nieuws/risicotaxatie-instrument-hkt-r-digitaal-beschikbaar

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). HCR-20 V3 Historical. *Clinical, Risk Management (Version 3): Professional guidelines for evaluating risk of violence*. Vancouver, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Ellwood, P. M. (1988). Outcomes management. A technology of patient experience. *The New England Journal of Medicine, 23*, 1549-56. doi: 10.1056/NEJM198806093182327

Goethals, K. R., & Van Marle, H. J. C. (2012). Routine outcome monitoring in de forensische psychiatrie: een lang verhaal in het kort [Routine outcome monitoring in forensic psychiatry: A long story cut short]. *Tijdschrift voor Psychiatrie, 54 (2),* 179 – 183. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/issues/450/articles/9335

Grevatt, M., Thomas-Peter, B., & Hughes, G. (2004). Violence, mental disorder and risk assessment: can structured clinical assessments predict the short-term risk of inpatient violence? *Journal of Forensic Psychiatry & Psychology*, *15*, 278-292. doi: 10.1080/1478994032000199095

Gunderman, R. B., & Chan, S. (2013). The 13-Point Likert Scale. *Academic radiology*, *20*(11), 1466-1467. doi: http://dx.doi.org/10.1016/j.acra.2013.04.010

Jeffery, R., & Woolpert, S. (1974). Work furlough as an alternative to incarceration: assessment of its effects on recidivism and social cost. *The Journal of Criminal Law and Criminology, 65*, 405-415. Retrieved from: http://scholarlycommons.law. northwestern.edu/cgi/viewcontent.cgi?article=5908&context=jclc

Lambert, M. J., Whipple, J. L., Hawkins, E. D., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003) Is it time for clinicians to rouintely track patient outcome? A meta- analysis. *American Psychological Association, D12,* 288 – 301. doi: 10.1093/clipsy.bpg025

Larson, E. B. (2013). Building trust in the power of "big data"research to serve the public good. *The Journal of the American Medical Association, 23,* 2443 - 2444. Retrieved from http://jama.jamanetwork.com/

LeClair, D. P., & Guarino-Ghezzi, S. (1991). Does incapacitation guarantee public safety? Lessons from the Massachusetts furlough and prerelease programs. *Justice Quarterly, 8*, 9 – 36. doi: 10.1080/07418829100090891

Lyall, M., & Bartlett, A. (2010). Decision making in medium security: Can he leave? *The Journal of Forensic Psychiatry & Psychology, 21,* 887 – 901. doi: 10.1080/14789949.2010.500740

Ogloff, J. R. P. (2011). Offender rehabilitation: From "nothing works" to what next? *Australian Psychologist, 37,* 245 – 252. doi: 10.1080/00050060210001706936

O'Shea, L. E., & Dickens, G. L. (2014). Predictive validity of the START for unauthorized leave and substance abuse in a secure mental health setting: A Pseudo-prospective cohort study. *International Journal of Nursing Studies, 52,* 970 - 979. doi: 10.1016/j.ijnurstu.2015.02.007

Rijksoverheid. (n.d.) Verlof tijdens tbs [leave during TBS]. Retrieved from: https:// www.rijksoverheid.nl/onderwerpen/tbs/inhoud/verlof-tijdens-tbs

Rijksoverheid. (n.d.[1]). Na de tbs. [After TBS]. Retrieved from http://www.rijksoverheid.nl/onderwerpen/tbs/na-de-tbs

Ross, T., Querengässer, J., Fontao, M. I., & Hoffman, K. (2012). Predicting discharge in forensic psychiatry: The legal and psychosocial factors associated with long and short stays in forensic psychiatric hospitals. *International Journal of Law and Psychiatry, 35,* 213 – 221. doi: 10.1016/j.ijlp.2012.02.011

Schuringa, E., Heininga, V. E., Spreen, M., & Bogaerts, S. (2018). Concurrent and predictive validity of the instrument for forensic treatment evaluation: From risk assessment to routine, multidisciplinary treatment evaluation. *International Journal of Offender Therapy and Comparative Criminology, 62,* 1281-1299. doi: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*:2, 127-144. doi: 10.1080/15228932.2014.897536

Spreen, M., Brand, E., Ter Horst, P., & Bogaerts, S (2014). *Handleiding en Methodologische Verantwoording HKT-R, Historisch, Klinische en Toekomstige - Revisie*. [Manual and methodological justification HKT-R, Historical, Clinical and Future – Revised] Dr. van Mesdag kliniek: Groningen

Tengström, A., Hodgins, S., Müller-Isberner, R., Jöckel , D., Freese, R., Özokyay, K., & Sommer, J. (2006) Predicting Violent and Antisocial Behavior in Hospital Using the HCR-20:The Effect of Diagnoses on Predictive Accuracy. *International Journal of Forensic Mental Health, 5*:1, 39-53. doi: 10.1080/14999013.2006.10471229

Ter Horst, P., Jessen, A., Bogaerts, S., & Spreen, M. (2015). Behandelduur tot eerste toekenning tbs-verlof in een Nederlandse forensische psychiatrische kliniek; dossieronderzoek naar invloed van patiëntkenmerken en delict. [clinical decision making with regard for granting of escorted leave for forensic patients detained by court order in a Dutch psychiatric clinicl. Role of gender, disorder and the type of offence in the procedure]. *Tijdschrift voor Psychiatrie, 57,* 314 - 322. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/issues/489/articles/10559

Van den Brink, R. H. S., Hooijschuur, A., Van Os, T. W. D. P., Savenije, W., & Wiersma, D. (2010). Routine violence risk assessment in community forensic mental health care. *Behavioural Sciences and the Law, 28,* 396 – 410. doi: 10.1002/bsl.904

Van der Feltz-Cornelis, C. M., Andrea, H., Kessels, E., Duivenvoorden, H. J., Biemans, H., & Metz, M. (2014). Shared decision making in combinatie met ROM bij patiënten met gecombineerde lichamelijke en psychische klachten; een klinisch-empirische verkenning [Does routine outcome monitoring have a promising future? A investigation into the use of shared decision-making combined with ROM for patients with a combination of physical and psychiatric symptoms]. *Tijdschrift voor psychiatrie*, *56*(6), 375-384. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/56-2014-6-artikel-vanderfeltz-cornelis.pdf

Van der Veeken F. C. A., Bogaerts, S., & Lucieer, J. (2017) Patient profiles in Dutch forensic psychiatry based on risk factors, pathology, and offense. *International Journal of Offender Therapy and Comparative Criminology, 61*, 1369-1391. doi: 10.1177/0306624X15619636

Van Nieuwenhuizen, C. H., Bogaerts, S., De Ruiter, E. A. W., Bongers, I. L., Coppens, M., & Meijers, R. A. A. C. tbs-behandeling geprofileerd; een gestructureerde casussenanalyse [*Profiling TBS-treatment: A structured cases analysis]*. Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) Ministerie van Veiligheid en Justitie. 2011. Retrieved from http://www.wodc.nl/onderzoeks database/voor studie-onderzoek-tbs.aspx

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

Vojt, G., Thomson, L. D. G., & Marshall, L. A.(2013). The predictive validity of the HCR-20 following clinical implementation: does it work in practice? *The Journal of Forensic Psychiatry & Psychology, 24*:3, 371-385. doi:10.1080/14789949.2013.800894

Ward, T., Mann, R. E., & Gannon, T. A.(2005). The good lives model of offender rehabilitation: Clinical implications. *Agression and Violent Behavior, 12,* 87 – 107. doi:10.1016/j.avb.2006.03.004

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

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Short-Term Assessment of Risk and Treatability (START)* (Version1.1). Coquitlam, BC, Canada: British Columbia Mental Health and Addiction Services

Werkgroep Risicotaxatie Forensische Psychiatrie [Taskforce Risk assessment Forensic Psychiatry]. (2002). *Handleiding HKT-30 versie 2. Risicotaxatie in de forensische psychiatrie* [Guide HKT-30 version 2. Risk Assessment in Forensic Psychiatry]. Den Haag, The Netherlands: Ministerie van Justitie, Dienst Justitiële Inrichtingen.

Wilson, C. M., Desmarais, S. L., Nicholls, T. L., & Brink, J. (2010). The role of client strengths in assessments of short-term violence risk. *International Journal of Forensic Mental Health Services, 9,* 282 – 293. doi: 10.1080/14999013.2010.534694

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

Whipple, J. L., & Lambert, M. J. (2011). Outcome measures for practice. *Annual review of clinical psychology*, *7*, 87-111. doi: 10.1146/annurev-clinpsy-040510-143938

CHAPTER SIX

Forensic Psychiatric Treatment Evaluation The Clinical Evaluation of Treatment Progress with Multiple Forensic Routine Outcome Monitoring Measures

Frida C. A. van der Veeken
Jacques Lucieer
Stefan Bogaerts

This is an adapted version of the article published as: Van der Veeken, F. C. A., Lucieer, J., & Bogaerts, S. (2018). Forensic psychiatric treatment evaluation: The clinical evaluation of treatment progress with repeated forensic routine outcome monitoring measures. *International Journal of Law and Psychiatry, 57*, 9-16. doi: 10.1016/j.ijlp.2017.12.002

Abstract

The likelihood of recidivism is considered to be the most important outcome measure in forensic psychiatry. Therefore, forensic psychiatric treatment focuses on the reduction of the risk of recidivism by treating dynamic risk and protective factors, aiming to reduce risk factors while enhancing protective factors during treatment.

The goal of this study is to assess treatment progress with the Instrument for Forensic Treatment Evaluation (IFTE) in a Dutch forensic psychiatric center ($N = 240$). Latent Class Analysis was conducted to reconfirm previously found patient profiles. Patient profiles were based on risk factors, psychopathology, and offense type. Repeated measures ANOVAs were conducted to assess treatment progress for the whole patient group, for high and low risk patients, and for patients who had been in treatment for a period longer and shorter than one year.

Latent Class Analysis has not reconfirmed the previously found profiles, therefore a repeated measures ANOVA was not conducted on profile level. On group-level, no significant progress was found. Though, patients with low protective behavior, low resocialization scores, and high problem behavior scores displayed significant treatment progress. Patients with low problem behavior showed a significant increase of problematic behavior and patients with high protective behavior a decrease of protective behavior. Results indicated an interaction effect between time of admission and the factor *resocialization skills*, however this effect was not found for the other two factors.

Results imply that higher risk patients can show more treatment progress than lower risk patients.

Introduction

The Risk-Need-Responsivity model (RNR model; Bonta & Andrews, 2007) is one of the leading theoretical models in the treatment of offenders. The Risk principle implies that the frequency and intensity of treatment and the level of security should match the nature and severity of criminogenic needs (i.e., dynamic risk factors directly related to criminal behavior) (Bonta & Andrews, 2007). The Need principle emphasises that the criminogenic needs related to the offending behavior should be addressed in treatment. The Responsivity principle implies that treatment must be in accordance with offenders' characteristics, learning styles, and abilities (Bonta & Andrews, 2007; Ward, Melser, & Yates, 2007). With crime-related needs reduced, skills enhanced, and treatment matched to the level of risk and patient characteristics, offenders can thus be prepared for rehabilitation.

The main treatment objectives − criminogenic needs − are related to self-regulation skills (Bonta & Andrews, 2007; Ward et al., 2007). Self-regulation is the offender's ability to alter deviant behavior and responses (Baumeister & Vohs, 2007), involving, for instance, withholding behavior and behavior aiming to attain a desired goal (Rothman, Baldwin, Hertel, & Fuglestad, 2011). Self-regulation failure can result in disruptive or deviant behavior caused by limited self-control (Wagner & Heatherton, 2015) and uncontrolled impulses (Bogaerts, Vervaeke, & Goethals, 2004; Bogaerts, Vanheule, & DeClercq, 2005). Strengthening self-regulatory skills, therefore, is important to control problematic behavior and to enhance protective behavior and resocialization skills (Ward et al., 2007).

A core problem in forensic psychiatry, however, is the heterogeneity of the forensic population in terms of type of offense, psychopathology, and risk factors. De Jonge, Nijman, and Lammers (2009) recommend that future studies should focus on differences in psychopathology and offense types in relation to treatment progress. Patient profiles can contribute to homogeneity and can support researchers and clinicians to provide insight into the levels of risk, criminogenic needs, and self-regulatory skills for different patient groups. Several studies have examined patient profiles based on psychopathology, risk factors, and/or crimes committed (Yiend, Freestone, Vazques-Montes, Holland, & Burns, 2013; Van Nieuwenhuizen et al., 2011; Bogaerts & Spreen, 2011: Van der Veeken, Bogaerts, & Lucieer, 2017).

Recently, Van der Veeken et al., (2017) identified four patient profiles by clustering axis I and II psychopathology, type of offense, and historical and dynamic risk factors. The first antisocial profile includes patients with mainly a cluster B personality disorder, with 50.6% being diagnosed with an antisocial personality disorder and a substance use disorder (SUD) (53%). Patients in this profile show high historical risk factors, such as criminal an-

tecedents, a problematic history of school and work, and lower dynamic risk factors, such as hostility and problematic treatment attitude. They show higher Psychopathy checklist-revised (PCL-R; Hare, 1991) scores (M = 24.19, SD = 6.16), with the factor two score (M = 12.54, SD = 2.95) higher than the factor one score (M = 9.18, SD = 3.44). The second mixed profile displays frequent cluster B diagnoses (45% one cluster B personality disorder and 15% two cluster B personality disorders or one cluster B and one cluster A disorder), often in co-morbidity with psychotic disorders (49%) or SUDs (38% primary SUD, 56% secondary SUD). They display high historical and dynamic risk factors and show high PCL-R scores (M = 24.16, SD = 7.04). Both the first and the second profile display a mix of offenses, such as homicide, assault, and violent property offenses.

The third profile consists of patients with mainly a pervasive developmental disorder (14% primary diagnosis, 10% secondary diagnosis), paedophilia (24% primary diagnosis, 9% secondary diagnosis), or other (31%), such as an affective disorder, paraphilia, or a dysthymic disorder. Patients in this profile display low historical risk factors and high dynamic risk factors, particularly limited *empathic skills, social skills*, and *crime responsibility*. They are often convicted of child sexual abuse or homicide. They show lower PCL-R scores (M = 15.72, SD = 5.91), but, if we look at the four-facet model of the PCL-R (Hare & Neumann, 2005), their affective facet score is high (M = 6.33, SD = 1.56).

The fourth profile includes patients with psychotic disorders (23% primary diagnosis, 5% secondary diagnosis), with a low cluster B comorbidity (19%). Some offenders have been convicted of homicide (42%) and show low historical risk factors such as *criminal history* and *previous violation of rules*, low dynamic risk factors, particularly *impulsivity* and *hostility*, and low PCL-R scores (M = 13.19, SD = 6.13) (Van der Veeken et al., 2017).

Treatment evaluation at the patient-profile level could provide insight into treatment progress for the identified patient groups differing in pathology, risk factors, and offense type. To our knowledge, this would be the first study on forensic treatment progress related to patient profiles and in line with the recommendations made by De Jonge, Nijman and Lammers (2009). Routine outcome monitoring (ROM) is developed to routinely evaluate a patient's treatment progress over time (Ellwood, 1988). ROM has been widely studied in general mental healthcare but is still in its infancy in forensic psychiatry. As monitoring problem behavior, protective behavior, and resocialization skills is a core mission in evaluating treatment progress in forensic psychiatry, the Instrument for Forensic Treatment Evaluation (IFTE) has been developed (Schuringa, Spreen, & Bogaerts, 2014). The IFTE assesses protective behavior, problem behavior, and resocialization skills (Schuringa et al., 2014), and can be used to monitor criminogenic needs and the development of self-regulation skills, such as protective behavior and resocialization skills. Recent studies have shown good

psychometric properties of the IFTE (Schuringa et al., 2014, Van der Veeken, Lucieer, & Bogaerts, 2018) and a moderate predictive contribution in the prediction of the likelihood of future incidents during inpatient treatment (Van der Veeken, Bogaerts, & Lucieer, 2016; Schuringa, Heininga, Spreen, & Bogaerts, 2018).

The general goal of this study was to assess treatment progress for the whole group of forensic psychiatric patients over six measurement periods. We expected that treatment progress would have differed between different patient profiles. Before that, we wanted to reassess the profiles previously found in the study of Van der Veeken et al. (2017) based on historical and dynamic risk factors, psychopathology, and type of offense.

In line with Andrews and Dowden's (2006) findings that patients with a higher risk profile benefit more from treatment, we expected that patients who show high problem behavior, or low protective behavior, and less appropriate resocialization skills would have shown more treatment progress than patients who show low problem behavior, high protective behavior, and appropriate resocialization skills at the first IFTE assessment.

Finally, the study by Nijman, De Kruyk, and Van Nieuwenhuizen (2004), showed that most behavioral changes occur during the first 14 months of treatment. We expected to find the same pattern in our patient population. ROM assessments were implemented for all patients in two forensic psychiatric centers, regardless how long they had already been in treatment. Therefore, we also wanted to assess if patients who had been in treatment for a period shorter than 12 months at the first ROM assessment would show more progress on problem behavior, protective behavior, and resocialization skills than patients who had been in treatment for a longer period since the first ROM assessment, similar to the results found by Nijman et al. (2004).

Method

Procedure

This study was conducted in two forensic psychiatric centers (FPCs) in the Netherlands. Patients who are going through an FPC are involuntary admitted with a TBS-measure (disposal by order of the state, meaning; terbeschikkingstelling) and receive inpatient care (Van Marle, 2002). In the case of an offense with a minimal penalty of four years, a person can receive a TBS-measure when a mental disorder is present at the time of the offense and the mental disorder is related to the offense; when the risk of reoffending is high, and when the offender cannot be held fully accountable for the crime (De Boer, Whyte, & Maden, 2008). A TBS-measure can be extended every one or two years by a judge when the risk of recidi-

vism is still problematic. The main goal of a TBS-measure is to rehabilitate and integrate the patient back into society with a low risk of recidivism. Patients receive care and treatment by a multidisciplinary team, including psychiatrists, psychologists, social workers, art and psychomotor therapists and others. Routine outcome monitoring (ROM) has been implemented in both FPCs since 2011. During implementation, all therapists were informed about the procedure and received information about the questionnaires.

Treatment evaluation with the IFTE occurs every four to six months for all patients, parallel to the multidisciplinary treatment evaluation meetings. For most patients, several evaluators (e.g., head of treatment, psychologist, and coaches) filled in the IFTE, obtaining a multi-perspective view of the patients' changes compared with the previous scores four to six months earlier. The treatment team could access the IFTE with a shortened instruction in a patient's individual document, and subsequently the assessments could be conducted in the patients' electronic patient file (EPF). The treatment team receives a report of the ROM measurement in preparation of the treatment evaluation meeting.

IFTEs with a minimum period of three months between two evaluations and evaluations rated by at least one psychologist, psychiatrist, or coach were included in this study. All data used in this study concern primary treatment information and were retrieved from patient files. Additional informed consent was not required. Data were anonymized for the analyses, and APA ethical guidelines were considered during the conduction of this study.

Patient profiles

Patient profiles were clustered with the same variables as the profiles constructed by Van der Veeken et al. (2017): historical and dynamic risk factors, derived from the Historical Clinical Future-30 items (HKT-30; Werkgroep risicotaxatie forensische psychiatrie [Risk Assessment Task Force in Forensic Psychiatry], 2002); psychopathology, described by DSM-IV-TR axis I diagnoses and severity of personality disorders (axis II); and index offenses (Van der Veeken et al., 2016). HKT-30 scores between the first six to 23 months after admission were included in the analyses. Previous data used in the profile study by Van der Veeken et al. (2017) were supplemented with new data, and diagnoses were adapted in the case of new diagnostic information.

Participants

Three hundred and twenty-one patients were included for Latent Class Analysis. Mean age of patients was 38.6 (*SD* = 10.54; range = 20-69) at the moment of their first risk assessment. Most patients were convicted of homicide (33%), followed by physical assault (19%), sexual offenses other than child sexual abuse (14%), child sexual abuse (13.4%), property offenses with violence (9.7%), arson (8.1%), property offenses (1.2%), or other (1.6%).

On axis I of the DSM-IV-TR (American Psychiatric Association [APA], 2000), patients were diagnosed with a psychotic disorder (32.7%), SUD (34.3%), paedophilia (7.2%), a pervasive developmental disorder (5.9%), attention deficit (hyperactivity) disorder (1.6%), a mood disorder (2.7%), sexual or physical abuse of child (2.1%), or sexual or physical abuse of adult (1.8%), or other (8.3%). Only 3.4% of the patients did not have an axis I disorder. On axis II of the DSM-IV-TR (APA, 2000), patients were diagnosed with antisocial personality disorder (34.6%), personality disorder not otherwise specified (NOS) (35.5%), borderline personality disorder (3.7%), narcissistic personality disorder (3.4%), mental disability (3.1%), or other (4.2%). Of all patients, 15.3% did not have an axis II diagnosis, or the diagnosis was deferred.

Measurements

Risk assessment

The Dutch Historical Clinical Future 30-items (HKT-30; Werkgroep risicotaxatie forensische psychiatrie [Risk Assessment Task Force in Forensic Psychiatry], 2002) is the most frequently used risk assessment instrument in the Netherlands. It is the predecessor of the revised HKT-R, validated nationwide, which should be used mandatory by all Dutch psychiatric centers since 2015 to annually report the likelihood of violent recidivism (Spreen, Brand, Ter Horst, & Bogaerts, 2014; Bogaerts, Spreen, Ter Horst, & Gerlsma, 2017). The HKT-30 is scored by one or two clinicians, based on behavioral observations, staff interviews, and legal and clinical reports.

The Historical domain includes 11 static risk factors, such as *history of school and work*, *history of substance use* and *previous treatment*. The Clinical domain contains 13 clinical factors, describing risk behavior observed during the year previous to the assessment, such as *impulsivity, empathy* and *social skills*. The Future domain consists of six factors, relating to situations outside the clinic which influence near future risk, such as *social network, daily activities*, and *material indicators* (such as housing). All items are scored on a five-point scale, ranging from zero: no risk (i.e., the patient has adequate skills or a supportive network) to four: highly problematic.

Previous studies have shown that the HKT-30 is predictive of future violent offending (Hildebrand, Hesper, Spreen, & Nijman, 2005). Interrater reliability with Spearman correlation coefficient could be assessed for the clinical and future items for 81 to 83 HKT-30 assessments apart from the item 'Crime responsibility' ($N = 65$). All items included in the Latent Class Analysis showed an $r_s > .60$, apart from the items 'Skills outside the institu-

tion', r_s = .58. The internal consistency assessed with Cronbach's alpha, is α = .68 for the historical scale, α = .77 for the clinical scale, and α =.79 for the future scale.

Psychopathology

All diagnoses were assessed by clinicians trained in using the DSM-IV-TR (APA, 2000). Axis I diagnoses were divided into pervasive developmental disorders, SUDs, psychotic disorders, paedophilia, other, or no disorder. The severity of personality disorders was measured according to the item 'personality disorder' as defined in the HKT-30. The last available score was used in the analyses. A score of zero implies 'no signs of personality traits'; a score of one indicates 'signs of pathological features'; a score of two indicates 'one or more personality disorders (PSDs) but no cluster B'; a score of three indicates 'one or more PSD with a cluster B PSD'; and, finally, a score of four indicates 'two or more PSD with two cluster B PSD or one cluster B PSD and a cluster A PSD'.

Offense type

Index offenses were retrieved from patient files and cover property offenses, property offenses with violence, arson, physical assault, homicide (including murder and attempted murder), sexual offenses other than child sexual abuse, child sexual abuse, or other.

Prospective treatment progress

Participants

Two hundred and forty patients were included in this study. All were male inpatients with a TBS-measure, meaning they were admitted to a forensic psychiatric center by order of the state with a minimum penalty of four years as they committed a crime caused by a psychiatric disorder (Van Marle, 2002). Mean age of patients was 40.47 (*SD* = 10.36, range = 22 – 72). Most patients were diagnosed with SUD (41.3%) on axis I of the DSM-IV-TR (APA, 2000), followed by schizophrenia (19.2%), paedophilia (8.3%), psychotic disorders other than schizophrenia (8.4%), pervasive developmental disorder (4.6%), mood disorder (4.2%), attention deficit hyperactivity disorder (2.1%), or other (9%). Only two point nine percent did not have an axis I disorder, or the diagnosis was deferred.

On axis II, most patients were diagnosed with a personality disorder not otherwise specified (NOS) (39.2%), followed by an antisocial personality disorder (30%), borderline (5.8%), narcissistic (4.6%), or other (4.5%); 15.9% were not diagnosed with an axis II disorder, or the diagnosis was deferred. Most patients had committed homicide (37.5%),

physical assault (16.3%), sexual offense other than child sexual abuse (13.8%), child sexual abuse (15.8%), a property offense with violence (7.9%), arson (6.3%), or other (2.5%).

Measurements

Instrument for Forensic Treatment Evaluation

The IFTE is an evaluation instrument that assesses problem behavior with nine items (such as *impulsivity, recent substance use,* and *antisocial behavior*), protective behavior with eight items (such as *problem insight* and *coping skills*) and resocialization skills with five items (such as *balanced daytime activities* and *social skills*). All items can be scored on a 17-point scale with five anchor points (Schuringa et al., 2014). A 17-point scale allows scoring between two anchor points or just above or below an anchor point (Schuringa et al., 2014). All items contain the 'Not enough information' (N.E.I.) option, which can be chosen if an assessor does not have enough information to evaluate an item. The IFTE is available in Dutch and has currently not been translated to other languages. The internal consistency assessed with Cronbach's alpha for the IFTEs in this study is α = .83 for the factor *protective behavior* (*N* = 23), α = .81 for the factor *problem behavior* (*N* = 143) and, α = .83 for the factor *resocialization skills* (*N* = 178). The items *medication use, psychotic symptoms, skills to prevent substance use, drug use, skills to prevent physically aggressive behavior* and *skills to prevent sexually deviant behavior* are optional if applicable to a patient and were, therefore, not assessed for all patients. Factor scores (resocialization skills, problem behavior, and protective behavior) used in this study, therefore, were calculated with the exclusion of these items. Interrater reliability and test-retest reliability were assessed in another study, with intra-class correlation coefficients, with a two-way randomized model for consistent measures and average measures. Interrater reliability varies from ICC = .50 to ICC = .89. Test-retest reliability varies from ICC = .57 to ICC = .92 (Van der Veeken, Bogaerts, & Lucieer, 2018).

Statistical analyses

Latent Class Analyses with risk factors, psychopathology and type of offense

Static and dynamic risk factors as displayed in Table 2, psychopathology (i.e., primary axis I diagnoses and severity of personality disorder), and offenses were loaded into latent gold 4.5 (Vermunt & Magidson, 2005) in order to confirm the four previously identified patient profiles (Van der Veeken et al., 2017). LCA was conducted with the HKT-30 risk factors

displayed in Table 2 as indicators, primary axis I diagnosis and severity of personality disorder as active covariates, and index offense as inactive covariate.

A lower Bayesian information criterion (BIC) indicates which model provides the best fit. The bootstrap P value should be >.05, and the entropy R squared and reduction of errors should indicate how well the model predicts class membership and should reach a value near one (Vermunt & Magidson, 2005). Apart from these values, the bivariate residuals between items should have a maximum value of two, to control for local independencies.

Prospective treatment progress

IFTEs were loaded into SPSS 23. With a repeated measures ANOVA test, treatment progress from measurement one to six was assessed for the whole group on factor level (i.e., *problem behavior, protective behavior,* and *resocialization skills*) and item level of the IFTE. To assess whether patients with high problem behavior, low protective behavior or low resocialization skills would show more progress than patients with high functioning scores, patients were divided into high- and low-functioning groups based on median scores for the three IFTE factors. Low-functioning means that a patient shows higher scores for problem behavior and lower scores on protective behavior and resocialization skills at the first measurement. High-functioning means that a patient shows lower scores on problem behavior and higher scores on protective behavior and resocialization skills. Then repeated measures ANOVAs were conducted for the high- and low-level groups on factor level.

Finally, a Mixed repeated measure ANOVA test was conducted to compare treatment progress for patients who had been in treatment for less than 12 months at the first IFTE assessment and patients who had already been in treatment for more than 12 months at the first IFTE assessment.

A *Z*-score between -2 and 2 for the Skewness and Kurtosis was maintained to control for the normality of the distribution for the IFTE factor and item scores and for the score differences between scores as we want to study treatment progress. Results were supplemented with a non-parametric test when the scores did not reach the set conditions for normal distribution. In the case of the violation of Mauchly's test of sphericity, the corrections provided by SPSS were consulted. For all repeated measures ANOVA the Greenhouse-Geisser or Huynh-Feldt corrections were used when Mauchly's test of sphericity was significant. The Greenhouse-Geisser correction was used when the value of Greenhouse-Geisser was below .75, in the case of a higher value, the Huynh-Feldt correction was used.

Results

Latent Class Analysis with risk factors, psychopathology and type of offense

Latent Class Analysis from three to six classes resulted in a best fitting model with four classes. Table 1 displays fit statistics for a three- to six-class solution and Table 2 displays included HKT-30 scores for the identified profiles. The BIC value was the lowest for the four-class model, indicating that the four-class solution was the best fitting model. However bivariate residuals displayed high values (range = .00 - 38.75). Even after correcting for high local depencies, bivariate residuals differ between 0.0 and 10.47, indicating that the profiles show a high variance within profiles. Therefore, the model did not display a proper fit, and therefore we could not assess treatment progress for differing treatment profiles within this study.

Table 1

Latent Class Analysis fit statistics

No. of classes	BIC (L^2)	Npar	L^2	*df*	*p**	Class error	Entropy	Reduction of errors (Lambda)
3	9420.0	120	10352.7	165	.19	.05	.88	.91
4	**9355.3**	**147**	**10135.3**	**138**	**.17**	**.06**	**.88**	**.91**
5	9375.1	174	10002.5	111	.11	.06	.89	.91
6	9426.9	201	9904.8	84	.07	.06	.90	.91

**p* after bootstrapping, BIC = Bayesian Information Criterion, Npar = Number of parameters, *df* = degrees of freedom

Table 2
Group HKT-30 means and standard deviations for the four profiles

	Class 1 (32%)		Class 2 (26%)		Class 3 (23%)		Class 4 (19%)	
Variables	Mean	(*S.E.*)	Mean	(*S.E.*)	Mean	(*S.E.*)	Mean	(*S.E.*)
Legal history	3.40	(.1)	3.23	(.1)	1.82	(.2)	2.00	(.2)
Violation of terms	3.60	(.1)	3.68	(.1)	1.22	(.2)	1.88	(.3)
Past care	3.41	(.1)	3.11	(.1)	1.96	(.2)	2.06	(.2)
History of school and work	3.51	(.1)	3.42	(.1)	1.99	(.2)	2.34	(.2)
Past substance use	3.61	(.1)	3.54	(.1)	1.70	(.2)	2.63	(.2)
Psychopathy	2.33	(.1)	2.23	(.2)	1.28	(.2)	1.05	(.2)
Problem insight	2.42	(.1)	3.37	(.1)	3.31	(.1)	2.11	(.1)
Impulsivity	2.10	(.1)	2.43	(.1)	1.21	(.2)	0.77	(.1)
Empathy	2.39	(.1)	3.45	(.1)	3.18	(.1)	1.76	(.1)
Hostility	1.98	(.1)	2.74	(.1)	1.62	(.1)	0.79	(.1)
Social and relational skills	2.15	(.1)	3.04	(.1)	2.68	(.1)	1.41	(.1)
Treatment attitude	1.51	(.1)	2.69	(.1)	2.27	(.2)	0.93	(.1)
Crime responsibility	2.08	(.1)	3.01	(.1)	2.76	(.1)	1.50	(.2)
Coping skills	2.44	(.1)	3.44	(.1)	2.98	(.1)	1.79	(.1)
Skills	2.58	(.1)	3.38	(.1)	2.74	(.1)	1.96	(.1)
Social network	2.67	(.1)	3.03	(.1)	2.65	(.1)	2.01	(.1)
Stressing circumstances	3.48	(.1)	3.92	(.04)	3.50	(.1)	2.49	(.1)

Prospective study on treatment progress for the entire patient group

The repeated measures ANOVA on factor level for the whole group of patients did not show a positive significant change over time for *protective behavior* (F (4.11) = 2.22, p = .065), *problem behavior* (F(4.17) = 1.34, p = .255) and *resocialization skills* (F (3.22) = .176, p = .92). Table 3 displays the mean factor scores over time. As the mean progress was low, we conducted a frequency analysis to assess the differences in IFTE scores between T0 and T5 for the whole group of patients. The frequency analyses showed that patients showed both progress and decline on IFTE factor scores for *problem behavior,* range = -28 - 39.66 (*SD* = 14.20), for *protective behavior*, range = -32.5 - 31 (*SD* = 12.51), and for *resocialization skills,* range = -33.25 - 23 (*SD* = 12.20).

For the factor *problem behavior*, the average time between two assessments was 5.39 months (range mean periods = 4.67 - 5.72, range *SD* = 1.50 - 2.34). For the factor *protective behavior*, the mean period between two assessments was 5.34 months (range mean periods = 4.68 - 5.68, range *SD* = 1.46 - 2.37).

For the factor *resocialization skills*, the mean period between two assessments was 5.14 months (range mean periods = 4.64 - 5.61, range *SD* = 1.05 - 2.15). For all factors, the Range of months between two assessments was three to 13 months.

The mean period between assessment one and six was 28.42 months for *protective behavior* (*SD* = 3.97, range = 18 - 37), 28.72 months for *problem behavior* (*SD* = 3.64, range = 20 - 37), and 27.54 months for *resocialization skills* (*SD* = 3.44, range = 20 – 33).

Repeated measures ANOVA on item level for the whole patient group for six measurements displayed a marginal positive significant effect for *coping skills* ($F(4.19) = 4.13$, $p = .002$), ($\chi^2(5) = 18.1$, $p = .003$)), and *Skills to prevent physical aggressive behavior* ($F(3.72) = 3.11$, $p = .02$) (see Table 3), $\chi^2(5) = 17.96$, $p = .003$). While not all score differences between assessments displayed a normal distribution, results where complemented with a Friedman's ANOVA. No significant change over time was found for the other items.

Table 3

Treatment progress measured with the IFTE for the whole group of patients on factor level and significant items: N, means, and standard deviations

	T0 Mean (*SD*)	T1 Mean (*SD*)	T2 Mean (*SD*)	T3 Mean (*SD*)	T4 Mean (*SD*)	T5 Mean (*SD*)
Problem behavior *N* = 58	39.52 (14.41)	39.16 (16.58)	39.96 (16.57)	41.48 (16.52)	37.59 (14.63)	38.07 (16.25)
Protective behavior *N* = 59	38.47 (10.84)	41.04 (13.14)	40.79 (11.92)	40.77 (11.11)	41.41 (11.76)	42.80 (11.44)
Resocialization skills *N* = 28	58.89 (17.08)	58.91(19.23)	58.08 (18.33)	59.31 (16.16)	59.53 (16.14)	59.61 (18.34)
Coping skills *N* = 70	7.91 (2.88)	8.61 (3.34)	8.52 (3.09)	8.45 (3.18)	9.07 (3.14)	9.19 (3.27)
Skills to prevent physical aggressive behavior *N* = 37	11.21 (2.59)	11.34 (3.48)	12.39 (3.07)	11.97 (3.05)	12.76 (3.10)	12.57 (3.81)

Prospective treatment progress for low- and high-functioning groups

To test treatment progress for patients with low-functioning and high-functioning scores on *problem behavior, protective behavior* and *resocialization skills*, patients were classified in a low- and high-scoring group per factor, namely low and high problem behavior, low and high protective behavior, and low and high resocialization skills. Median scores displayed a

cut-off value for *problem behavior* of 36 (range = 8 - 80.5, *M* = 37.79, *SD* = 16.13), a cut-off value of 39 (range = 4.5 - 68, *M* = 38.37, *SD* = 12.46) for *protective behavior*, and a cut-off value of 61.25 (range = 5 - 85, *M* = 57.10, *SD* = 15.73) for *resocialization skills*. A repeated measures ANOVA test was computed on factor level for both low- and high- scoring groups on each factor for five assessment moments (T0 – T4). Analyses for six measurements would lead to a lower N (*N* = 9 to 32) per factor. Table 4 displays mean scores.

The group with high problem behavior showed a strong significant progress, meaning an important decrease in problematic behavior ($F(4) = 6.41$, $p < .01$). The group with low problem behavior showed a significant increase in problematic behavior over time ($F(3.57) = 3.57$, $p = .01$). The group with low protective behavior displayed a significant positive progress over time ($F(3.52) = 8.09$, $p < .01$). The group with high protective behavior displayed a significant decrease over time ($F(2.99) = 3.77$, $p = .01$).

Considering resocialization skills, the group with low resocialization skills showed a significant progress in these skills ($F(4) = 3.32$, $p = .02$), and the group with high resocialization skills also displayed a significant decrease of resocialization skills ($F(2.60) = 2.93$, $p = .05$).

However, for low problem behavior (Difference T4 - T3 *Z* Kurtosis = 2.79; Difference T1 - T0 *Z* Kurtosis = 10.05, Difference T2 - T1 *Z* Skewness = 2.63; difference T1 - T0 = 10.05 and Difference T4 - T0 *Z* Skewness = 2.32), not all score differences between assessments were normally distributed. This also accounts for high protective behavior (Difference T 2 - T1 *Z* Skewness = -4.73, *Z* Kurtosis = 9.57, difference T4 - T0 = *Z* Skewness = -2.68 *Z* Kurtosis = 2.12) and low protective behavior (Difference T4 - T3 *Z* Kurtosis = 2.69), and for high resocialization skills (Difference T4-T3 *Z* Kurtosis = 3.83, Difference T 2 - T1 *Z* Skewness = -4.67, *Z* Kurtosis = 8.08, Difference T1 - T0 *Z* Skewness = -4.06, *Z* Kurtosis = 5.26 and Difference T4 - T0 *Z* Skewness = -3.47, *Z* Kurtosis = 2.42) and low resocialization skills (Difference T1 - T0 *Z* Skewness = -2.20, Difference T4 - T0 *Z* Kurtosis = 3.13). Therefore we ran a Friedman's ANOVA for low problem behavior, high problem behavior, low protective behavior, high protective behavior, low resocialization skills, and high resocialization skills to complement the results.

No significant change has been found for low problem behavior ($\chi^2(4) = 6.65$, $p = .16$), high protective behavior ($\chi^2(4) = 7.02$, $p = .14$), and high resocialization skills ($\chi^2(4) = 6.77$, $p = .15$) with Friedman's ANOVA. A significant change was found for high problem behavior, $\chi^2(4) = 20.33$, $p < .001$, low protective skills $\chi^2(4) = 18.40$, $p = .001$ and low resocialization skills $\chi^2(4) = 10.88$, $p = .03$.

Table 4

Mean scores and standard deviations at T0 – T4 for high and low factor scores

Factor	T0 Mean (*SD*)	T1 Mean (*SD*)	T2 Mean (*SD*)	T3 Mean (*SD*)	T4 Mean (*SD*)
High problem behavior *N* = 42	50.06 (11.88)	48.39 (14.28)	45.81 (15.42)	46.67 (14.44)	41.55 (13.73)
Low problem behavior *N* = 41	25.96 (6.19)	27.54 (11.39)	31.77 (13.90)	31.37 (13.21)	29.47 (12.34)
Low protective behavior *N* = 41	28.89 (7.82)	31.84 (11.73)	33.67 (12.21)	37.15 (12.02)	36.24 (13.43)
High protective behavior *N* = 52	48.51 (5.89)	48.45 (8.44)	46.03 (8.95)	44.46 (10.50)	46.08 (9.70)
Low resocialization skills *N* = 16	40.59 (14.82)	41.15 (18.58)	43.69 (21.95)	45.45 (16.22)	49.25 (15.64)
High resocialization skills *N* = 30	69.39 (5.01)	69.67 (8.85)	65.79 (11.76)	64.53 (13.84)	65.77 (12.15)

Prospective treatment progress at the start of treatment

To examine if patients in their first 12 months of treatment showed more progress than patients who had been in treatment for more than 12 months at the first IFTE assessment, we conducted a repeated measures ANOVA test controlled for date of admission. Four assessments were included in the analysis because a six-assessment analysis (*N* = 5 to 47 per group) would yield a smaller sample than a four-assessment analysis (*N* = 16 to 85 per group). Repeated measures ANOVA showed a significant within subjects interaction effect for date of admission and the factor *resocialization skills* (F (3.71) = 2.47, p = .02), other factors did not show a significant interaction effect, p > .05. However the assumption of normality was violated for the factor *resocialisation skills* (Z Skewness = -3.55 - 5.22). Results indicated a better functioning score for the patient group that had received more years of treatment at the first IFTE assessment. Therefore, we also conducted an independent samples T-test to compare patients with a first IFTE assessment in the first year of treatment with patients who had been in treatment for a longer period at the first IFTE assessment (see Table 5).

Results displayed significantly higher *protective behavior* scores (t(218) = 3.48, p = .001) and *resocialization skills* scores (t(176) = 2.56, p = .01) at T0 for the patient group with a first assessment a year or longer after admission than the group with a first assessment within a year after admission (see Table 5). Mann Whitney test also displayed significant difference for resocialisation skills (Z Skewness = -4.89, U = 2406.5, p = .01). Even though patients who had been in treatment longer did seem to display less problematic behavior, the difference was not significant (t(212) = -.83, p = .41).

Analyses on item level controlled for date of admission for five assessments displayed no significant effects for time with the repeated measures ANOVA, apart from the item *antisocial associates* (F(3.85) = 31.1, p = .001). However, an independent samples T test was again conducted to assess differences between the two patient groups at T0.

Significant differences were found at T0 between the patient groups who had been admitted for a period of 12 months or shorter and those who had been admitted for a longer period at first IFTE assessment, for *problem insight* ($t(139.15) = 2.52$, $p = .01$), *cooperation with treatment* ($t(142.83) = 2.92$, $p = .004$ (Z Skewness = -2.38, Z Kurtosis = -2.49, $U = 5049$, $p = .005$)), *responsibility for the crime* ($t(223) = 2.82$, $p = .005$ (Z Skewness = -2.15, $U = 4481.5$, $p = .009$)), *coping skills* ($t(148.04) = 3.25$, $p = .001$), *balanced daytime activities* ($t(142.53) = 2.70$, $p = .008$ (Z Skewness = -4.67, $U = 5406.5$, $p = .031$)), *labor skills* ($t(93.76) = 2.59$), $p = .011$ (Z Skewness = -5.63, $U = 3263$, $p = .020$)), *skills to prevent substance use* ($t(65.85) = 3.23$, $p = .002$, Z Skewness = -3.60, $U = 1772.5$, $p = .002$)). Independent samples t-test also displayed a significant difference for *skills to prevent physically aggressive behavior,* however Mann Whitney analysis was not significant ($t(80.68) = 2.31$, $p = .024$, (Z Skewness = -5.54, Z Kurtosis = 2.79, $U = 2915.5$, $p = .093$)). Means are displayed in Table 5.

Table 5

Means and standard deviations at T0 for first IFTE assessment within a year after admission and assessment at 12 months or longer after admission: N, means, standard deviations, and mean period between admission and T0

	N	T0 within 12 months after admission *M* = 3.41 (3.03)* *M (SD)*	*N*	T0 more than 12 months after admission *M* = 46.81 (26.32)* *M (SD)*
Problem behavior	65	39.18 (17.91)	149	37.19 (15.32)
Protective behavior	73	34.32 (13.25)	147	40.37 (11.58)
Resocialization skills	49	52.28 (17.00)	129	58.93 (14.89)
Problem insight	82	8.45 (4.13)	153	9.78 (3.36)
Cooperation with treatment	85	9.02 (4.99)	152	10.86 (3.93)
Responsibility for the crime	77	9.35 (3.91)	148	10.85 (3.70)
Coping skills	85	7.58 (3.40)	153	8.99 (2.81)
Balanced daytime activities	85	9.52 (4.97)	153	11.21 (3.92)
Labor skills	61	10.28 (5.84)	135	12.46 (4.52)
Skills to prevent substance use	46	9.18 (4.81)	111	11.72 (3.52)
Skills to prevent physically aggressive behavior	58	10.99 (4.32)	119	12.43 (2.78)

*Mean period in months of assessment after admission

Discussion

The goal of this study was to assess if the IFTE could assess treatment progress in a Dutch forensic psychiatric population. Latent Class Analysis was conducted with historical and dynamic risk factors, psychopathology, and type of offense in order to identify patient profiles.

Repeated measures ANOVA tests were conducted to assess treatment progress for the whole group of patients and to assess whether this progress differed between patients with high problem behavior scores, low resocialization skills scores, or low protective behavior scores, and if treatment progress differed for length of treatment period. The four emerging profiles − the antisocial profile, the mixed profile, the maladaptive disordered profile, and the psychotic first offender profile − are comparable to the profiles found in the previous study by Van der Veeken et al. (2017), however bivariate residuals indicated that the model did not fit, as the model displayed high variability within profiles. This implies that these profiles and the profiles previously identified include high variability of patient characteristics within the four profiles. These results could indicate that it could be very difficult to identify more homogeineous groups within the heterogeinic forensic psychiatric population. Even though we can identify profiles based on diagnosis and offense, as found by Van Nieuwenhuizen et al. (2011), it is likely that patients within these groups also differ on level of risk and protective factors.

A repeated measures ANOVA was conducted for six measurement periods for the whole group of patients. The results in this study did not display significant progress over time for the factors *protective behavior, problem behavior* and *resocialization skills.* Significant differences in IFTE scores over time have been found for the protective items *coping skills* and *skills to prevent physically aggressive behavior.*

This is not in accordance with previous results found in studies assessing progress with risk assessment tools. Olsson, Strand, Kristiansen, Sjöling, and Asplund (2013), for example, found a small but significant change in clinical Historical Clinical Risk- 20 items (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) items for lack of insight, negative attitudes, active symptoms of major illness, and unresponsive treatment for a period of nine months and for all clinical items in a period of 43 months. A study by De Jonge et al. (2009) found a change in risk factors over time when assessed with the HKT-30 in a Dutch forensic population.

However, the IFTE is designed to assess individual progress, stagnation, and deterioration. The 17-point scale is possibly so sensitive to change that it detects not only progress but also decline. As the frequency analysis showed, there are both high decreases and in-

creases in IFTE scores, leading to a mean change near zero. This makes it difficult to assess treatment progress for a heterogeneous group of patients. The start of IFTE assessment may also have influenced the non-significant differences. IFTE assessments started in 2011 and 2012, regardless of the length of time patients had already been admitted and the duration of treatment patients had already received.

Further, a repeated measures ANOVA was conducted at five assessment periods for low- and high-functioning groups per factor, based on the median scores. As expected, patients with lower functioning scores showed significant progress on all three factors which is in line with expectations based on the risk principle of the RNR model (Andrews & Dowden, 2006). The treatment effect would be larger with higher risk patients (Polaschek, 2011) or, in this case, patients with high problem behavior, low protective behavior, or low resocialization skills. Polaschek (2011) found similar results, in which high risk offenders would benefit more from treatment than medium risk offenders, but this was based on recidivism outcomes. Lewis, Olver, and Wong (2012) also found a significant change with the violence risk scale (VRS; Harris, Rice, & Quinsey, 1993) in a forensic population scoring above the normative sample.

This seems to imply that higher risk patients have more progress to gain, as indicated by the RNR model (Andrews & Dowden, 2006), and benefit from treatment. However, results for the resocialization factor should be interpreted with care as these analyses include a low number of patients. Even more, repeated measures ANOVA for patients with higher-functioning scores on *protective behavior, problem behavior* or *resocialization skills* also showed a significant decrease in protective behavior and an increase in problem behavior. This could imply a certain amount of 'regression' to the mean, but the decrease in protective behavior seemed to be lower than the increase in protective behavior for the low-functioning group. The same accounts for the low problematic groups and high resocialization groups. Even more, the non-parametric assessment did not display significant change for patients with 'higher' functioning scores and low problem behavior.

There may also be several other 'confounding' factors that influence the found trend of decrease in well-functioning behavior. In the recent past, for example, the duration of a TBS-measure has risen sharply to approximately nine years (Nagtegaal, Van der Horst, & Schönberger, 2011). The question is if it is beneficial to treat patients for such a long time in a clinical setting, or if this might possibly counteract treatment effects. Hildebrand and De Ruiter (2012), for example, found that patients can 'become more interpersonally aggressive, competitive exploitive, dominant and assertive during treatment.' Our results, however, do not answer this question, and future studies will have to look into this effect.

Finally, progress at four measurement periods was assessed for patients who had been in treatment for less than 12 months at the first assessment and patients who had been in treatment for a longer period at their first assessment. In contrast with the study by Nijman et al. (2004), we did not find a difference in progress between these two groups for all factors: patients with a first assessment within 12 months after admission did not show more progress than patients who had been in treatment for a longer period for the factor *problem behavior* and *protective behavior*. Results however indicated an interaction effect for the factor *resocialization skills*, however these results should be interpreted with care as the assumption of normality was violated and the group concerns a rather low N. Also, results indicated that patients with IFTE assessments in the first 12 months of treatment showed less protective behavior and fewer resocialization skills than patients who had been in treatment for a longer period at the moment of their first assessment.

Even though these results should be interpreted with care, they could imply that patients who were in treatment longer, developed more skills than patients who are just starting their treatment. However, we did not control for diffusing factors that could possibly influence a difference in scores. The number of patients admitted to a forensic psychiatric hospital with a TBS-measure has been decreasing for several years now; patients who would have been committed to a forensic psychiatric center in the past would now perhaps be committed to a penitentiary institution.

Limitations

A limitation of this study is that not all IFTE assessments were performed at the start of treatment for all patients. ROM was implemented for the whole group of patients in both clinics in 2011 and 2012 (Van der Veeken, Bogaerts, & Lucieer, 2012), including patients who had already been in treatment for some time. The number of patients with an assessment at the start of treatment was considerably smaller. Also, it would have been better if the groups defined in this study consisted of larger samples.

Especially the analyses comparing different groups of patients include rather small samples and therefore it is important that future studies replicate the analyses. Larger samples would lead to more robust analyses. All IFTE assessments, moreover, were linked to multidisciplinary treatment evaluations. These evaluations were held once every six months at the start of this study, and their frequency was increased to once every four months during this study. Because of the dynamics of the clinical setting, not all evaluations were held exactly within this time period, as treatment evaluations might have been postponed or brought forward due to special circumstances. This led to different time periods between assessments.

While data collection occurred between September 2011 and March 2015, it is also known that a TBS-measure and internal treatment lasts longer than four years (Van Nieuwenhuizen et al., 2011; Nagtegaal et al., 2011). This means that, together with the low number of patients with start-of-treatment assessments, we have hardly any data of patients both at the start and at the end of treatment. The fact that forensic psychiatric treatment takes a long time possibly influences our results. We decided not to score the IFTE retrospectively as the IFTE is an observational tool used by multiple raters, and scoring based on file information would provide a different type of assessment than scoring based on the IFTE instructions.

At the time of this study, no translations are known to the authors. This means that the capabilities of the IFTE have only been studied within Dutch FPCs. Therefore, results cannot be generalized to forensic settings in other countries. Future studies should focus on the use of the IFTE in different settings, other than an FPC, and in countries, other than the Netherlands.

Clinical implications

Results demonstrate that the IFTE indicates more progress for patients who show high problem behavior, low protective behavior, and/or low resocialization skills. With due caution, this implies that patients who may be considered high risk can gain more treatment progress. Even more higher problematic scores or lower protective or resocialization scores, could be indicative for treatment goals at the start of treatment. Treatment progress is defined according to the RNR model, and the IFTE assesses problematic and protective behavior and resocialization skills derived from recidivism-related risk assessments. As risk items are related to recidivism, it can be useful to assess change on these factors. Lewis et al. (2012) found that a reduction in risk factors of the VRS is related to less recidivism, which implies that it makes sense to assess treatment progress on these factors. Hanson, Harris, Scott, and Helmus (2007) found different results not supporting this implication, but this study was conducted with sexual risk assessment tools in a sexually delinquent population.

The results show limited progress for the whole patient group, possibly caused by the limited duration of the study relative to the treatment time in a Dutch forensic psychiatric center and the high inter-patient variation in progress and decline. Hildebrand and De Ruiter (2012) observed that treatment progress may possibly be gained at a slower pace, considering the long treatment periods. However, the individual IFTE reports used in treatment, do assess significant change for individual patients, with the N = 1 analysis (Spreen, Timmerman, Ter Horst, & Schuringa, 2010). In this way, it is possible to assess clinically relevant treatment progress, and the IFTE may serve the purpose of routine outcome monitoring, to determine the appropriate treatment plan and the need for further treatment, and

to assess whether the proposed treatment progress has been attained or not for an individual patient (Barwick & Urajnik, 2014).

Future research

Future research should focus on treatment progress from the start till the end of forensic psychiatric treatment. Even though this would take more years, we could gain more insight into treatment processes in this way. ROM measurements could also be used in a randomized controlled trial design to assess progress for different treatment modules (Nagtegaal et al., 2011). As Nugter and Buwalda (2012) suggest, ROM could be used to gain insight into separate treatment elements in order to evaluate how separate elements contribute to the whole treatment. As mentioned in the method section, therapists have access to the IFTE reports. With the access to these reports, therapists can assess treatment goals related to risk management and resocialization. We have not evaluated if the progress found is due to certain treatment modules, therapies, or therapist related characteristics, such as responsivity (e.g., Kunst, Bogaerts, & Winkel, 2009). At the moment the effect of feedback with the IFTE to patients on treatment progress is being studied in a sample in an FPC within the Netherlands (Muchall & Bogaerts, nd). Future research would have to look into treatment progress assessed with the IFTE in relation to different treatment modules or therapies. This could also include progress during different leave modules, as these are a crucial part of resocialization. IFTE reports are designed particularly for individual treatment evaluations, which allows the IFTE to be used for single case designs, which are strong in assessing the effect of treatment (Nagtegaal et al., 2011). These results would not be generalizable, but when repeated often, this could give us insight into treatment processes within a highly heterogeneous population.

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders – DSM-IV-TR.* Arlington, VA: American Psychiatric Association.

Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A Meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology, 50,* 88 – 100. doi: 0.1177/0306624X05282556

Barwick, M. A., & Urajnik, D. J. (2014). Training and maintaining system-wide reliability in outcome management. *Journal of Child and Family Studies, 23,* 85 – 94. doi: 10.1007/s10826-012-9694-x

Baumeister, R. F., & Vohs, K. D. (2007). Self-regulation, ego depletion, and motivation. *Social and Personality Psychology Compass, 1,* 115-128. doi: 10.1111/j/1751-9004.2007.00001.x

Bogaerts, S., & Spreen, M. (2011). *Persoonlijkheidspathologie, slachtofferschap vóór het 18e levensjaar, huiselijk geweld en delict informatie in een klinische groep terbeschikkinggestelden* [Personality pathology, victimization before the age of 18, domestic violence and crime information in a clinical group of inpatients]. Den Haag, The Netherlands: Onderzoek in opdracht van het Ministerie van Justitie en Veiligheid, WODC.

Bogaerts, S., Spreen, M., Ter Horst, P., & Gerlsma, C. (2017). Predictive validity of the HKT- R risk assessment tool: Two and 5-year violent recidivism in a nationwide sample of Dutch forensic psychiatric patients. *International Journal of Offender Therapy and Comparative Criminology,* 1 - 12. doi: 10.1177/0306624X17717128

Bogaerts, S., Vanheule, S., & DeClercq, F. (2005). Recalled parental bonding, adult attachment style, and personality disorders in child molesters: a comparative study. *The Journal of Forensic Psychiatry & Psychology, 16,* 445-458. doi:10.1080/ 14789940500094524

Bogaerts, S., Vervaeke, G., & Goethals J. (2004). A Comparison of Relational Attitude and Personality Disorders in the Explanation of Child Molestation. Sexual Abuse: A journal of Research and Treatment, 16: 37-47. doi: 10.1023/B:SEBU.0000006283. 30265.f0

Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and treatment (User Report No. 2007-06). Ottawa, Ontario: Public Safety Canada. Retrieved from https://cpoc.memberclicks.netassets/Realignment /risk _need_2007-06_e.pdf

De Boer, J., Whyte, S., & Maden, T. (2008). Compulsory treatment of dangerous offenders with severe personality disorders: a comparison of the English DSPD and Dutch TBS systems. *The Journal of Forensic Psychiatry & Psychology, 19,* 148 - 163. doi: 10.1080/14789940701830726

De Jonge, E., Nijman, H. L. I., & Lammers, S. M. M. (2009). Gedragsverandering tijdens tbs-behandeling: een multicenteronderzoek [Behavioural changes during forensic psychiatric treatment: a multicenter study]. *Tijdschrift voor Psychiatrie, 51,* 205 – 215. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/articles_ 2764pdf.pdf

Ellwood, P. M. (1988). Outcomes management. A technology of patient experience. *The New England Journal of Medicine, 23*, 1549-56. doi: 10.1056/NEJM198806093182327

Hanson, R. K., Harris, A. J. R., Scott, T., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The dynamic supervision project* (User Report 2007-05). Ottawa: Public Safety and Emergency Preparedness Canada.

Hare, R. D. (1991) The psychopathy checklist revised. Multi-Health Systems, Toronto.

Hare, R. D., & Neumann, C. S. (2005). Structural Models of Psychopathy. *Current Psychiatry Reports, 7;* 57-64. doi:10.1007/s11920-005-0026-3

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism in mentaly disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20,* 315 – 335. Retrieved from: http://psycserver.psyc.queensu.ca/quinsey/pdf/Prediction/Harrisetal1993CJB-Recidivism.pdf

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. Expertise Centrum Forensische Psychiatrie, Utrecht: 2005.

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. doi: 10.1016/j.ijlp.2012.04.001

Kunst, M., Bogaerts, S., & Winkel, F. (2009). Peer and inmate aggression, type D-personality and posttraumatic stress among Dutch prison workers. *Stress and Health: Journal of the International Society for the Investigation of Stress,* 25, 387-395. doi: 10.1002/smi.1247

Lewis, K., Olver, M. E., & Wong, S. C. P. (2012). 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. doi: 10.1177/1073191112441242.

Muchall S., & Bogaerts, S. *Routine Outcome Monitoring (ROM) Feedback en behandelmotivatie bij forensisch psychiatrische patiënten met een hoog en laag risicoprofiel.* nd.

Nagtegaal, M. H., Van der Horst, R. P., & Schönberger, H. J. M. (2011). *Inzicht in de verblijfsduur van tbs-gestelden - Cijfers en mogelijke verklaringen.* Meppel:Boom Juridische uitgevers.

Nijman, H., De Kruyk, C., & Van Nieuwenhuizen, C. (2004). Behavioural changes during forensic psychiatric (TBS) treatment in the Netherlands. *International Journal of Law and Psychiatry, 27*, 79 – 85. doi: 10.1016/j.ijlp.2003.12.001

Nugter, M. A., & Buwalda, V. J. A. (2012). Achtergronden en gebruiksmogelijkheden van ROM in de GGZ [Background and possible use of ROM in mental health care]. *Tijdschrift voor Psychiatrie, 54,* 111 – 120. Retrieved from: http://www. tijdschriftvoorpsychiatrie.nl/assets/articles/TvP12-02_p111-120.pdf

Olsson, H., Strand, S., Kristiansen, L., Sjöling, M., & Asplund, K. (2013). Decreased risk for violence in patients admitted to forensic care, measured with the HCR-20. *Archives of Psychiatric Nursing, 27,* 191 – 197. doi: 10.1016/j.apnu.2013.03.004

Polaschek, D. L. L. (2011). High-intensity rehabilitation for violent offenders in New Zealand: Reconviction outcomes for high- and medium-risk prisoners. *Journal of interpersonal Violence, 26,* 664 - 682. doi: 10.1177/0886260510365854.

Rothman, A. J., Baldwin, A. S., Hertel, A. W., & Fuglestad, P. T. (2011). Self-regulation and behaviour change: Disentangling behavioural initiation and behavioural maintenance. Handbook of Self-regulation: Research, theory, and applications. Ed. Vohs, K. D. & Baumeiser, R. F., The Guilford Press. New York.

Schuringa, E., Heininga, V. E., Spreen, M., & Bogaerts, S. (2018). Concurrent and predictive validity of the instrument for forensic treatment evaluation: From risk assessment to routine, multidisciplinary treatment evaluation. *International Journal of Offender Therapy and Comparative Criminology, 62,* 1281-1299. doi:10.1177/ 0306624X1667 6100

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,* 127-144. doi: 10.1080/1522893 2.2014.897536

Spreen, M., Brand, E., Ter Horst, P., & Bogaerts, S (2014). *Handleiding en Methodologische Verantwoording HKT-R, Historisch, Klinische en Toekomstige – Revisie* [Manual and methodological justification HKT-R Historical, Clinical and Future-Revised]. Dr. van Mesdag kliniek: Groningen.

Spreen, M., Timmerman, M. E., Ter Horst, P., & Schuringa, E. (2010), Formalizing clinical decisions in individual treatments: Some first steps. *Journal of Forensic Psychology Practice, 10*, 285 – 299. doi. 10.1080/15228932.2010.481233

Van der Veeken, F. C. A., Bogaerts, S., & Lucieer, J. (2012). Routine outcome monitoring in forensic psychiatry: FPC 2landen and FPC De Kijvelanden. In T. I. Oei & M. S. Groenhuijsen (Eds.), *Progression in forensic psychiatry* (pp. 355-375). Alphen aan den Rijn, The Netherlands: Kluwer.

Van der Veeken, F. C. A., Bogaerts, S., & Lucieer, J. (2017). Patient profiles in Dutch forensic psychiatry, based on risk factors, pathology and offense. *International Journal of Offender Therapy and Comparative Criminology, 61*, 1369-1391. doi:10.1177 /0306624X15619636

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 Practice.* doi: 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, 12.* doi:10.1371/journal.pone.0160787 Van Marle, H. J. C. (2002). The Dutch Entrustment Act (TBS): Its principles and innovations. *International Journal of Forensic Mental Health, 1*, 82 - 92. doi: 10.1080/14999013.2002.10471163

Van Nieuwenhuizen, C. H., Bogaerts, S., De Ruiter, E. A. W., Bongers, I. L., Coppens, M., & Meijers, R. A. A. C. tbs-behandeling geprofileerd; een gestructureerde casussenanalyse [*ProfilingTBS-treatment: A structured cases analysis]*. Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) Ministerie van Veiligheid en Justitie. 2011. Retrieved from http://www.wodc.nl/onderzoeks database/voor studie-onderzoek-tbs.aspx

Vermunt, J. K., & Magidson, J. (2005). Technical guide for Latent Gold 4.0: Basic and Advanced. Belmont Massachusetts: Statistical Innovations Inc.

Wagner, D., & Heatherton, T. F. (2015). Self-regulation and its failure: The seven deadly threats to self-regulation. In M. Mikulincer & P. R. Shaver (Eds.), *Handbook of personality and social psychology vol 1: Attitudes and social cognition (*pp. 805-42) doi: 10.1037/1434-026

Ward, T., Melser, J., & Yates, P. (2007). Reconstructing the risk-need-responsivity model: A theoretical elaboration and evaluation. *Agression and Violent Behaviour, 12,* 208 - 228. doi: 10.1016/j.avb.2006.07.001

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

Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Task Force in Forensic Psychiatry]. (2002). *Handleiding HKT-30 versie 2. Risicotaxatie in de forensische psychiatrie* [Guide HKT-30 version 2. Risk Assessment in Forensic Psychiatry]. Den Haag, The Netherlands: Ministerie van Justitie, Dienst Justitiële Inrichtingen.

Yiend, J., Chambers, C., Burns, T., Doll, H., Fazel, S., Kaur, A., ... Fitzpatrick, R. (2011). Outcome measurement in forensic mental health research: An evaluation. *Psychology, Crime and Law, 17,* 277-292. doi: 10.1080/10683160903203938

Yiend, J., Freestone, M., Vazauqes – Montes, M., Holland, J., & Burns (2013). The clinical profile of high-risk mentally disordered offenders. *Social Psychiatry and Psychiatric Epidemiology, 48*, 1169 – 76. doi: 10.100

CHAPTER SEVEN

General Discussion

This thesis focuses on routine outcome monitoring (ROM) conducted with the Instrument for Forensic Treatment Evaluation (IFTE) in two Dutch forensic psychiatric centers (FPCs), where patients are admitted involuntarily by order of the state after committing a crime. This thesis aims to examine if the IFTE, which consists of 22 items and 3 factors, is suitable for ROM in the clinical forensic setting. The research questions concerned the psychometric values of the IFTE, together with its factorial structure and its predictive validity. More specifically, this thesis aims to study if the IFTE can aid clinical decision-making in treatment, and whether the IFTE can assess change on the important forensic treatment indicator of "risk," translated into problem behavior, protective behavior, and resocialization skills.

Forensic psychiatric treatment focuses on the rehabilitation of forensic patients and the prevention of future recidivism. Such treatment, therefore, is embedded in the Risk-Need-Responsivity model (RNR) model, which emphasizes that treatment should focus on patients' individual criminogenic needs and responsivity factors (Bonta & Andrews, 2007). In addition to the RNR model, the Good Lives Model (Ward & Gannon, 2006) is also used in the forensic field, but the emphasis is on the RNR model in this thesis. The IFTE is designed to evaluate the important treatment indicators related to the RNR model.

Forensic ROM, however, does not only concern treatment evaluation of risk factors and protective factors (Grann et al., 2005; Fitzpatrick et al., 2010). Mental health quality, quality of life, (social) functioning, and psychosocial adjustment are also important treatment outcomes next to recidivism (Fitzpatrick et al., 2010; Yiend et al., 2011; Shinkfield & Ogloff, 2014).

Chapter 2, therefore, describes ROM tools composed for social therapeutic, supportive, and mildly intellectually disabled patients. The tools include instruments that assess the indicators quality of life, symptomatic, and daily functioning, in addition to the use of the IFTE to monitor forensic treatment progress and the HKT-30 (Historisch Klinisch Toekomst- 30 items; Werkgroep Risicotaxatie Forensische Psychiatrie, 2002 [Historical Clinical Future - 30 items; Risk Assessment Task Force in Forensic Psychiatry]) and later the HKT-R (Historisch Klinisch Toekomst - revised; Spreen, Brand, Ter Horst, & Bogaerts, 2014 [Historical Clinical Future - revised]) to assess risk. The remaining chapters focus on the usability of the IFTE as a forensic ROM tool and the identification of patient profiles to provide insight into specific patient groups and differences in treatment progress.

Summary of findings

Forensic patient profiles

Because forensic psychiatry ministers to a very heterogeneous population, latent class analysis (LCA) was used in **chapter 3** to identify homogeneous patient profiles. For the development of patient profiles, three types of factors were used: diagnoses, type of offence, and static and dynamic risk factors. Next, LCA was computed with inclusion of these factors, and four patient profiles were identified (**chapter 3**): the antisocial profile, the mixed profile with multiple problems, the maladaptive disordered profile, and the psychotic first offender profile.

These patients profiles were reassessed in **chapter 6** to examine whether they differed in treatment and could give insight into differences in progress on the selected treatment indicators. As the results indicated that the model did not fit, however, we did not assess differences in treatment change between the four profiles. Even though the profiles and comparisons between profiles on PCL-R total, factor, and facet scores in **chapter 3** appeared to accord with results in previous studies, the results found in **chapter 6** implied that it could be difficult to assess "homogeneous" profiles. Profiles could still be heterogeneous. Possibly, too many factors were included in the analysis, or a different combination of factors could be included. However, the profiles in **chapter 3** remain important as they give insight into the dynamics that therapists must cope with in forensic psychiatric treatment. The variability in profiles makes it difficult to assess specific treatment change for patients on similar factors, as described in the profiles. Assessments of intra-individual change could give more insight into this matter, compared to group evaluations (O'Brien & Daffern, 2016).

The results also emphasize the importance of a patient-centered approach, as several inpatient factors can influence treatment outcomes (Sedgwick, Young, & Kumari, 2016). As also described by the RNR model, it is important always to consider the individual level of risk, needs, strengths, and responsivity factors that must be incorporated into treatment (Dowden, Antonowicz, & Andrews, 2003; Andrews & Dowden, 2006; Polaschek, 2012). Patient wards, for example, are allocated on the basis of required guidance or psychopathology (Van Nieuwenhuizen et al., 2011), and a patient-centered approach is important in these patient wards. The main guidelines in Dutch forensic psychiatry, for psychotic offenders, for example, underline this importance and value possibilities for adjusting treatment to patient-specific indicators (Van Erp, Van Vugt, Van der Veeken, Van Boxtel, & Van Rooijen, 2018).

Usability of the IFTE: ROM merits

As the IFTE is a relatively new instrument in forensic care, psychometric research was reported in **chapter 4.** The interrater reliability, test-retest reliability, and internal consistency

were assessed, together with an explorative factor analysis to test the clinically used factor structure as described by Schuringa, Spreen, and Bogaerts (2014). Agreement for the test-retest and interrater reliability was assessed with the intra-class correlation coefficient. As not all items showed a normal distribution, results were complemented with Spearman's correlation coefficient. Test-retest results were assessed for the first two ROM assessments and showed a substantial agreement between measurements for most items and a somewhat lower value for the item *psychotic symptoms*. The interrater reliability was assessed between raters from two different disciplines, where most items also showed substantial agreement, apart from the item *antisocial associates*. These results are confirmed by results of Schuringa et al. (2014), who found moderate to almost perfect interrater reliability and test-retest values.

Internal consistency assessment with Cronbach's alpha showed good values, though the factor analysis should be interpreted with care. Similar to the study by Schuringa et al. (2014), the results support the factor *resocialization skills,* but the explorative factor analysis in **chapter 4** indicated four factors, instead of three factors. The items *antisocial associates, skills to prevent drug use,* and *drug use* loaded on a fourth factor. The factors *problem behavior* and *protective behavior* were less enhanced, and cross-loadings were found for several items. In conclusion, therefore, the results in this thesis do not fully enhance the clinically used factor analysis.

Both in **chapter 4** and in the study by Schuringa et al. (2014), the factor *protective behavior* has been less supported by statistical results. Schuringa et al. (2014), however, suggest a clinically fitting solution. Psychotic symptoms, for example, show a higher factor loading on the *resocialization skills* factor in both studies. In treatment, psychotic symptoms can be considered problematic, and Schuringa et al. (2014), therefore, suggest placing this item in the *problematic behavior* factor. Based on the considerations made by Schuringa et al. (2014), we chose to evaluate the clinically used factor structure in this thesis, as we studied the IFTE as it is being used in treatment.

Other important merits that a ROM tool should possess are validity and the ability to assess change over time (Slade, Thornicroft, & Glover, 1999; Schoen & Derksen, 2011). A ROM tool should assess relevant treatment indicators to inform patients and clinicians about the treatment direction (decision-making). It is also important for a ROM instrument to have sufficient predictive values because negative interfering events, such as inpatient aggression or violation of urine drug screening, and important steps in treatment, such as leave modalities, can influence treatment outcome (**chapter 5**).

Therefore, the predictive validity for important positive and negative treatment outcomes was assessed with Receiver Operating Characteristics (ROC) analyses (**chapter 5**). Low problem behavior, high protective behavior, and high resocialization skills scores were

considered predictive of positive treatment outcomes, such as permitted leave modalities (guided, unguided, and transmural leave). Leave, a positive treatment outcome, is crucial for resocialization, and leave decisions are carefully weighed. Patients who received leave modalities showed better IFTE factor scores than patients who were not permitted to go on leave. However, IFTE scores were marginally predictive of leave modalities. A difficulty in the positive treatment outcome analyses in **chapter 5** is that it is not certain whether the current leave approvals were granted at the most efficient point in treatment. Descriptive research shows that leave more often started at a later time point in treatment in Dutch FPCs in the 2002-2009 period (Nagtegaal, Van der Horst, & Schönberger, 2011).

High problem behavior, low protective behavior, and low resocialization skills scores were considered predictive of violation of urine drug screenings and inpatient aggression. Inpatient aggression and violations of urine drug screenings are important outcomes that can impede treatment. IFTE item and factor scores were also mildly related to future drug violations, apart from the items *compliance with rules, skills to prevent substance use,* and *substance use.* ROC analyses for patients with a personality disorder (NOS) and patients with a personality disorder (NOS) with a co morbid substance use disorder showed a reasonable to moderate predictive validity for inpatient aggression, and Area Under the Curve (AUC) values seemed to improve for patients with a personality disorder and co-morbid substance use disorder. Lower scores on resocialization and higher scores on problem behavior items appeared to be particularly related to inpatient aggressive incidents, and results indicate that deficient coping skills are also an important consideration in treatment.

Schuringa, Heininga, Spreen, and Bogaerts (2018) found similar results for items related to future drug use, and they enhance the predictive validity of *problem behavior,* and more specifically *impulsivity, skills to prevent physically aggressive behavior, antisocial behavior,* and *hostility* in relation to future violence. The factor *problem behavior* even appears to be a significant predictor of short-term violence in different patient treatment groups (e.g., patients with a personality disorder, patients with a pervasive developmental disorder, patients who have committed a sexual offence, patients who cope with psychotic vulnerability, and mildly intellectually disabled patients; Schuringa, Spreen, & Bogaerts (2018[1])). Considering validity, the IFTE also has a moderate to strong correlation with corresponding HKT-30 items and the corresponding therapy measures, work attendance and illegal drug use (Schuringa et al., 2018). Together with the studies by Schuringa et al. (2018) and Schuringa et al. (2018[1]), the results in **chapter 5** indicate that the IFTE item and factor scores are related to future inpatient behavior. Even though predictive validity has not been assessed for future outpatient recidivism, the results tentatively indicate that the IFTE items and factors are related to the need principle of the RNR model.

The ability to assess treatment change is also an important ROM merit, and the assessed items and factors, therefore, need to be dynamic and therefore changeable. The IFTE items are especially designed to assess treatment change. Because the treatment of forensic patients often takes very long and change of behavior is slow, the IFTE was designed with a 17-point scale to assess treatment change. The ability to measure treatment change at four points was reported in **chapter 3**. Patients were divided into a low and high group for each of the three factors (problem behavior, protective behavior, and resocialization skills), and dependent sample T-tests were computed to assess the effect size in-between assessments. A small effect over time was found for patients with high problematic behavior scores and low resocialization skills; a small to medium effect was found for patients with high protective behavior scores; and a small to large effect was found for patients with low protective behavior, low problematic behavior, and high resocialization skills scores. This indicates that the IFTE can show change over time.

Chapter six focuses on measuring treatment change with the IFTE for the whole patient group, for patients who were in treatment for a period shorter than one year at the first assessment, compared to patients who were in treatment for a longer period at the first assessment, and for patients with high functioning and low functioning scores. Six measurement points could be included in the repeated measures ANOVA. No significant change was found, however, for the whole patient group. Looking at the distribution of score differences, we do notice that patients in the whole group showed both progress and deterioration.

As ROM assessments started in 2011 and 2012 for all patients, and Nijman, de Kruyk, and Van Nieuwenhuizen (2004) showed that patients show more progress in the first period of treatment, a comparison was made between patients who started ROM at the beginning of treatment and patients who had been in treatment for a longer period of time at the first ROM assessment. A repeated measures ANOVA with four measurement points showed that treatment progress did not differ for patients who had been in treatment shorter than 12 months at the first assessment compared to patients who had been in treatment for a longer period. Patients who had been in treatment for a longer period than 12 months at the first assessment did show significantly better protective behavior and more resocialization skills at that first assessment.

Finally, as the RNR literature states that high-risk patients could benefit more from treatment, treatment progress was also studied for high and low problematic scores on the three factors. Similar to the analyses in **chapter 3**, progress was assessed for high-scoring patients and low-scoring patients on problem behavior, protective behavior, and resocialization skills, this time for five measurement points. Both groups, again, showed change,

with patients with higher functioning scores showing deterioration on the three factors, and patients with low functioning scores showing positive progress on the three factors. Results were complemented with non-parametric analyses, which confirmed significant change in the intended direction for high problem behavior, low protective behavior, and low resocialization scores, showing positive progress. However, parametric results could not confirm significant deterioration for low problem behavior, high protective behavior, and high resocialization skills.

Chapters 3 and **6** show a trend with patients with more problematic scores gaining progress, and patients with high functioning scores showing regress. Patients who have been in treatment for a longer period appear to show better functioning scores than patients with a first assessment at the beginning of treatment, but our results could not confirm the results found by Nijman et al. (2004).

Several factors could influence the results found in **chapters 3** and **6**. Even though some studies do find progress (De Vries Robbé, De Vogel, Douglas, & Nijman, 2015; Morrisey, Beeley, & Milton, 2014; O'Brien & Daffern, 2016), other studies also indicate that treatment progress could be gained at a slow pace in forensic care or involuntary treatment (Hildebrand & de Ruiter, 2012; Kortrijk, 2013), possibly explaining why progress was not found for the whole group of patients and the whole group of patients with an assessment at the beginning of treatment. Even though, similar to our results, some studies found that high-risk patients appear to show more progress (Polaschek, 2012; Lewis, Olver, & Wong, 2012), a study by Hildebrand and De Ruiter (2012) also showed that patients can show increased rates of problem behavior during treatment.

As progress was not evaluated throughout the entire treatment period, trends in our study do give indications that treatment should be evaluated throughout the entire treatment to identify possible fluctuations and variations in treatment progress or, as some patients show a deterioration, to study the possibility of a "ceiling." The RNR model also states that treatment intensity should be altered to the level of risk. A higher treatment intensity, compared to the level of risk, could even lead to the opposite effect, namely that of aggravating problem behavior (Bonta & Andrews, 2007). Even though patients in a forensic psychiatric institution do show a substantial number of risk factors, it should be considered whether this is the case in the forensic clinical population.

Limitations

This study has been conducted in a clinical setting, which has the benefit that we studied the IFTE as it is actually used in clinical practice. Even though field studies have the benefit that we can study instruments as they are used, this also comes with some limitations. ROM assessments are conducted in preparation of the routine treatment evaluation discussions. Appointments for such patient discussions, however, have been rescheduled many times, and discussions may be postponed for various reasons, but they may also be brought forward if decisions have to be made at short notice. This meant that there were different time periods between assessments, which would not have been the case in a study design that was solely implemented for scientific research. In the course of this study, moreover, the frequency of treatment evaluation discussions was upgraded from every six months to every four months due to internal reconsiderations. IFTE assessments were also first assessed in Excel documents, and later in the electronic patient files, leading to a differing layout of the 22 items and the 17-point scale (Appendix A and B). This study has been conducted within two FPCs with patients with a tbs measure, results are not generalisable to other forensic psychiatric patients. Also, patients residing at FPC de Kijvelanden and FPC 2landen are all male, no women were included in this study. Therefore, results are not generalisable to female patients residing in an FPC.

As the ROM data in this thesis were primary treatment information and feedback reports that were provided to all treatment teams and were accessible to patients upon request, we could not design a randomized controlled trial to assess the use of ROM in treatment. Another limitation concerns the workload in forensic psychiatric settings, which is very high and led several therapists to observe that they had little time to fill out ROM instruments. Sometimes therapists had to assess the IFTE "in-between other tasks," or the IFTE was not assessed at all, or only by one therapist. As mentioned, a composite score is likely to be more reliable than a one-person score. Previous ROM studies also noted that it takes a long time to implement ROM and that a change in the evaluation cultures is required (Smith, & Baxendine, 2015).

ROM studies report the difficulty of implementing ROM and the skepticism of therapists, nationally and internationally (De Jong, 2012; Kortrijk, 2013; Buwalda, 2013; Roe, Gelkopf, Gornemann, Baloush-Kleinman, & Shadmi, 2015). Clinicians in several studies use ROM to a limited extent only (Boswell, Kraus, Miller, & Lambert, 2013; Slade et al., 1999; Tasma et al., 2016). Important barriers to using ROM are other priorities, time, and motivation (De Jong, 2012; Boswell et al., 2013). The clinicians' attitude towards ROM and feedback could also influence the use of ROM. Studies report that clinicians inclined to

rely on extrinsic feedback use ROM and its feedback more often than clinicians inclined to rely on intrinsic feedback (De Jong, 2012[1]; Buwalda, 2013). Studies also report that some clinicians do not see ROM as an addition to their clinical view or feel that ROM could impede their autonomy (Buwalda, 2013). These are all factors that have been brought forward by clinicians during this study, and they may possibly have impeded its results.

It is important that continuous attention is paid to the use, usability, and user-friendliness of ROM, and that time should be invested in training and informing clinicians. Administrative pressure, imposed by management, insurance companies, or ministries could also bias the results (Delespaul, 2015; Kortrijk, 2013), and it is highly important, therefore, to ensure that ROM benefits treatment and that clinicians experience these benefits so they will use it for their patients rather than look upon it as an "administrative obligation." Even though we tried to inform all therapists at the start of the ROM implementation and organized information meetings after that, many employees left the organization and new employees arrived during our investigation. Patients themselves also moved between wards, leading to shortened observation periods before their next assessment.

ROM assessments in this study were implemented for all patients regardless of their date of admission. Most patients had already received treatment. In addition, as mentioned, treatment for patients with a TBS order takes a long time. This meant that we could not assess treatment progress from the beginning till the end of treatment. The relatively short treatment period that was assessed and the inclusion of all participants also meant that we could not conclude whether the change assessed with the IFTE was due to treatment, or that other confounding factors were involved. Especially the trend found in the low functioning and high functioning patient groups needs further study to control for a possible "regression to the mean."

Even though we included all patients in both FPCs (except those residing in a ward for intellectually disabled patients) in **chapters 4, 5**, and **6**, the sample size in our study was rather small. Increasing the sample size can give greater power, and for the factor analysis (**chapter 4**) and latent class analysis (**chapters 3** and **6**) in particular, a higher number of participants would have been better. The same applies to the treatment evaluation studies (**chapter 6**). The Greenhouse-Geisser, Huyn-Feldt, and independent sample T-test corrections also had to be interpreted for some analyses (**chapters 5** and **6**), possibly due to the small sample size. As a higher number of participants gives better power, this could produce more robust analyses. This study was conducted in FPCs with patients remaining in treatment for a long time and under continuous guidance. The results of this thesis, therefore, are not generalizable to other forensic settings, such as forensic outpatient treatment.

As we wanted to assess the IFTE as it is used in clinical practice, the clinically used factor structure as studied by Schuringa et al. (2014) and based on clinical considerations was

used in the analyses throughout this thesis. The results in **chapter 4**, however, support the factor analysis only to a limited extent. A larger factor analysis study could conclude if the clinically used factor structure shows the best fit. However, this could also lead to different results as we found a four-factor solution. This could indicate that, with regard to factor scores, results found in this thesis should partly be reassessed.

Finally, this thesis focuses on the IFTE, but the use of other instruments in forensic psychiatry should also be assessed. It is of the utmost importance to assess the use of the Dynamic Risk Outcome Scales (DROS), the Manchester Assessment of Quality of Life (MANSA), and the Health of the Nations Outcomes Scales (HoNOS) as these have been prescribed for all forensic mental health facilities with a legal title (DJI/MVenJ, 2015).

Clinical implications

Our results indicate that the IFTE can be used in forensic psychiatric treatment. The multidisciplinary approach of the IFTE is important for clinical practice. As multiple disciplines assess the IFTE to gain a composite score, patients' behavior is assessed from multiple perspectives, which creates a more complete image without the burden of a consensus scoring method. The number of raters can contribute to scoring accuracy (Wu, Whiteside, & Neighbours, 2007). Wu, Whiteside, and Neighbours (2007) compared individual measures, composite scores, and consensus scores and found that both composite scores and consensus scores were more accurate than individual measures.

Because the IFTE was developed to assess change in treatment, a more dynamic scale was used: the 17-point scale (Schuringa et al., 2014). As previously mentioned throughout this thesis, the IFTE was designed to allow treatment monitoring in relation to the RNR principles. Our results indicate that the IFTE is indeed related to future behavior, as could be expected from a ROM tool that assesses RNR principles. This is essential in forensic psychiatric treatment, in which risk management and prevention are essential (Doyle & Logan, 2012; Schuringa et al., 2014). The assessed items are behaviors or symptoms that can be related to cognitions or psychiatric diagnoses (Harris, & Rice, 1997). These behaviors should be dealt with in each patient in treatment in order to alter behavior.

With the IFTE, therapists can assess on item and factor level whether treatment is evolving in the right direction **(chapter 2**; Appendix C and D), decide if treatment plans should be reconsidered, and monitor problem behavior, resocialization skills, and protective behavior. However, clinicians should be aware that AUC values in this thesis are not high enough to rely on IFTE scores alone when assessing chance for inpatient incidents.

As we mentioned before, forensic ROM does not only concern treatment evaluation of risk factors and protective factors (Grann et al., 2005; Fitzpatrick et al., 2010), but important forensic treatment outcomes next to recidivism also include mental health, quality of life, (social) functioning, and psychosocial adjustment (Fitzpatrick et al., 2010; Yiend et al., 2011; Shinkfield & Ogloff, 2014; Van Nieuwenhuizen & Nijman, 2009). The ROM instruments described in **chapter 2**, together with the IFTE can be used as plural information sources for evaluating forensic patients.

As ROM is a fairly new method in forensic psychiatric care, it is important to consider current developments. Currently, the use of the HoNOS, DROS, or MANSA are now obligatory in forensic psychiatry (Dienst Justitiële Inlichtingen/Ministerie van Justitie en Veiligheid [Office for Judicial Institutions/Ministry of Justice and Security] (DJI/MjenV), 2015; 2017) together with more specific instruments for sexual offenders and patients with a substance use disorder. For patients with a personality disorder, the Stip-5 (Semi-structured interview for Personality functioning DSM-5) is now recommended to assess treatment progress (Oudejans, Spits, Nugter, & Van Bokkem, 2015). Schel et al. (2017) advise the use of the Forensic Inpatient Quality of Life Questionnaire: short version (FQL-SV; Schel, Bouman, Vorstenbosch, & Bulten, 2016) for assessing quality of life in inpatient treatment.

For risk assessment of mildly intellectually disabled patients, Nijman, Didden, and Hesper (2017) advised the Adaptive Questionnaire for the Intellectually Disabled [Adaptieve Vragenlijst Verstandelijke Beperking (AVVB; Jonker, Kruisdijk, Goedhard, & Nijman, 2016)] and the Scale for Emotional Development [Schaal voor Emotionele Ontwikkeling voor mensen met een verstandelijke beperking-revised, (SEO-R²; Morisse & Dosen, 2016)] together with the DROS or HKT-R for risk management. This could inform ROM for this specific population as described in **chapter 2.**

The goal of forensic treatment in an FPC is to rehabilitate patients into society. As described in **chapter 3**, patients are mostly not merely diagnosed with a substance use disorder, personality disorder, or sexual disorder. Treatment indications are based on both criminogenic needs, responsivity, psychopathology, and offense type. The reduction of problem behavior and the improvement of protective behavior and resocialization skills are important measurable indicators in forensic psychiatry. As the choice of outcome is driven by that which is supposed to change (Yiend et al., 2011), risk and protective factors should be assessed at the base. It would be advisable to assess the ROM indicator of "risk" more often than once a year to assess this goal. The IFTE is designed to do so. Therefore, the IFTE, which assesses protective and problem behavior and resocialization skills and is related to future risk behavior, is at the heart of forensic ROM. It should be investigated if

the MANSA or the FQL-SV and other obligatory tools are appropriate in this setting to be used together with the IFTE.

Currently, the use of the 14 clinical items of the HKT-R to assess treatment change in forensic psychiatry is being studied by the Dr. S. van Mesdag Clinic, FPC de Woenselse Poort and FPC de Kijvelanden in cooperation with kwaliteit forensische zorg (KFZ, n.d. [quality forensic care]). On a routine basis, however, the IFTE could offer a more dynamic assessment, and Schuringa et al. (2018[1]) conclude that the IFTE can be used in treatment for different patient groups in forensic treatment. When applicable (as in substance use disorder, personality disorder, and sexual disorder), specific diagnostic instruments could be added to the IFTE assessments in combination, or when treatment focuses on the measured indicators. It is important to safeguard the size of a ROM set as a data set that is considered too large or takes too much time would inhibit its use (Lambert & Hawkins, 2004). The possibility to add ROM tools for a period during treatment could be a solution.

In conclusion, the IFTE can contribute to patient-specific RNR-based treatment evaluation, but to do so it is important that we improve our understanding of within-patient treatment change. Other ROM instruments could expand these assessments with other specific patient factors that could influence treatment or treatment outcomes, such as quality of life and diagnosis-specific ROM tools. Such assessments could contribute to inpatient decision-making and aid inpatient treatment evaluation.

Future studies

This thesis focused on the reliability, predictive validity, and the ability of the IFTE to assess change. Several other ROM merits should be further investigated. Future research may focus on the effectiveness of feedback compared to no feedback in forensic treatment. At the current moment, studies do not find a conclusive answer to whether ROM feedback is beneficiary in treatment (Lambert et al., 2003; Delgadillo et al., 2018), and several factors such as diagnoses and motivation should be considered in studies (Muchall & Bogaerts, n.d; De Jong, Segaar, Ingenhoven, Van Buschbach, & Timmerman, 2017). An RCT approach with two conditions can be chosen (patients who receive real feedback and patients who receive neutral feedback). In addition to group level research, it is also possible to opt for single case experimental designs (such as an ABA design), in which realistic and neutral feedback follow each other and alternate to examine changes in behavior over time.

For the benefit of the study of treatment evaluation, it is advised to study treatment progress for the whole course of treatment and during resocialization outside the clinical

setting. Currently, progress for the entire group of patients was not found in this thesis. By evaluating the whole treatment, it is possible to re-assess results for treatment as a whole and for trends found in this study, such as progress found in low functioning patient groups and deterioration found in high functioning patients. It is important to study how these trends develop over a longer period, to assess what these trends mean, and to establish what the effect is over time, including time after release from the forensic psychiatric center.

Treatment progress in this study was studied with a repeated measures ANOVA to assess to what level patients would show progress. Analyses with mixed model designs could study the different patterns of change during treatment (Krueger, 2004). Studies in psychiatry also often use the *index for clinical significance* (Hafkenscheid & Van Os, 2016) to assess whether change is of clinical value. In these analyses, change is related to scores of a "normal population," and a patient's final score would then fall in or near the area of the "normal population" (Hafkenscheid & Van Os, 2016). Considering the long treatment duration and the low number of patients with measurements from the start till the end of treatment, however, it would not be realistic to use this assessment. Therefore, we assessed whether the group of patients showed a significant change on the IFTE scores and what this change was. What is more, for risk assessment tools and the IFTE in particular, norm scores are not known to the authors. Even though we did not use the reliable change index (Jacobson, & Truax, 1991 in Hafkenscheid, & Van Os, 2016), it could be used in the future to assess the proportion of changed patients in the institution.

More importantly, future studies should assess whether the IFTE items and factor scores are predictive of future recidivism and especially if changes measured with the IFTE throughout treatment are predictive of future recidivism (Kroner & Yessine, 2013; Douglas & Skeem, 2005). As ROM has not been studied widely in forensic psychiatric care, we have only limited results on successful treatment progress. Many studies evaluate risk of recidivism after treatment with offense ratings after release (Coid, Hickey, Kahtan, Zhang, & Yang, 2007; Dolan & Doyle, 2000). Some studies have evaluated change during forensic psychiatric treatment with risk assessment tools in relation to recidivism (De Vries Robbé et al., 2015; Spreen, Brand, Ter Horst, & Bogaerts, 2014). The study by De Vries Robbé et al. (2015), however, evaluated change on pre- and post-treatment scores in relation to recidivism. Spreen et al. (2014) assessed change in relation to recidivism during important treatment phases (pre-treatment, start of treatment, unguided leave, probation, and unconditional leave). They found that recidivists and non-recidivists show the same change during treatment, but that the recidivists' baseline is higher and remains higher over time. Kroner and Yessine (2013) found that only change measured on antisocial associates was related to recidivism in a sample of offenders. It is important to assess whether changes

found with the IFTE are related to the most important treatment outcome, which is the prevention of recidivism and rehabilitation.

Future studies could also use latent growth analysis to assess what factors influence treatment progress. This could also benefit further study of the trend found in **chapter 6**. It is likely that more factors are related to different treatment outcomes in the forensic psychiatric patient population. Heap (2003), for example, compared discharged with undischarged patients and found that undischarged patients were younger at admission and showed more reductions in leave status and more incidents during treatment.

As mentioned in **chapter 5**, it is not certain if the current leave modules (guided, unguided, transmural, etc.) that are used in Dutch forensic psychiatry start at the proper moment in treatment. An evaluation of IFTE scores for patients who have successfully and unsuccessfully proceeded through treatment and leave modules without and with re-convictions could inform us about IFTE scores related to positive treatment outcomes. Current ROM feedback studies (De Jong, 2012; Lambert et al., 2003; Lambert, 2010) describe a kind of warning system in ROM reports, signaling patients that are not on track. With such "warning" systems, patients who are not progressing can be identified at an earlier point in treatment. This could also be studied in forensic psychiatric care.

Concluding remarks

As ROM has received much attention in mental healthcare over the past years, ROM in forensic mental healthcare has been less studied. International reviews imply that recidivism is the most important treatment outcome (Fitzpatrick et al., 2010; Yiend et al., 2011), and that forensic ROM should focus on this important forensic treatment outcome. Instruments assessing dynamic risk factors can be promising in forensic outcome monitoring (Fittzpatrick et al., 2010; Yiend et al., 2011; Goethals & Van Marle, 2012; Bouman & Bulten, 2016). This also corresponds with the need principle of the RNR model, describing that treatment should focus on criminogenic needs. The IFTE was developed to assess important forensic treatment needs in a more timely manner. The results discussed in this thesis indicate that the IFTE can be used in forensic psychiatric clinical care as the IFTE shows acceptable psychometric values and predictive validity. A trend displaying treatment change for high- and low-scoring patients was found, but, as mentioned, this trend requires further study.

The psychometric and predictive validity results of the IFTE do not appear to be inferior to other well-studied forensic tools. Considering the short assessment time, the IFTE

appears to be an appropriate ROM tool in forensic care. However, as mentioned, the ability to assess change should be further studied. The IFTE is not intended to replace current risk assessments, but more short-time measures could complement the use of risk assessment tools. The use of the IFTE also fits in with the important ROM treatment indicators prescribed by DJI/MVenJ: the systematic assessment of problem severity, the systematic assessment of risk, and the assessment of change in risk of recidivism (DJI/MVenJ, 2017).

In this way, the use of the IFTE and additional ROM instruments could aid treatment quality together with the consideration of other treatment indicators, such as continuity of care, recidivism after treatment, early termination of treatment, aggressive incidents, and patient satisfaction with treatment (DJI/MVenJ, 2017). Individual use of the IFTE and additional ROM instruments could enhance both the need and the responsivity principle and could enhance a more patient-centered approach in which treatment is fitted to an individual's needs (Maassen, Schrevel, Dedding, Broerse, & Regeer, 2017).

One of the current criticisms of forensic psychiatric care is that treatment takes too long. Long treatment periods involve high costs, and the question is if treatment duration corresponds to the proportionality principle (Nijman, Lammers, Vrinten, & Bulten, 2017). The costs of a TBS measure are high (approximately 1.5 million euros per treatment; Nagtegaal, Goethals, & Meynen, 2016), but treatment results in lower serious recidivism rates compared to the serious recidivism rates of ex-convicts (Nagtegaal et al., 2016). Studies should assess if forensic ROM assessments lead to the timely adjustment of treatment plans and risk management procedures, thus benefiting treatment duration, and if this change on ROM indicators is related to lower recidivism rates. Only then can ROM assessments contribute to shortening treatment and, hence, to reducing costs.

The use of ROM for benchmarking is currently the subject of debate. Though Barendregt claims that criticism is partly based on misconceptions of the term benchmarking (Barendregt, 2015), objections to benchmarking have become more pronounced, with people saying they do not mind ROM being used as a quality assessment tool but that they are against it being used as a benchmarking tool. This has led to a "Stop ROM" petition (https://www.stopbenchmark.com), interim proceedings, and a critical report in 2017 (Hafkenscheid & Van Os, 2018). Their main criticisms are that data are being shared without permission, that the shared data, even if pseudonymized, can still be traced back to an individual, that quality improvement, though promised, has not yet emerged, and that benchmarking is not scientific but time-consuming and expensive (https://www. stopbenchmark.com).

Hafkenscheid and Van Os (2018) say that we cannot simply assume that ROM feedback enhances treatment, nor that ROM measures treatment outcome. They conclude that

ROM in its current form should be abandoned and that professionals should build differentiating ROM systems that do justice to the complexity of clinical reality and that can be used to evaluate intended care (Hafkenscheid & Van Os, 2018). ROM was developed as a tool to enhance treatment effectiveness but is now often used to achieve imposed percentages (Schoevers & Beekman, 2017). Instruments and assessment points are often selected on the basis of national guidelines and mandatory percentages (Barendregt, 2015; Tiemens & Van Sonsbeek, 2017). Schoevers and Beekman observe that it befits our profession to evaluate our actions, and they emphasize the possibility to learn from outcomes. Substantive arguments should determine how ROM is used to benefit patient treatment in mental healthcare in the Netherlands (Schoevers & Beekman, 2017).

The IFTE and forensic ROM primarily serve to benefit treatment and to inform treatment teams and patients; they should not be made to serve management and policy purposes. Future studies should further examine the use of forensic ROM with the IFTE, because with more information on treatment prognosis, the use of feedback, and the relation with recidivism, the IFTE and forensic ROM could be used as a compass in treatment.

References

Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology, 50,* 88 – 99. doi: 10.1177/0306624X05282556 Barendregt, M. (2015). Benchmarken en andere functies van ROM: back to basics. [Benchmarking and other functions of ROM: back to basics]. *Tijdschrift voor Psychiatrie, 57,* 517-525.

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and treatment* (User Report No. 2007-06). Ottawa, Ontario: Public Safety Canada. Retrieved from: https://cpoc.memberclicks.net/assets/Realignment/risk_need_2007-06_e.pdf

Boswell, J. F., Kraus, D. R., Miller, S., & Lambert, M. J. (2013). Implementing routine outcome monitoring in clinical practice: Benefits, challenges and solutions. *Psychotherapy Research, 25,* 6 - 19. doi: 10.1080/10503307.2013.817696

Bouman, Y. H. A. & Bulten, B. H. (2016). ROMmen met risicotaxatie in de forensische psychiatrie [To rom with risk assessment in forensic psychiatry]. *PsyXpert, 4,* 25- 35

Buwalda, V. J. A. (2013). Routine outcome monitoring in dutch psychiatry: Measurement, instruments, implementation and outcome. (Dissertation VU medisch centrum).

Coid, J., Hickey, N., Kahtan, N., Zhang, T., & Yang, M. (2007). Patients discharged from medium secure forensic psychiatry services: reconvictions and risk factors. *British Journal of Psychiatry, 190,* 223 - 229. doi: 10.1192/bjp.bp.105.018788

De Jong, K. (2012)[1]. De rol van de behandelaar: de 'vergeten' factor in ROM. *Tijdschrift voor Psychiatrie, 54,* 197 - 201.

De Jong, K. (2012). A chance for change: building an outcome monitoring feedback system for outpatient mental health care. (Dissertation Leiden University). Ridderkerk grafisch bedrijf, Ridderkerk.

De Jong, K., Segaar, J., Ingenhoven, T., Van Buschbach, J., & Timman, R. (2017). Adverse effects of outcome monitoring feedback in patients with personality disorders: a randomized controlled trial in day treatment and inpatient settings. *Journal of Personality Disorders, 31, 1* - 21. doi: 10.1521/pedi_2017_31_297

Delespaul, P. A. E. G. (2015). Routine outcome measurement in the Netherlands - A focus on benchmarking. *International Review of Psychiatry, 27,* 320-328. doi: 10.3109/09540261.2015.1045408

Delgadillo, J., De Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., .McMillan, D. (2018). Multi-site randomised controlled trial of outcome feedback technology used to support the psychological treatment of depression and anxiety. *The Lancet Psychiatry*

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

Dienst Justitiële Inrichtingen, Ministerie van Veiligheid en Justitie. (2015). *Kernset prestatie-indicatoren forensische psychiatrie. Forensische geestelijke gezondheidszorg, verslavingszorg en zorg voor mensen met een (licht) verstandelijke beperking*. Directie Forensische zorg. Den Haag.

Dienst Justitiële Inrichtingen, Ministerie van Veiligheid en Justitie. (2017). *Kernset prestatie-indicatoren forensische psychiatrie. Forensische geestelijke gezondheidszorg, verslavingszorg en zorg voor mensen met een (licht) verstandelijke beperking*. Directie Forensische zorg. Den Haag.

Dolan, M., & Doyle, M. (2000). Violence risk prediction: Clinical and actuarial measures and the role of the psychopathy checklist. *British Journal of Psychiatry, 177,* 303 - 311. doi: 10.1192/bjp.1774.303

Douglas, K. S. & Skeem, J. L. (2005). Violence risk assessment. Getting specific about being dynamic. *Psychology, Public, Policy, and Law, 11,* 347 - 383. doi: 10.1037/1076-8971.11.3.347

Dowden, C., Antonowicz, D., & Andrews, D. A. (2003). The effectiveness of relapse prevention with offenders: A meta-analysis. *International Journal of Offender Therapy and Comparative Criminology, 47,* 516-528. Doi: 10.1177/ 0306624X03253018

Doyle, M. R. M. N., & Logan, C. D. (2012). Operationalizing the assessment and management of violence risk in the short term. *Behavioral Sciences and the Law, 30,* 406 - 419. doi: 10.1002/bsl.2017

Fitzpatrick, R., Chambers, J., Burns, T., Doll, H., Fazel, S., Jenkinson, C., Yiend, J. (2010). A systematic review of outcome measures used in forensic mental health research with consensus panel opinion. *Health Technology assessment*, 18, 1 – 91. doi. 10.3310/hta14180

Goethals, K., & Van Marle, H. J. C. (2012). Routine outcome monitoring in de forensische psychiatrie: Een lang verhaal in het kort. [Routine outcome monitoring in forensic psychiatry: A long story cut short] *Tijdschrift voor Psychiatrie, 2*, 179-183. Retrieved from: http://www.tijdschriftvoorpsychiatrie.nl/issues/450/articles/9335

Grann, M., Sturidsson, K., Haggård-Grann, U., Hiscoke, U. L., Alm, P-U., Denevik, M., Gumpert, C., Hallqvist, J., Hallquist, T., Kullgren, G., Långström, N., Lotterberg, M., Nordström, K., Ståhle, B., & Woodhouse, A. (2005). Methodological development: Structured outcome assessment and community risk monitoring (SORM). *International Journal of Law and Psychiatry, 28,* 442 - 456. doi: 10.1016/j.ijlp.2004.03.011

Hafkenscheid, A., & Van Os, J. (2016). Wat iedereen die betrokken is bij ROM zich over de metingen moet realiseren. [ROM measurements in mental health care: users need to be aware of the problems and pitfalls]. *Tijdschrift voor Psychiatrie, 58,* 388 - 396. Harris, G. T., & Rice, M. E.. (1997). Risk appraisal and management of violent behavior. *Psychiatric Services, 48,* 1168 - 1176

Heap, M. (2003). Differences in the progress of discharged and undischarged patients in a medium secure unit: a pilot study. *Journal of Psychiatric and Mental Health Nursing, 10,* 534 - 543.

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. doi: 10.1016/j.ijlp.2012.04.001 Jacobson, N. S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12-19. doi: 10.1037/0022-006X.59.1.12

Jonker, F., Kruisdijk, S., Goedhard, L., & Nijman, H. (2016). *Adaptieve Vragenlijst Verstandelijk Beperking.* Den Dolder: Wier.

KFZ (n.d.). K-factoren en behandelverloop (Call 2014-33). Retrieved from: https://kfz.nl/resultaten/call-2014-33

Kortrijk, H. E. (2013). Use of routine outcome monitoring data for evaluating assertive community treatment. (Dissertation, Erasmus University Rotterdam). Ipskamp drukkers, The Netherlands

Kroner, D. G. & Yessine, A. K. (2013). Changing risk factors that impact recidivism: in search of mechanisms of change. *American Psychological Association, 37,* 321-336. doi: 10.1037/lhb0000022

Krueger, C. (2004). A comparison of the general lineair mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research for Nursing, 6,* 151 – 157. doi: 10.1177/1099800404267682

Lambert, M. J. (2010). Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice. , (pp. 109-134). Washington, DC, US:American Psychological Association, xiv, 293 pp. http://dx.doi.org/10.1037/12141- 005

Lambert, M. J., & Hawkkins, E. J. (2004). Measuring outcome in professional practice: Considerations in selecting and using brief outcome instruments. *Professional Psychology: research and Practice, 35,* 492 - 499. doi: 10.1037/0735-7028.35.5.492

Lambert, M. J., Whipple, J. L., Hawkins, E. D., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003) Is it time for clinicians to rouintely track patient outcome? A meta- analysis. *American Psychological Association, D12,* 288 - 301

Lammers, S., Vrinten, M., & Nijman, H. Te lang in de tbs? Een onderzoek naar patienten die meer dan 15 jaar in de tbs-behandeling zijn. Den Haag: WODC en Nijmegen: Radboud Universiteit; 2015.

Lewis, K., Olver, M. E., & Wong, S. C. P. (2012). 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. doi: 10.1177/1073191112441242.

Maassen, E. F., Schrevel, S. J. C., Dedding, C. W. M., Broerse, J. E. W., & Regeer, B. J. (2017) Comparing patients' perspectives of "good care" in Dutch outpatient psychiatric services with academic perspectives of patient-centred care. *Journal of Mental Health, 26*, 84-94, doi: 10.3109/09638237.2016.1167848

Morisse, F., & Dosen, A. (red.) (2016). *SEO-R2 . Schaal voor emotionele ontwikkeling voor mensen met een verstandelijke handicap – Revised2*. Antwerpen/Apeldoorn: Garant. Morrissey, C., Beeley, C., & Milton, J. (2014). Longitudinal HCR-20 scores in a high-secure psychiatric hospital. *Criminal Behaviour and Mental Health, 24,* 169 - 180. doi: 10.1002/cbm.1893

Muchall S., & Bogaerts, S. *Routine Outcome Monitoring (ROM) Feedback en behandelmotivatie bij forensisch psychiatrische patiënten met een hoog en laag risicoprofiel.* Submitted.

Nagtegaal, M. H., Goethals, K., & Meynen, G. (2016). De tbs-maatregel: kosten en baten in perspectief. [Mandatory treatment of forensic psychiatric patients in the Netherlands: costs and benefits in perspective]. *Tijdschrift voor psychiatrie, 58,* 739-745. Nagtegaal, M. H., Van der Horst, R. P., & Schönberger, H. J. M. *Inzicht in de verblijfsduur van bsgestelden: Cijfers en mogelijke verklaringen*. 2011. Boom juridische uitgevers, Meppel.

Nijnam, H., Didden, R., & Hesper, B. *Handreiking Risicomanagement LVB.* (2017). Kwaliteit forensische zorg. Retrieved from: www.kfz.nl

Nijman, H., Lammers, S., Vrinten, M., & Bulten, E. (2017). (Te) lang in tbs? [(to) long in TBS?]. *Tijdschrift voor Psychiatrie, 59,* 9-19 .

Nijman, H., De Kruyk, C., & Van Nieuwenhuizen, C. (2004). Behavioural changes during forensic psychiatric (TBS) treatment in the Netherlands. *International Journal of Law and Psychiatry, 27*, 79 – 85. doi: 10.1016/j.ijlp.2003.12.001

O'Brien, K. & Daffern, M. (2016). Treatment gain in violent offenders: the relationship between proximal outcomes, risk reduction and violent recidivism. *Psychiatry, Psychology and Law.* doi: 10.1080/13218719.2016.1209804

Oudejans, S., Spits, M., Nugter, A., & Van Bokkem, S. Onderzoek naar een geschikt instrument voor het meten van de behandelvoortgang bij forensische patiënten met een persoonlijkheidsstoornis. 2015, Kwaliteit forensische zorg. Retrieved from: www.kfz.nl

Polaschek, D. L. L. (2012). An appraisal of the risk-need-responsivity (RNR) model of

offender rehabilitation and its application on correctional treatment. *Legal and Criminological Psychology, 17,* 1 -17. doi: 10.1111/j.2044-8333.2011.02038.x

Roe, D., Gelkopf, M., Gornemann, M. I., Baloush-Kleinman, V., & Shadmi, E. (2015). Implementing routine outcome measurement in psychiatric rehabilitation services in Israel. *International Review of Psychiatry, 27,* 345 - 353. doi: 10.3109/09540261.2015.1025722

Schel, S. H. H., Bouman, Y. H. A., Vorstenbosch, E. C. W., & Bulten, B. H. (2016). Development of the forensic inpatient quality of life quistionnaire: short version (FQL-SV). *Quality of Life Research,* online publication: doi: 10.1007/s11136-016- 1461-9

Schel, S. H. H., Bouman, Y. H. A., Schaftenaar, P., Smeekens, M. V., Smeijers, D., & Bulten, B. H. (2017). *Aanbevelingen en handvatten voor het verbeteren van de objectieve en subjectieve kwaliteit van leven van forensisch psychiatrische patiënten.* [recommendations and tools to improve the objective and subjective quality of life of forensic psychiatric patients]. Kwaliteit forensische zorg. Retrieved from: www.kfz.nl

Schoen, R., & Derksen, J. (2011). Praktische vergelijking van uitkomstinstrumenten in de geestelijke gezondheidszorg. *Directieve therapie, 31*, 209 – 248. doi: 10.1007/s12433-011-0012-3

Schoevers, R. A. & Beekman, A. T. F. (2017). Van 'stop ROM' naar 'hier met die ROM!' [Not 'stop ROM' but 'give me my ROM!']. *Tijdschrift voor Psychiatrie, 59,* 245-246.

Schuringa, E., Heininga, V., & Spreen, M. (2011). De N = 1 statistiek achter het patiënt volg systeem in het FPC Dr. S. van Mesdag. GGzet Wetenschappelijk, 15, 70-77.

Schuringa, E., Heininga, V. E., Spreen, M., & Bogaerts, S. (2018). Concurrent and predictive validity of the instrument for forensic treatment evaluation: From risk assessment to routine, multidisciplinary treatment evaluation. *International Journal of Offender Therapy and Comparative Criminology, 62,* 1281-1299. doi: 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,* 127 - 144. doi: 10.1080/15228932.2014.897536

Schuringa, E., Spreen, M., & Bogaerts, S. (2018[1]) Routine outcome monitoring in de tbs. De voorspellende waarde van het instrument voor forensische behandelevaluatie (IFBE) voor intramurale geweldsincidenten bij verschillende doelgroepen. *Tijdschrift voor psychiatrie, 60,* 662-671.

Sedgwick, O., Young, S., Das, M., & Kumari, V. (2016). Objective predictors of outcome in forensic mental health services- a systematic review. *CNS spectrums, 21,* 430-444. doi: 10.1017/S1092852915000723.

Shinkfield, G., & Ogloff, J. (2014). A review and analysis of routine outcome measures for forensic mental health services. *International Journal of Forensic Mental Health, 13,* 252 - 271. doi: 10.1080/14999013.2014.939788

Shinkfield, G., & Ogloff, J. (2015). Use and interpretation of routine outcome measures in forensic mental health. *International Journal of Mental Health Nursing, 24,* 11-18. doi: 10.1111/inm.12092

Slade, M., Thornicroft, G., & Glover, G. (1999). The feasibility of routine outcome measures in mental health. *Social Psychiatry and Psychiatric Epidemiology, 34,* 243 - 249.

Smith, M., & Baxendine, S. (2015). Outcome measurement in New Zealand. *International Review of Psychiatry, 27,* 276 – 285. doi: 10.3109/09540261.2015.1023783

Spreen, M., Brand, E., Ter Horst, P., & Bogaerts, S. (2014). *Handleiding, Historische Klinische Toekomst-Revisie versie 2014 [Manual, Historical, Clinical, Future, Revised version 2014].* The Netherlands, Groningen: Stichting FPC Dr. S. van Mesdag

Tasma, M., Swart, M., Wolters, G., Liemburg, E., Bruggeman, R., Knegtering, H., & Castelein, S. (2016). Do routine outcome monitoring results translate to clinical practice? A cross-sectional study in patients with a psychotic disorder. *BioMed Central, online,* doi:10.1186/s12888-016-0817-6

Tiemens, B. G. & Van Sonsbeek, M. (2017). ROM als tomtom: Als routine outcome monitoring echt routine zou zijn. [ROM as a GPS navigation system: How can ROM be used routinely in treatment procedures?]. *Tijdschrivt voor Psychiatrie, 59,* 489-493.

Van Erp, N., Van Vugt, M., Van der Veeken, F., Van Boxtel, M., & Van Rooijen, S. (2018). *Herstelondersteunend handelingsplan psychotische aandoeningen in de forensische zorg.* Kwaliteit forensische zorg. Retrieved from: www.kfz.nl

Van Nieuwenhuizen. C. H., Bogaerts, S., De Ruijter. E. A. W., Bongers, I. L., Coppens, M., & Meijers, R. A. A. C. tbs-behandeling geprofileerd; een gestructureerde casussenanalyse [*Profiling TBS-treatment: A structured cases analysis]*. Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) Ministerie van Veiligheid en Justitie. 2011. Retrieved from http://www.wodc.nl/onderzoeks database/voor studie-onderzoek-tbs. aspx

Van Nieuwenhuizen, C., & Nijman, H. (2009). Quality of life of forensic psychiatric inpatients. *International Journal of Forensic Mental Health, 8,* 9-15. doi:10.1080/14999010903014671

Ward, T., & Gannon, T. A. (2006). Rehabilitation, etiology, and self-regulation: the comprehensive good lives model of treatment for sexual offenders. *Agression and Violent Behavior, 11,* 77-94. doi: 10.1016/j.avb.2005.06.001

Werkgroep Risicotaxatie Forensische Psychiatrie [Risk Assessment Task Force in Forensic Psychiatry]. (2002). *Handleiding HKT-30 versie 2. Risicotaxatie in de forensische psychiatrie* [Guide HKT-30 version 2. Risk Assessment in Forensic Psychiatry]. Den Haag, The Netherlands: Ministerie van Justitie, Dienst Justitiële Inrichtingen.

Wu, S., Whiteside, U., & Neighbors, C. (2007). Differences in inter-rater reliability and accuracy for a treatment adherence scale. *Cognitive Behaviour Therapy*, 4, 230-239. doi. 10.1080/16506070701584367

Yiend, J., Chambers, J. C., Burns, T., Doll, H., Fazel, S., Kaur, A., Fitzpatrick, R.(2011). Outcome measurement in forensic mental health research: An evaluation. *Psychology, Crime and Law, 17 (3),* 277 – 292. doi: 10.1080/10683160903203938

Samenvatting en conclusie

Doel van het proefschrift

Inzicht in de verandering van de ernst van de problematiek bij forensisch psychiatrische patiënten is nodig om de effectiviteit van de behandeling te kunnen vaststellen. Het doel van dit proefschrift is om te onderzoeken of de forensische Routine Outcome Monitoring (ROM) met het Instrument voor Forensische Behandel Evaluatie (IFBE), als kompas kan dienen voor de behandeling en kan helpen om inzicht te krijgen in de evaluatie en voortgang van de behandeling van forensisch psychiatrische patiënten.

ROM staat voor metingen die periodiek bij patiënten worden gedaan om een beeld te krijgen van de toestand van hun functioneren en maakt behandelvoortgang inzichtelijk voor zowel behandelaren als patiënten De behandeling kan zo nodig op geleide van deze metingen worden bijgestuurd. Binnen de forensische psychiatrie is het met name van belang dat het risico op recidive wordt teruggedrongen. De Risk-Need-Responsivity (RNR) principes schrijven voor dat dynamische criminogene behoeften, ofwel veranderbare risicofactoren, door de behandeling worden gereduceerd en dat rekening wordt gehouden met de responsiviteit van de patiënt en de behandelaar. Om die reden is het zeer belangrijk dat een ROM-instrument wordt gebruikt in de forensische psychiatrie welke de behandeling en veranderbaarheid van deze risicofactoren- en beschermende factoren inzichtelijk maakt.

Het IFBE is specifiek ontwikkeld om forensisch relevante behandelindicatoren te monitoren. Het instrument wordt multidisciplinair gescoord en heeft 22 items en maakt gebruik van een 17-puntschaal. Het IFBE bestaat uit drie componenten, namelijk probleemgedrag, beschermend gedrag en resocialisatievaardigheden. Het algemeen doel van een behandeling in een forensisch psychiatrisch centrum (FPC), waar dit onderzoek plaatsvond, is om probleemgedrag te verminderen en beschermend gedrag en resocialisatie vaardigheden te verbeteren. Het IFBE is gekozen om het behandelverloop te monitoren. Om te onderzoeken of het IFBE daadwerkelijk als kompas in de behandeling kan dienen, is onderzoek naar de psychometrische eigenschappen van het instrument en de veranderbaarheid ervan gedaan.

FPC de Kijvelanden en FPC 2landen zijn in 2011 gestart met het gebruik van ROM in de behandeling. Dit zijn beide instellingen waar patiënten met een tbs-maatregel (maatregel terbeschikkingstelling met dwangverpleging), zijn opgenomen en worden behandeld. De ROM-data in deze studie betreffen de ROM gegevens die primair voor de behandeling zijn afgenomen.

ROM in de tbs

Als eerste wordt in hoofdstuk 2 geïnventariseerd en beschreven welke ROM-instrumenten worden toegepast bij patiënten met een tbs-maatregel. Drie instrumenten worden beschreven voor drie verschillende groepen patiënten in FPC 2landen en FPC de Kijvelanden: pati-

enten die verblijven in een sociaal therapeutisch milieu, patiënten die verblijven in een supportief milieu en patiënten die verblijven in een milieu ten behoeve van licht verstandelijk beperkten (LVB). Voor de eerste twee milieus (sociaal therapeutisch en supportief) wordt het IFBE gebruikt als ROM-instrument maar worden ook andere instrumenten (de Health of the Nations Outcomes Scales, Manchester Assessment of Quality of Life en wanneer geïndiceerd voor het supportieve milieu de Positive and Negative Syndrome Scale-remissie tool en de Sociale Redzaamheidsschaal voor verstandelijk beperkten van hoger niveau) toegepast om het symptomatisch functioneren, dagelijks functioneren en de kwaliteit van leven inzichtelijk te maken.

Patiëntprofielen in de tbs

Forensisch psychiatrische patiënten met een tbs-maatregel vormen een zeer heterogene groep wat betreft psychopathologie, risico- en beschermende factoren en de delicten die ze hebben gepleegd. Om orde te scheppen in deze heterogeniteit zijn in hoofdstuk 3 patiëntprofielen onderzocht. Met latente klasse analyse is onderzocht of er onderscheidbare patiëntprofielen kunnen worden geïdentificeerd op basis van historische en dynamische risicofactoren, psychopathologie en het type delict, die gebruikt kunnen worden om een behandeling of zorgpad te indiceren. De resultaten leiden tot vier patiëntprofielen: het "antisociale profiel", het "gemengde profiel met meerdere problemen", het "ontwikkelingsverstoorde en affectief verstoorde profiel" en het "psychotic first offender profiel".

Deze vier patiëntprofielen vertonen overeenkomsten met eerder gevonden patiëntprofielen door Van Nieuwenhuizen et al. (2011) en Bogaerts en Spreen (2011) en impliceren dat er verschillende risicofactoren, diagnoses en delictsvormen zijn voor verschillende groepen van patiënten. In hoofdstuk 6 is aandacht besteed aan de vraag of deze profielen ook door de tijd heen als gevolg van behandeling kunnen veranderen.

Het IFBE als ROM-instrument

In het onderzoek is aandacht besteed aan de vraag of het IFBE bruikbaar is als ROM-instrument. Er is gekeken naar betrouwbaarheid, validiteit en veranderbaarheid.

Betrouwbaarheid, veranderbaarheid en de factorverdeling

De psychometrische kwaliteiten van het IFBE zijn onderzocht in hoofdstuk 4. De interbeoordelaarsbetrouwbaarheid, de test-hertest betrouwbaarheid, de interne consistentie en de veranderbaarheid van het IFBE zijn samen met de factorstructuur onderzocht. De resultaten laten zien dat de interbeoordelaarsbetrouwbaarheid met uitzondering van de items *psychotische symptomen* en *vaardigheden ter voorkoming van fysiek agressief gedrag* voldoende

is, mits het IFBE multidisciplinair wordt gescoord. De test-hertest betrouwbaarheid is op de items *oriëntatie op slechte personen of groepen* en *vaardigheden ter voorkoming van fysiek agressief gedrag* na, eveneens voldoende. Met name het item *vaardigheden ter voorkoming van fysiek agressief gedrag* vraagt om een betere toelichting bij het item. De interne consistentie van de drie factoren is goed. Zoals ook uit eerder onderzoek blijkt laden de afzonderlijke items voldoende op de factor *resocialisatie vaardigheden.* Echter wordt in plaats van drie factoren, vier factoren gevonden en laden meerdere items op meerdere factoren (kruisladingen). Dit betekent dat de factorstructuur van het IFBE in dit onderzoek met de nodige voorzichtigheid moet worden geïnterpreteerd. In de studies in dit proefschrift is wel vastgehouden aan de factorstructuur zoals oorspronkelijk bepaald door Schuringa, Spreen en Bogaerts (2014). In toekomstig onderzoek is het wel belangrijk aandacht te besteden aan de factorstructuur van het IFBE.

Het IFBE heeft ook tot doel verandering op probleemgedrag, beschermend gedrag en resocialisatie vaardigheden over de tijd te meten. Uit het onderzoek blijkt dat het IFBE in staat is om verandering aan te tonen voor patiënten die hoog problematisch gedrag of weinig vaardigheden op een van de factoren vertonen aan het begin van de behandeling. Patiënten die reeds weinig probleemgedrag vertonen of patiënten die reeds hoog scoren op beschermend gedrag of resocialiserende vaardigheden laten een trend zien waarin zij achteruitgaan op de betreffende factor.

Concluderend zijn de psychometrische waarden voor de meeste items goed. De factorstructuur vereist echter aanvullend toekomstig onderzoek. Het onderzoek heeft laten zien dat het IFBE in staat is verandering te meten op de drie factoren.

Het IFBE en klinische besluitvorming

In hoofdstuk 5 is ingegaan op de vraag of het IFBE behandelbeslissingen kan ondersteunen en of de items en factoren voorspellend zijn voor belangrijke behandeluitkomsten (verlof, agressie en het schenden van urine drugs screening (UDS) protocollen of het gebruik van middelen). Incidenten tijdens de behandeling, zoals agressie zijn belangrijke voorspellers voor problematisch gedrag na de behandeling. Verlofmodules zijn belangrijke positieve stappen in het resocialisatie traject. Voor de gehele groep patiënten is de voorspellende waarde onderzocht voor het goedkeuren van begeleid verlof, onbegeleid verlof en transmuraal verlof. De resultaten laten zien dat het IFBE matig voorspellend is voor de positieve behandeluitkomst verlof. Patiënten die met verlof mogen, scoren significant hoger op een aantal IFBE factoren in vergelijking met patiënten die niet op verlof mogen. Als negatieve behandeluitkomsten werden algemene agressie (materiële agressie, verbale agressie en fysieke agressie), fysieke agressie en het schenden van regels en voorwaarden (het gebruik van middelen of het

schenden van UDS protocollen) onderzocht bij patiënten met een primaire psychotische stoornis, persoonlijkheidsstoornis of een aan middelen gebonden stoornis, patiënten met een primaire persoonlijkheidsstoornis en patiënten met een persoonlijkheidsstoornis en een co-morbide aan middelen gebonden stoornis. De voorspellende waarde van het IFBE voor het schenden van regels en voorwaarden is marginaal. Enkel de items die over middelengebruik gaan en *schending van regels en voorwaarden,* laten een redelijk voorspellende waarde zien bij alle drie de groepen. De meeste items laten een marginale tot redelijke voorspellende waarde zien voor algemene agressie en fysieke agressie in de drie groepen. Een aantal items laat zelfs een goede voorspellende waarde zien. De factoren *probleemgedrag* en *resocialisatie vaardigheden* zijn het meest gerelateerd aan toekomstige agressie.

Wij kunnen concluderen dat het IFBE en een aantal items redelijke tot goede voorspellende waarden laat zien voor toekomstige incidenten en kan bijdragen aan de signalering van aandachtspunten in de behandeling. De voorspellende waardes zijn echter niet hoog genoeg om op individueel niveau enkel van het IFBE uit te gaan.

Hoewel de voorspellende waarden voor verlof matig zijn, scoren patiënten die met verlof mogen wel hoger op een deel van de factoren. Op dit moment kan het IFBE beperkt gebruikt worden voor de besluitvorming voor verlofmodaliteiten. De vraag is ook of verlof op het ideale moment in de behandeling wordt aangevraagd, mede omdat de afgelopen jaren, met name na het akkoord van Lunteren, wisselend is gedacht over de criteria voor het toekennen van verlof. Toekomstig onderzoek naar de voorspellende waardes van incidenten tijdens verlof en het verloop van de voortgang tijdens verlof, zouden kunnen bijdragen aan het gebruik van het IFBE bij verlofaanvragen.

Behandelvoortgang gemeten met het IFBE

Hoofdstuk 6 gaat in op de vraag of het IFBE verandering in kaart kan brengen. In navolging van hoofdstuk 3 is allereerst geprobeerd om behandelvoortgang voor de vier gevonden patiëntprofielen te onderzoeken. Op basis van latente klasse analyse bleek dat de factoren niet onafhankelijk waren van elkaar waardoor niet werd voldaan aan de assumpties van onafhankelijkheid. Dit was een reden om geen behandelvoortgang te onderzoeken voor de vier patiëntprofielen. Toekomstig onderzoek zou mogelijk meer duidelijkheid kunnen geven over verschillende factoren in de relatie tot het verloop van de behandeling.

Om verandering met het IFBE te onderzoeken is eerst gekeken naar de gehele studiepopulatie. Analyses laten geen significante verandering voor de gehele groep zien wat afwijkt van eerdere studies. Vervolgens is opnieuw onderzocht of patiënten die hogere scores hebben op probleemgedrag meer voortgang tonen in vergelijking met patiënten die lagere scores op probleemgedrag vertonen. Dezelfde redenering geldt voor beschermende factoren

en resocialisatie vaardigheden. In lijn met de resultaten in hoofdstuk 4, tonen patiënten zowel vooruitgang als achteruitgang. Patiënten die laag scoren op beschermend gedrag of resocialisatie vaardigheden of hoog scoren op probleemgedrag, laten een positieve verandering zien. Patiënten die hoog scoren op beschermend gedrag of resocialisatievaardigheden of laag scoren op probleemgedrag, laten een trend zien waarin zij achteruit lijken te gaan. Deze achteruitgang wordt echter niet significant ondersteund door de aanvullende non-parametrische toetsen. Het verschil in voortgang en achteruitgang lijkt overeenkomsten te tonen met de RNR-principes, waarin patiënten met een hoger risico meer profijt kunnen hebben van een intensieve behandeling en waarbij dit averechts kan werken bij patiënten met een laag risico. In de toekomst zou onderzoek gedaan kunnen worden naar baseline scores op de drie factoren zodat behandeling op maat kan worden gegeven. Het is van belang de gevonden trend verder te onderzoeken, gedurende de gehele behandeling.

Eerder onderzoek laat zien dat patiënten meer behandelvoortgang tonen tijdens de eerste periode van de behandeling. Om die reden is ook onderzocht of patiënten die gestart zijn met ROM-metingen in het eerste jaar van de behandeling meer voortgang vertonen, dan patiënten die later zijn gestart met ROM-metingen. De resultaten laten geen verschil in gedragsverandering zien maar laten wel zien dat patiënten die langer dan een jaar in behandeling zijn significant hogere scores op beschermend gedrag en resocialisatievaardigheden vertonen. Dit kan erop wijzen dat patiënten gedurende de behandeling meer vaardigheden hebben verkregen. De resultaten kunnen echter ook beïnvloed worden door andere factoren, zoals een mogelijke verandering in aard en ernst van de problematiek van de populatie. Gedurende de afgelopen jaren is het aantal tbs-patiënten dat instroomt gedaald, dit kan betekenen dat patiënten die voorheen opgenomen zouden zijn met een tbs-maatregel, nu een gevangenisstraf krijgen zonder tbs-maatregel.

Discussie en conclusie

Tot slot is in hoofdstuk 7 ingegaan op de bevindingen en conclusies en zijn beperkingen en sterktes van het onderzoek besproken samen met aanbevelingen voor toekomstig onderzoek. De resultaten laten zien dat het IFBE verandering kan meten op *probleemgedrag, beschermend gedrag* en *resocialisatievaardigheden*. De interbeoordelaarsbetrouwbaarheid, test-hertest betrouwbaarheid en interne consistentie zijn voldoende voor de meeste items. De voorspellende waardes indiceren dat de IFBE items en factoren gerelateerd zijn aan risicovol gedrag wat in lijn is met studies die elders zijn uitgevoerd. Het IFBE lijkt daarmee factoren te meten die gerelateerd zijn aan de RNR-principes en kan daarom gebruikt worden om de behandeling

over tijd te evalueren. De factorstructuur dient echter met voorzichtigheid te worden geïnterpreteerd en de veranderbaarheid van het IFBE is niet gedurende de gehele behandeling onderzocht. Hoewel de waarde van de profielen niet bevestigd wordt in hoofdstuk 6, geven de profielen in hoofdstuk 3 wel een beeld van de populatie binnen de tbs klinieken.

Deze studie heeft een aantal beperkingen. Allereerst betreft het een studie met ROM-data die primair voor de behandeling zijn afgenomen. Dit heeft als gevolg dat tijdstippen tussen metingen konden variëren door de dynamiek binnen de setting en dat er geen randomised controlled trial studie is uitgevoerd aangezien er niet met een controlegroep kon worden gewerkt. Ook is ROM ten tijde van deze studie voor alle patiënten geïmplementeerd, ongeacht de opnamedatum en het moment van afname. Behandelaren konden de ROM afnames niet als kompas of hulpmiddel, maar als administratieve last zien, wat het gebruik, mede als de gevolg van de hoge werkdruk, mogelijk heeft belemmerd. Binnen de voorzieningen kampte men tevens met overplaatsingen van patiënten, was er de sluiting van FPC 2landen en was er een grote doorstroom van behandelaren. Een grotere onderzoeksgroep zou deze studie eveneens ten goede zijn gekomen. Als laatste is deze studie uitgevoerd in een setting waar alleen mannen verblijven. De resultaten zijn daarom ook niet generaliseerbaar naar vrouwen en hebben enkel betrekking op tbs-patiënten.

Het is van belang dat toekomstige studies ingaan op de andere kwaliteiten die een ROM-instrument moet hebben. Dit gaat onder andere om het effect van ROM-feedback met het IFBE. Het is ook belangrijk dat onderzocht wordt of de voortgang, gemeten met het IFBE, gerelateerd is aan toekomstige recidive. Daarnaast dient het IFBE gedurende het gehele traject onderzocht te worden, met aandacht voor de verandering binnen de individuele casus en de diverse factoren die de behandeling kunnen beïnvloeden. Als laatste dient onderzocht te worden in welke mate andere ROM-instrumenten het IFBE kunnen aanvullen.

Dit proefschrift laat zien dat het IFBE gebruikt kan worden in de forensisch psychiatrische behandeling als ROM-instrument. Het IFBE bestaat uit drie factoren die gerelateerd zijn aan de Need-principes van het RNR model. Het IFBE zou ook gebruikt kunnen worden als signaleringsinstrument en zou bij kunnen dragen aan het tijdig signaleren van patiënten die geen voortgang of zelfs achteruitgang vertonen. Op die manier zou het IFBE bij kunnen dragen aan een korter traject, tijdige wijziging van het behandelplan of geïnformeerde besluitvorming. Tenslotte is het niet onbelangrijk te vermelden dat het IFBE gebruiksvriendelijk is en weinig tijd in beslag neemt om te scoren. Het is van belang dat ROM niet als verplichting wordt gezien, maar omarmd wordt door behandelaren en gebruikt wordt als instrument ten bate van de behandeling. Meer kennis over het gebruik van feedback en de relatie van de behandelvoortgang met recidive kan verder bijdragen aan het gebruik van het IFBE en de forensische ROM als kompas in de behandeling.

APPENDIX A

IFBE

Instrument voor
Forensische
Behandel
Evaluatie

Naam Patiënt:	
Naam Beoordelaar:	
Functie Beoordelaar:	
Datum van invullen:	

Instructie

Plaats een 'x' in het vakje onder de score naar keuze. Zo ook bij de score N.G.I. en N.V.T.
In principe gaat men uit van de score 0, 1, 2, 3 en 4 welke worden toegelicht. Tussen elke twee scores zijn drie 'tussenmogelijkheden wanneer bijvoorbeeld score 2 te laag is, maar score 3 te hoog, kan voor kan voor een tussenscore gekozen worden.

Vind je dat de patiënt de afgelopen periode over het algemeen is verbeterd?

Verslechterd	Niet verbeterd	Een beetje verbeterd	Veel verbeterd

Schuringa, E., Bogaerts, S. & Spreen, M (2010)

IFBE

Instrument voor Forensische Behandel Evaluatie

1 Uw probleeminzicht of probleembesef?

Iemand met probleeminzicht heeft inzicht op de eigen mentale processen en de invloed hiervan op het gedrag. Bij
probleembesef heeft de patiënt last van de problemen die zijn gedrag oproept (hij beseft dat hij een probleem heeft) maar hij heeft

geen inzicht in wat zijn gedrag stuurt of hoe hij invloed op zijn gedrag kan uitoefenen.

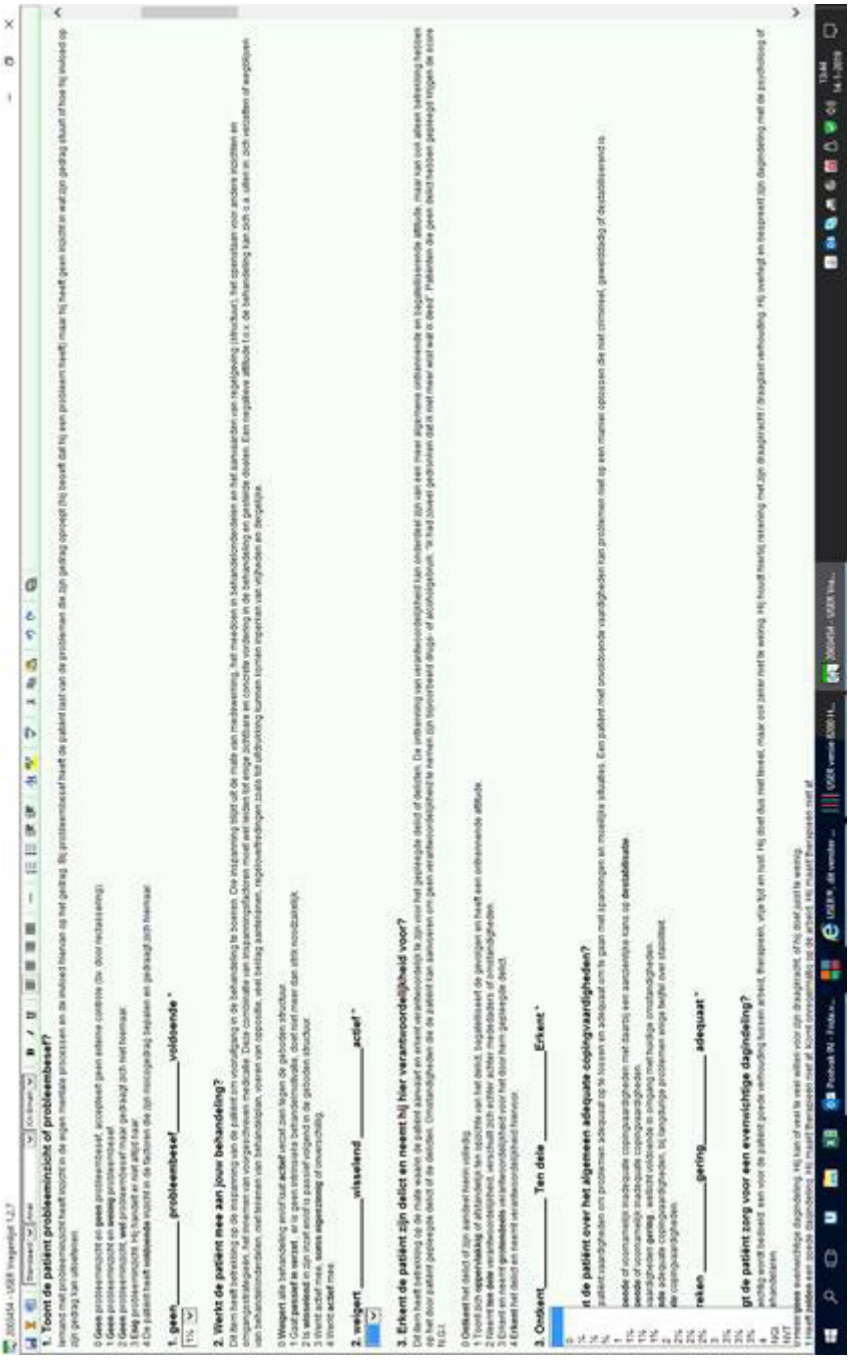
N.G.I.

Geen					Probleembesef										voldoende
0				1				2				3			4

0	Geen probleeminzicht en geen probleembesef, accepteert geen externe controle.
1	Geen probleeminzicht en weinig probleembesef.
2	Geen probleeminzicht, wel probleembesef maar gedraagt zich niet hiernaar.
3	Enig probleeminzicht. Hij handelt er niet altijd naar.
4	De patiënt heeft voldoende inzicht in de factoren die zijn risicogedrag bepalen en gedraagt zich hiernaar.

APPENDIX B

IFTE Question format in the EPD



APPENDIX C

Terugkoppelingguidelines

Belangrijk!

Om hoog betekent stijging, op alle factoren! Hoe hoger hoe beter de resocialisatie factoren en beschermende factoren **hoe minder** probleemgedrag!

Grijs vlak is behandeldoel: voortgang gewenst: zie pijltjes in verandering: omhoog is gewenst!

Terugkoppeling is altijd ter motivatie en transparantie naar je patiënt toe. Lage scores zijn factoren waar nog iets te leren is!

Samenvatting

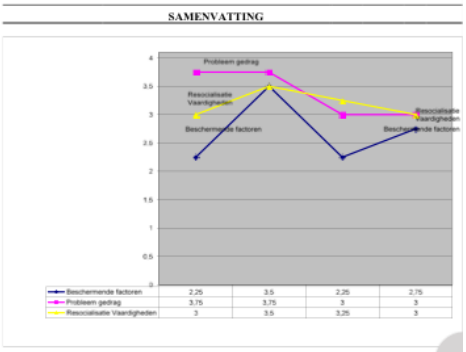
- IFBE is opgedeeld in: beschermend-, probleem gedrag en resocialisatie factoren

Samenvatting

- Geeft een samenvatting over de gemiddelde VOORTGANG op items in deze factoren

Samenvatting

- De gemiddelde voortgang kan je kort benoemen: gemiddeld heb je voortgang geboekt op de factor(en).....



Factor

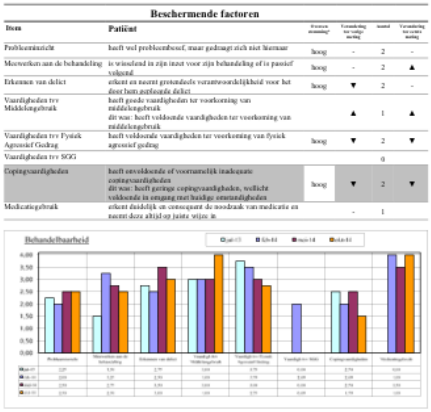
- Per item kan je laten zien (tabel) wat de voortgang is. en beschrijven wat er gescoord is (tekst).
- Bij achteruitgang kan je benoemen waar de patient vandaan komt, wat hij wel kan en vragen waarom het minder goed gaat.

Feedback

- **Licht vooral met voorbeelden (van de afdeling) toe waarom een patiënt een bepaalde score heeft: wees concreet en transparant!**
- Het is de teamscore: het team ziet dit, niet jij alleen.

Feedback

- Soms willen we meer gedrag zien, dit kan achteruitgang betekenen op bv probleemgedrag: benoem dat dit goed is zodat we er wel mee aan de slag kunnen



Zet de afspraken die je maakt door in de bejegening → Pas waar nodig het signaleringsplan aan! → Schrijf doelen op voor de BB

APPENDICES

Curriculum Vitae

Frida Cornelia Antonia van der Veeken was born on January 19th, 1988 in Breda. Frida started studying psychology at Tilburg University in 2006 and earned her Bachelor's degree in 2010. Frida's interest in Forensic Psychology increased during her Bachelor study, which resulted in the attendance of the Master in Forensic Psychology. During this master she started an internship at FPC de Kijvelanden where she started to work on Routine Outcome Monitoring (ROM) and the instrument for Forensic Treatment evaluation for a master's thesis. Due to this internship, her interest in the forensic psychiatric population was increased, along with an interest in research. Both these interests and the experience with ROM for her Master's thesis resulted in the start of a PhD in cooperation with Tilburg University. For a year Frida also worked as a lecturer at Tilburg University and guided students with their bachelor's and master's theses.

In 2016, she started working with adolescents as a psychologist at what was then known as "Het Dok" Breda and Tilburg. This is an outpatient treatment facility of FPC de Kijvelanden and is now part of "Fivoor". While working as a psychologist, she attended several trainings including a Cognitive Behavioural Therapy Basis Course.

Publications and presentations

Van der Veeken, F. C. A., Bogaerts, S., Lucieer, J. (2012). Routine Outcome Monitoring in forensic psychiatry: FPC 2landen and FPC de Kijvelanden. In Oei, T.I. & Groenhuijsen, M.S. Editors. Progression in Forensic Psychiatry. Alphen aan den Rijn: Kluwer;2012. p 355-375.

Van der Veeken, F. C. A., Bogaerts, S., & Lucieer, J. (2016). Patient profiles in Dutch forensic psychiatry, based on risk factors, pathology and offense. *International Journal of Offender Therapy and Comparative Criminology, 61,* 1369-1391. doi: 10.1177/ 03066 24X15619636

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.* doi: 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, 12.* doi:10.1371/journal.pone.0160787

Van der Veeken, F. C. A., Lucieer, J., & Bogaerts, S. (2018). Forensic psychiatric treatment evaluation: The clinical evaluation of treatment progress with repeated forensic routine outcome monitoring measures. *International Journal of Law and Psychiatry, 57,* 9-16. doi: 10.1016/j.ijlp.2017.12.002

Van Erp, N., Van Vugt, M., Van der Veeken, F., Van Boxtel, M., & Van Rooijen, S. (2018). *Herstelondersteunend handelingsplan psychotische aandoeningen in de forensische zorg.* Kwaliteit forensische zorg. Retrieved from: www.kfz.nl

Presentations

33rd Congress of International Academy of Law and Mental Health (Amsterdam)
Presentation: Routine Outcome Monitoring in Forensic Psychiatry.
The Annual conference of the European Association of Psychology and Law 2013 (Coventry, UK)
Presentation: Patient Risk Profiles
The Annual Conference of the European Association of Psychology and Law 2014 (St Petersburg, Russia)
Poster presentation: Forensic Patient Profiles Based on Historical, Clinical, Future items, Psychopathology and Offence
Poster presentation: Routine Outcome Monitoring in Forensic Psychiatric Centers Among Different Forensic Groups (case evaluation)
Presentation: Routine Outcome Monitoring in Forensic Psychiatry
XIV Annual Meeting of the International Association of Forensic Mental Health Services 2014 (Toronto, Canada)
Poster presentation: Forensic Patient Profiles Based on Historical, Clinical, Future items, Psychopathology and Offence
Poster presentation: Routine Outcome Monitoring in Forensic Psychiatric Centers Among Different Forensic Groups (case evaluation)
34rd congress of International Academy of Law and Mental Health 2015, Vienna
Presentatie: Forensic treatment evaluation
Kwaliteit forensische zorg landelijke implementatie dag 2017, Utrecht
Presentation: handelingsplan psychotische aandoeningen in de forensische zorg
Festival forensische zorg 2018, Utrecht
Presentatie: handelingsplan psychotische aandoeningen in de forensische zorg

Dankwoord

Dit proefschrift had niet tot stand kunnen komen zonder de ondersteuning, de inzet en de begeleiding van anderen. Ik wil iedereen die, op wat voor manier dan ook, een bijdrage heeft geleverd aan dit proefschrift bedanken. Het zal niet mogelijk zijn iedereen persoonlijk te noemen, maar ik wil graag stilstaan bij het bedanken van een aantal mensen.

In het bijzonder ben ik mijn promotor en copromotor, Prof. Dr. Stefan Bogaerts en Dr. Jacques Lucieer zeer dankbaar. Zonder jullie inzet, kritische blik, vertrouwen en geduld had dit proefschrift nooit tot stand kunnen komen.

Beste Stefan, we hebben elkaar leren kennen tijdens het eerste jaar van de master forensische psychologie, waarbij jij verbonden bent als professor en hoofddocent. Na mijn masterthesis over het IFBE zijn we verder gegaan met het implementeren van ROM in FPC de Kijvelanden en daarnaast ook FPC 2landen. Tijdens dit project heb je me de kans en het vertrouwen gegeven om aan dit proefschrift te beginnen. Gedurende de jaren hebben we veel samengewerkt en heb ik veel geleerd. Je kritische blik en aansporingen om verder te kijken hebben een belangrijke bijdrage geleverd aan het eindresultaat. Naast het 'sparren' over het onderzoek hebben we ook altijd veel klinische aspecten van het werk kunnen bespreken en heb je ook altijd interesse getoond in mij als persoon. Je hebt me aangespoord mijzelf te ontplooien op wetenschappelijk gebied. Hierdoor ben ik niet alleen als wetenschappelijk onderzoeker gegroeid maar heb ik als psycholoog, werkend in een forensische instelling, ook ontzettend veel geleerd.

Beste Jacques, jij kon een factor van rust bieden, maar ook de lijn vasthouden die we voor ogen hadden. Daarnaast kon je als geen ander vragen stellen waardoor de resultaten in een verschillend perspectief kwamen te staan. Als ik het niet meer wist, wist je altijd de juiste dingen te zeggen en voelde ik me begrepen. Daarnaast zal ik de gesprekken op het gebied van onderzoek maar ook over het klinische reilen en zeilen ontzettend missen. Naast het werk in FPC de Kijvelanden heb je me ook de kans gegeven om te werken in FPC 2landen. Ik ben blij dat ik onderdeel heb mogen uitmaken van het mooie team op deze locatie die helaas gesloten is in 2015. Je hebt me niet alleen begeleid tijdens dit traject en bij alle onzekerheden die hier bij horen, je hebt me ook gestimuleerd bij de ontwikkeling als behandelend psycholoog, een ontwikkeling die ik graag voortzet.

Het bestuur en de directie van FPC de Kijvelanden en toenmalig FPC 2landen wil ik bedanken voor de mogelijkheid om de ROM te implementeren in beide klinieken en het onderzoek uit te voeren op beide locaties. Daarnaast wil ik alle collegae die jarenlang alle ROM-lijsten hebben ingevuld enorm bedanken voor hun medewerking. Ik hoop oprecht

dat het proefschrift bijdraagt aan de mogelijkheid om het IFBE als instrument ten bate van de behandeling te gebruiken. Daarnaast wil ik ook alle patiënten bedanken met wie ik de afgelopen jaren de IFBE zelfrapportage lijsten heb mogen afnemen en bespreken. Jullie inzichten hebben me enorm veel geleerd.

Tilburg University wil ik graag bedanken voor de mogelijkheid om te starten met het promotietraject om uiteindelijk te promoveren. Hierbij wil ik ook alle collegae van het departement ontwikkelingsstoornissen bedanken voor de goede gesprekken en de tips omtrent het uitvoeren van onderzoek. Jenny, Mirthe, Theo en Jelle, het was fijn dat ik af en toe even binnen kon lopen met vragen over het uitvoeren onderzoek en onzekerheden die hierbij horen. And of course, I also want to thank Carlo for his advice.

Tevens wil ik de leden van de commissie Prof. Dr. G. Meynen, Prof. Dr. F. Vander Laenen, Prof. Dr. I. S. J. G. Jeandarme en Dr. M. Spreen bedanken voor het beoordelen van het proefschrift en voor de bereidheid om zitting te nemen in de oppositie. Bedankt voor de investering en tijd die jullie hierin gestoken hebben.

Uiteraard wil ik ook mijn twee paranimfen Carla van der Veeken en Marloes Hartkoorn bedanken voor de ondersteuning tijdens de afgelopen jaren, maar ook zeker voor de ondersteuning bij de laatste loodjes. Carla, je hebt me enorm geholpen met het doorlezen van teksten en je wordt een hele leuke tante. Het is leuk om te zien hoe betrokken je bent. Marloes, we hebben veel goede gesprekken gehad op kantoor. De gezelligheid, de muziek en het luisterend oor waren ook erg belangrijk. Je hebt nog een taartje tegoed.

Ook wil ik Erwin Schuringa bedanken voor de informatie en de adviezen over het gebruik van het IFBE, zonder jouw werk hadden we de IFBE rapporten niet kunnen maken. Ik wil Henk Nijman tevens bedanken voor het advies over ROM voor licht verstandelijk beperkten en Wilco Emons voor de hulp bij de latente klasse analyse bij het laatste artikel.

Daarnaast wil ik graag mijn collegae van het vaste kantoor in de Kijvelanden noemen. Brigitte, Ysaline, Gerben, Simone, Almar, Daniëlle en Joyce, bedankt voor jullie interesse en steun gedurende de afgelopen jaren en natuurlijk ook de gezelligheid. Het kerstontbijt houden we erin! Nicole en Thijs jullie hebben mij ook enorm geholpen met adviezen over onderzoek en de hulp met Excel. Uiteraard wil ik ook al mijn collegae van het ambulant centrum Breda en ambulant centrum Tilburg bedanken voor jullie interesse en de steun bij de laatste loodjes en natuurlijk het feit dat we twee fijne teams hebben!

Gedurende dit traject heb ik een aantal stagiaires mogen begeleiden bij het schrijven van een masterscriptie, zij hebben tevens bijgedragen aan de dataverzameling. Debby, Renée, Marjolein, Floor en Jasper door jullie inzet tijdens jullie stage is de dataverzameling een stuk soepeler gegaan. Jasper, je hebt me ook goed geholpen na je stage. Daarnaast wil ik Viviënne Burgers bedanken voor haar inzet bij de laatste ordening van de dataset in dit onderzoek.

Buiten alle collegae en professionals die op mijn pad zijn gekomen bij de totstandkoming van dit proefschrift, wil ik uiteraard ook mijn lieve familie en vrienden bedanken. Zonder jullie steun en hulp en belangrijke afleiding was het een stuk moeilijker geweest. De boswandelingen, lunches, etentjes, borrels, fantastische weekendjes en goede gesprekken hebben enorm geholpen en een belangrijke afleiding geleverd. Koen, ik ben enorm blij met de kaft die je hebt ontworpen en de tijd die je daarvoor vrij gemaakt hebt! Tereza, het was enorm fijn dat je tijd wilde vrijmaken om feedback te leveren op de Nederlandse tekst. Daarnaast hebben jullie mij veel begrip getoond wanneer ik iets af moest zeggen als ik een deadline had. Dat is ook zeker iets waar mijn softbalteam mij ook bij geholpen heeft! Ik voel me bijzonder gezegend met al mijn lieve vrienden en familie. Gelukkig is het proefschrift nu afgerond en heb ik weer meer ruimte in de agenda (op een verlof na)! Nu kunnen we het afronden gaan vieren.

Daarnaast wil ik in het bijzonder mijn ouders bedanken voor alle steun die ze mij al die jaren hebben gegeven. Hoewel het belangrijk was dat ik mijn best deed, hebben jullie mij altijd de kans gegeven mijn eigen keuzes te maken. Door al die jaren heen hebben jullie altijd interesse getoond in wat ik deed en hebben jullie me altijd gesteund. Door jullie heb ik het vertrouwen gevonden mijn eigen weg te gaan en de kansen die ik kreeg te benutten. Bij de kleinste stapjes die ik maakte had mijn vader al het idee dat ik heel wat bereikt had voor mijn proefschrift, niemand was zo trots als jullie. Mam, je bent een van de sterkste vrouwen die ik ken, niemand doet je na wat jij nu doet. Ik ben ook trots op jullie en jullie gaan een fantastische opa en oma worden.

Als laatste wil ik natuurlijk Maurice bedanken voor alle steun van de afgelopen jaren. Je bent er altijd voor me geweest, vooral als het niet goed ging. Jij was degene die kon zeggen dat ik nu echt eerst even moest gaan zitten en rustig aan moest doen. Jij wist me altijd in het hier en nu te houden en je kan bij iedereen je rust overbrengen. De afgelopen jaren hebben we veel mooie dingen meegemaakt en mooie reizen gemaakt. Samen hebben we een huis gekocht, die jij vooral heel erg opgeknapt hebt met hulp van onze lieve vrienden. Nu beginnen we aan de volgende stap en ik weet zeker dat jij de liefste vader wordt.