Change during forensic treatment in psychopathic versus nonpsychopathic offenders

Farid Chakhssi[a]; Corine de Ruiter[b]; David Bernstein[b]

[a] Forensic Psychiatric Centre, "de Rooyse Wissel", Venray, The Netherlands [b] Department of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands

PLEASE SCROLL DOWN FOR ARTICLE

RESEARCH ARTICLE

Change during forensic treatment in psychopathic versus nonpsychopathic offenders

Farid Chakhssi[a]*, Corine de Ruiter[b] and David Bernstein[b]

[a]*Forensic Psychiatric Centre, "de Rooyse Wissel", Venray, The Netherlands;*
[b]*Department of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands*

Psychopathy in forensic psychiatric patients and other criminal offenders is associated with higher criminal recidivism rates. Moreover, many forensic mental health professionals believe that psychopaths are not amenable to treatment. The present study examines whether patients with psychopathy demonstrate change during forensic psychiatric treatment. Seventy-four personality disordered offenders who had been convicted for serious violence were rated on the the Hare Psychopathy Checklist-Revised and assessed repeatedly on risk-related behaviors during 20-months of inpatient forensic treatment. Group- and individual-level analyses showed no significant differences between psychopathic and non-psychopathic patients on adaptive social behavior, communication skills, insight, attribution of responsibility, and self-regulation strategies. However, a subgroup of psychopaths (22%) deteriorated during treatment with regard to physical aggression, whereas none of the non-psychopathic patients did ($p < 0.01$). Our findings demonstrate that, contrary to clinical lore, treatment does not make a majority of psychopaths worse, but there are significant differences between psychopaths and non-psychopaths in treatment responsiveness.

Keywords: psychopathy; antisocial behaviour; forensic mental health; treatment

Introduction

Psychopathy can be defined as a disorder of personality which includes predatory behavior, emotional detachment, callousness, impulsivity and persistent antisocial behavior (Hare, 2003; Patrick, 2006). Studies following up on psychopaths after incarceration or hospitalization show that they commit more serious violent crimes, and are more likely to recidivate than nonpsychopaths (Hemphill, Hare, & Wong, 1998; Hildebrand, Hesper,

Spreen, & Nijman, 2005; Leistico, Salekin, DeCoster, & Rogers, 2008). However, only a few studies have addressed the treatability, aimed at reducing violence risk, of psychopaths.

Approximately 24–35% of the population in Dutch forensic inpatient settings consists of psychopaths (Hildebrand & de Ruiter, 2004; Hildebrand et al., 2005) and 13–47% of European and North American forensic psychiatric samples consist of psychopaths (Patrick, 2006). Psychopathic patients pose a tremendous challenge for forensic treatment settings in safeguarding society, especially when forensic mental health professionals are faced with the widely held belief that psychopaths are untreatable (Cleckley, 1988) or that treatment even has adverse effects (see for instance, Hare, Clark, Grann, & Thornton, 2000; Rice, Harris, & Cormier, 1992; Seto & Barbaree, 1999). This belief has recently been challenged by Salekin (2002) who stated, on the basis of a review of 42 treatment studies, that there is no convincing scientific evidence for the belief that psychopathy is untreatable. More importantly, the review shows that although many mental health professionals and researchers adhere to Hare's definition of psychopathy (1991), only four of the 42 studies examined the responsiveness of psychopaths using the current diagnostic standard for assessing psychopathy: the Hare Psychopathy Checklist-Revised (PCL-R; Hare, 1991, 2003). Several authors have extensively reviewed PCL-R based research into psychopathy in relation to treatment involvement and found that the few studies addressing this issue provided insufficient evidence to support the view that psychopathy is immutable (see Harris & Rice, 2006; Skeem, Monahan, & Mulvey, 2002) or to support the view that treatment made psychopaths worse (D'Silva, Duggan, & McCarthy, 2004). On top of that, Seto and Barbaree (1999) who observed that treatment progress in psychopathic sex offenders ($n = 224$) was associated with increases in recidivism, reported in a follow-up study using a longer observation period and more recidivism data, that there was no evidence in the follow-up data that justified their preliminary conclusion that 'treatment causes harm' (Barbaree, 2005).

Results of a recent PCL-R based study, suggest that psychopaths are as likely to benefit from treatment as nonpsychopaths. A prospective study among 381 male offenders mandated to residential drug treatment found that psychopaths who received intensive treatment were over three times less likely to be rearrested at 1 year follow-up than psychopaths who received less intensive treatment (Skeem, 2008). Even though psychopathy was associated with misbehavior and less perceived progress during treatment in this study, the PCL-R scores did not moderate the effect of treatment dose on re-arrest rates. Similar findings were reported in an evaluation of treatment in 871 civil psychiatric patients (Skeem et al., 2002). In this study, psychopathy did not moderate the effect of treatment involvement and subsequent violence during a post-discharge follow-up of 10 weeks.

Moreover, these authors found that civil psychiatric patients with psychopathic traits were three times less likely to be violent when they received an adequate dose of treatment (more than seven sessions during a 10-week period). More recently, similar results have been found in a retrospective study by Olver and Wong (2009). After a 10-year post-treatment follow-up, in a sample of 156 PCL-R assessed sex offenders, they found that sex offenders who showed positive therapeutic responses were less likely to recidivate in violent and sexual crimes. Again, this relationship was not moderated by PCL-R scores and/or Sexual Violence Risk scores (Olver & Wong, 2009). Taken together, these findings suggest that some psychopaths are amenable to change, but also raise several further questions. First, while psychopaths as a group may show significant reductions in mean level of problems, how many psychopaths show clinically relevant change at an individual level? That is, how many improve, how many stay the same, and how many deteriorate? Second, what dynamic risk factors for violence account for the change in psychopaths? Although, psychopaths share a high likelihood of future criminal behavior and violence, their co-morbid psychopathology is far from homogenous (Brinkley, Newman, Widiger, & Lynam, 2004; Hildebrand & de Ruiter, 2004). Hence, the treatment response of psychopaths at a group level may not reflect the development of individual psychopaths during forensic treatment, and more individual focused analyses are necessary to investigate a possible differential treatment within the group of psychopaths. To determine whether individual psychopaths respond to treatment, we use the Reliable Change Index (RCI; Jacobson and Truax, 1991). The RCI was developed to evaluate client change during therapy, and is frequently used in therapy outcome research (Atkins, Bedics, McGlinchey, & Beauchaine, 2005; McGlinchey, Atkins, & Jacobson, 2002; Ogles, Lunnen, & Bonesteel, 2001). The RCI describes change of individuals during treatment not only in terms of reliable improvement but also in terms of no change and/or reliable deterioration. The RCI measures the reliability of change by taking repeated measurements and measurement unreliability of the instrument used into account. Compared to group level analyses, the RCI is a more rigorous methodology. To demonstrate reliable change, patients have to change beyond the degree of change that could be explained by measurement unreliability and repeated assessments alone.

In this prospective study, we examined change during long-term inpatient forensic treatment in a sample of male psychopaths and non-psychopaths classified according to Hare's PCL-R. To obtain a homo-geneous sample, only offenders with personality disorders according to DSM-IV (American Psychiatric Association, 2000) were included, and patients with major mental disorders were excluded. In light of recent research (Skeem, 2008; Skeem et al., 2002), we hypothesized that psychopaths would show the same treatment responsiveness as

nonpsychopaths. We predicted that there would be no group differences in treatment responsiveness, and that on an individual level, similar percentages of psychopaths and non-psychopaths would show improvement, no improvement, or deterioration according to the RCI. We assessed therapeutic change with the Behavioral Status Index (BEST-Index; Reed, Wood, & Robinson, 2000), a measure of dynamic, risk-related behaviors. We chose the BEST-index because it can be administered repeatedly during treatment to assess changes in multiple dimensions of forensic risk, and has shown good reliability and validity in previous studies with forensic patients (Chakhssi, de Ruiter, & Bernstein, 2009; Ross et al., 2008; Woods, 2000). Moreover, because the BEST-Index is completed by paraprofessional informants (i.e. psychiatric nurses), it is less susceptible to response bias, such as socially desirable responding, that often affects the responses of forensic patients to self-report questions.

Method

Setting

This study was conducted at Forensic Psychiatric Centre de Rooyse Wissel (dRW), a Dutch maximum security hospital for the treatment of mentally disordered offenders who are hospitalized under the Dutch judicial measure of 'TerBeschikkingStelling' (TBS). TBS is a mandatory treatment order imposed on offenders who have committed serious offences, carrying a punishment of at least 4 years imprisonment, and who suffer from a mental disorder according to DSM-IV (Axis I and/or Axis II). The TBS-order is prolonged as long as the court deems the patient a danger to society. The hospital has 229 residential treatment beds for male offenders, divided over three locations.

Treatment objectives and program

The overall treatment objective of dRW is to reduce future violence risk in mentally disordered offenders by providing inpatient treatment and, in most cases, a stepwise reintegration into society with the aid of correctional- and health- service providers. All patients reside in 10–12 beds high security wards where they are offered a supportive milieu by psychiatric nurses, including motivational interventions (Miller & Rollnick, 2002), work- and educational programs, individual and/or group psychotherapy by psychologists, psychotherapists, or creative arts therapists (e.g. drama, art, music, or movement therapists), pharmacological interventions by psychiatrists, where indicated, and support in maintaining and building constructive social networks by social workers. During this study, the treatment program was tailored to four offender groups: (1) offenders with schizophrenia and other psychotic disorders, (2) personality disordered offenders, (3) sex

offenders, and (4) intellectually disabled offenders. After an extensive observational and assessment period of 3–6 months, all personality disordered patients are enrolled in a multimodal treatment program with a cognitive behavioral approach which focuses on relapse prevention (e.g. Laws, Hudson, & Ward, 2000). During the observation period of the study ($M = 20$ months), the psychopathic- as well the non-psychopathic offenders followed the same treatment protocol. This consisted of group-based weekly sessions of 2 h led by two cognitive-behavioral therapists. The sessions, which take up to 19 months in length, focus on the patient's criminal behavior, taking responsibility for the committed crime(s), criminogenic needs and the chain of events that resulted into the offence(s), with a continuing emphasis on the patient's participation to formulate alternative strategies to the behaviors and cognitions which resulted in the offence(s). Given the heterogeneity of problematic behavior within personality disordered offenders, every patient's idiosyncratic criminogenic needs are addressed in successive (and sometimes parallel) therapeutic interventions. For example, group-based anger replacement training (Goldstein, Glick, Reiner, Zimmerman & Coultry, 1987) for reactive aggression problems, and/or individual schema focused therapy (Young, 1994) for distorted (antisocial) cognitions, and/or group-based social skills training (Liberman, DeRisi, & Mueser, 1989) and/or creative arts therapy (see Smeijsters & Cleven, 2006). During the study period, the average personality disordered patient received 2 h per week creative arts therapy, 1 h per week group-based social skills, 1 h per week individual psychotherapy, 1 h per week aggression replacement training and 20 h per week a work- and educational program. Also, during the study period psychotropic medications were used in 13.5% ($n = 10$) of the cases, evenly divided among the psychopaths ($n = 5$) and the nonpsychopaths ($n = 5$). Within the psychopathic group: 2 patients used antipsychotics (Pipamperon, 40 mg/day; Zuclopentixol, 30 mg/day), 1 patient used antidepressants (Trazodon, 100 mg/day), 1 patient used anticonsulvia (Depakine 1000 mg/day) and 1 patient used sedatives (Clorazepaatdikalium, 15 mg/day). Within the non-psychopathic group: 2 patients used antidepressants (Clomipramine, 150 mg/day; Paroxetine, 20 mg/day), 2 patients used anticonvulsive medication (Depakine 1000 mg/day; Depakine, 1500 mg/day) and 1 patient used psychostimulants (Methylfenidaat, 36 mg/day). All the medications were used primarily as an adjunct to the therapeutic interventions and were mostly prescribed to provide symptom relief.

Sample

The study sample consisted of 74 personality disordered male offenders, admitted to dRW under the TBS-order between March 1, 2000 and May 1, 2007. All patients who stayed at least 1 year at the hospital were included

and were assessed for psychopathy ($n = 160$) with the Psychopathy Checklist-Revised (PCL-R; Hare, 1991, 2003; Dutch version: Vertommen, Verheul, de Ruiter, & Hildebrand, 2002). In order to obtain a more homogeneous sample, patients with borderline intellectual functioning, psychotic, paraphilic and autism spectrum disorders ($n = 86$) were excluded from the study. The resulting sample consisted of 74 personality disordered offenders, and represented 46% of the total sample. These patients were divided into two groups, psychopathic offenders and nonpsychopathic offenders, based on the PCL-R cut-off score of greater than or equal to 26 (Hildebrand, de Ruiter, & de Vogel, 2004; Hildebrand, de Ruiter, & Nijman, 2004).

Characteristics of the study sample are presented in Table 1. Mean age of the patients was 36.0 years (SD = 8.9), and the mean length of stay in the hospital was 3.9 years (SD = 1.9). Almost 34 (45.9%) of the patients had committed or attempted homicide, 8 (10.9%) had committed sexual offenses, 26 (35.1%) had committed violent theft, robbery or assault and 6 (8.1%) had committed arson. The classification of personality disorders was conducted on the basis of the raw scores on the Structured Interview for DSM-IV Personality Disorders (Dutch version: de Jong, Derks, van Oel, & Rinne, 1996; SIDP-IV; Pfohl, Blum, & Zimmerman, 1995). SIDP-IV scores were unavailable for six patients; three patients refused to be interviewed and three patients had undergone a mental health assessment without a SIDP-IV. For these six patients, we conducted a chart review (including

Table 1.   Sample characteristics ($n = 74$).

Characteristic	Sample ($n = 74$)	PCL-R < 26 ($n = 47$)	PCL-R ≥ 26 ($n = 27$)
Age[a]	36.04 (8.85)	36.91 (8.87)	34.55 (8.78)
Time in treatment[a]	3.86 (1.93)	4.19 (2.01)	3.30 (1.68)
Main index offense[b]			
Homicide offense (incl. attempted)	45.9%	46.8%	44.4%
Sexual offense	10.9%	12.7%	7.4%
Violent theft, robbery or assault	35.1%	29.8%	44.4%
Arson	8.1%	10.6%	3.7%
Criminal history			
Prior convictions[b]	73.0%	66.0%	85.2%
Age at 1st conviction[a]	20.66 (6.17)	21.89 (7.00)	18.52 (3.58)*
Prior TBS[b]	6.8%	0%	18.5%**
Mental health history			
Prior contact with mental health services[b]	74.3%	72.3%	77.8%
Age at 1st mental health contact[a]	16.58(9.49)[c]	17.40 (10.03)[d]	15.27 (8.61)[e]

[a]Oneway ANOVA, [b]χ^2-test, [c]$n = 57$, [d]$n = 35$, [e]$n = 22$, *significant difference between the groups at $p < 0.05$. **significant difference between the groups at $p < 0.01$.

previous psychiatric and psychological reports) using the SIDP-IV criteria for DSM-IV personality disorders.

Materials

BEST-index

The BEST-Index (Reed et al., 2000; Dutch version: Van Erven, 1999) is a nurse-rated instrument for the assessment of risk-relevant behaviors among forensic psychiatric patients. The instrument contains 70 items divided a priori among three subscales: risk, insight, and communication & social skills. Detailed descriptions of the BEST-Index subscales and item examples are available elsewhere (Chakhssi et al., 2009; Reed et al., 2000; Ross et al., 2004; Woods, 2000). A previous factor analytic study among 291 Dutch forensic psychiatric patients from dRW revealed an underlying four-factor structure comprising 63 items (Chakhssi et al., 2009). The four factors of the Dutch version of the BEST-Index were: social skills, insight, interpersonal hostility, and physical violence. The social skills factor contains 23 items and is related to social skills and adaptive social behaviors. The insight factor consists of 21 items measuring the level of insight, into the nature of patient's problems; into antecedent events leading to their current situation; and into attribution of responsibility. The interpersonal hostility factor contains 12 items and is related to aggressive and dominant behavior in interpersonal contacts. The fourth factor was labeled physical violence, contains 7 items and is related to physically aggressive acts. Each item can be rated on a five-point scale ranging from 1 (=*worst case*) to 5 (=*optimal case*). Internal consistency of the Dutch version of the BEST-Index and the derived factors was excellent. Cronbach's α for the BEST-Index total score was 0.97, and for the BEST-Index factors internal consistencies were 0.96 (Social skills factor), 0.95 (Insight factor), 0.86 (Interpersonal hostility factor) and 0.74 (Physical violence factor), respectively. Interrater reliability for the Dutch version was also found satisfactory. In a sample of 182 raters (psychiatric nurses trained in the BEST-Index), the average measure intraclass correlation coefficient for the BEST-Index total score was 0.84, and for the BEST-Index factors the coefficients were 0.84 (Social skills factor), 0.80 (Insight factor), 0.82 (Interpersonal hostility factor), and 0.81 (Physical violence factor), respectively. Furthermore, the concurrent and predictive validity of the BEST-Index were supported. Moderate to strong correlations were found between the BEST-Index scales and the Historical Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997), a widely used and extensively validated instrument for assessing violence risk (for an overview see, Douglas, Guy, & Weir,

2006), as well as between the BEST-Index scores and inpatient violence (Chakhssi et al., 2009).

PCL-R

The Psychopathy Checklist – Revised (Hare, 1991, 2003; Dutch version: Vertommen et al., 2002) is a clinical construct rating scale designed to assess psychopathy in forensic populations. The PCL-R consists of 20 items, each rated 0 to 2 (0 = does not apply, 1 = applies to some extent, 2 = definitely applies). The PCL-R yields a dimensional total score between 0 and 40 indicating the degree to which the individual matches the prototypical psychopath. The PCL-R has been extensively studied and significantly predicts general and violent recidivism (Leistico et al., 2008), and it has become the most widely used instrument for assessing psychopathy. In the present study, PCL-R ratings were performed by mental health professionals (i.e. clinical psychologists, psychiatrists, and psychotherapists) who all attended certified training in the administration and coding of the PCL-R.

Procedure

Approval for the study was obtained from the hospital's executive board and the institutional research review committee.

BEST-index assessments

Psychiatric nurses were blind to the purpose of the study and to the PCL-R ratings, although they were not blind to descriptive information on psychopathic features used for treatment planning. The psychiatric nurses were all trained in using the BEST-Index and assessed every patient 6 months after admission, which generally corresponds with the start of the patient's treatment. During their hospital stay, the nurses assessed the patients at 6-month intervals. Every assessment was performed on their observation of patients' behaviors, and collateral information on relevant behaviors provided in the patient's charts, during the previous 6 months on the ward. The BEST-index scoring manual (Dutch version: Van Erven, 1999; Woods, 2000) offers detailed rating criteria for every item. Moreover, the scoring manual offers suggestions for acquiring additional information if items are difficult to score (e.g. conducting interviews with patients on relevant topics). Assessments were performed by at least the primary nurse (mentor) on the ward where the patient resided, and one other nurse from the same ward. If the patient was transferred to another ward, then different nurses assessed the patient at the subsequent time points. However, one nurse was always assigned as the primary nurse of the patient. All the

patients in the study sample had undergone at least four repeated BEST-Index assessments, which were used for the statistical analyses in the study. The baseline assessment (T1) was performed 4.7 months (SD = 1.2) after admission and the mean length of time between the baseline assessment (T1) and the last follow-up assessment used in this study (T4) was 1.68 years (SD = 0.41). The mean scores of the two nurses will be used in this study for the BEST-Index total score (based on 63 items) and factor scores. In total, 206 nurses performed one or more BEST-Index assessments in this study (M = 2.68). Interrater reliability for the raters in this study for the different time-points (T1–T4) was satisfactory. The average measure intraclass correlation coefficient based on absolute agreement for the BEST-Index total score (based on 63 items) ranged from 0.76 to 84 (M = 0.82). For the BEST-Index factors, the coefficients for the social skills factor ranged from 0.75 to 0.83 (M = 0.79), the coefficients for the Insight factor ranged from 0.73 to 0.83 (M = 0.80), for the interpersonal hostility factor they ranged from 0.80 to 0.82 (M = 0.82) and for the physical violence factor from 0.63 to 0.90 (M = 0.81). In the current sample, the reliability coefficients for the BEST-Index total score was 0.96, and for the BEST-Index factors internal consistencies were 0.95 (Social skills factor), 0.94 (Insight factor), 0.87 (Interpersonal hostility factor), and 0.70 (Physical violence factor), respectively.

PCL-R assessments

The PCL-R ratings for every patient were carried out in the course of treatment (Mean = 2.5 years after admission, SD = 1.88) by two trained and independent raters using file information only. Interrater reliability for the PCL-R total score was satisfactory at 0.85 (average measure ICC; n = 135). Patient files in the dRW contain criminal records, psychiatric- and psychological reports for the court, violence risk assessments, medical information, treatment progress reports and reports from psychiatric nurses, social workers, work- and education supervisors. Final PCL-R ratings were based on consensus between the two raters. The PCL-R consensus scores were used in all subsequent data analyses. The PCL-R total score is calculated as the sum of all the item scores, when at least 15 of the 20 items are rated. In case of five or less omitted items, the PCL-R total score is adjusted by prorating, according to the procedure stipulated in the PCL-R Manual (Vertommen et al., 2002). Thirty-eight patients had been assessed with the file-based procedure as well as with the interview-based procedure, and with different raters for the two procedures. The average time between the two procedures was 16.7 months (SD = 9.6). The PCL-R mean score for these 38 patients for the file-based protocol was 23.97 (SD = 3.83, range 15–32), and for the interview-based protocol 23.94 (SD = 4.47, range 13–34). A paired sample *t*-test revealed no significant differences between the two

procedures for the mean PCL-R total scores [$t(37) = 0.05$, $p = 0.96$], for the PCL-R Factor 1 scores [$t(37) = 0.00$, $p = 1.0$] and for the PCL-R Factor 2 scores [$t(37) = 0.84$, $p = 0.41$]. The average measure intraclass correlation coefficient between the scores derived from the file-based- and interview-based procedure was found to be 0.73 for the PCL-R total score, 0.68 for the PCL-R Factor 1 score, and 0.73 for the PCL-R factor 2 scores. Although, file-based procedures may not capture the complete PCL-R affective and interpersonal features, these preliminary findings suggest that the differences between either procedures most likely are small in our sample.

Statistical analysis

Differences in demographic and clinical variables between the psychopathic and nonpsychopathic offenders were tested using analysis of variances (ANOVA) and contingency tables with the Pearson chi-square test or Fisher exact tests, when appropriate. A repeated measures ANOVA was used to examine the changes on the BEST-Index total and subscale scores between T1 and T4. Group (psychopathic *versus* nonpsychopathic offenders) was used as between-subjects factor and the repeated assessments on the BEST-Index (T1, T2, T3, T4) as within-subjects factor.

Individual change

To examine individual change we used the RCI developed by Jacobson and Truax (1991). As stated earlier, the RCI addresses whether patients' change during treatment exceeds the change that would be expected based on measurement error alone. The RCI controls for the degree of difference in scores from pre- to post-test as a result of measurement unreliability. The RCI is defined as (Jacobson & Truax, 1991; p. 14):

$$RCI = \frac{X_2 - X_1}{\sqrt{2(SE)^2}}$$

where X_1 is a subject's pretest score, X_2 is a subject's post-test score and SE is the standard error of measurement. SE is calculated by multiplying the standard deviation of the pretreatment group with the square root of 1 minus the reliability of the assessment measure. The denominator, also described as S_{diff}, reflects 'the amount of difference which one could expect between two scores, obtained on the same test by the same individual, as function of measurement error alone' (Christensen & Mendoza, 1986; p. 307). Under the assumption of normality, 68% of change attributable to measurement error will fall within the range of -1 and 1 S_{diff}, and 95% of change attributable to measurement error will fall within the range of -1.96

and 1.96 S_{diff}. RCI's larger than 1.96 would be unlikely to occur ($p < 0.05$) without actual change. In this study, patients with a negative reliable change (RCI < -1.96) will be labeled as *reliably deteriorated*, patients within the band of no reliable change ($-1.96 \leq$ RCI ≤ 1.96) will be labeled as *uncertain change*, and patients with a positive reliable change (RCI > 1.96) will be labeled as *reliably improved* (see Appendix for an example of calculating RCI). All analyses are performed using the Statistical Package for the Social Sciences, version 13 (SPSS; SPSS Inc., 2006).

Results

In this sample, the most prevalent personality disorder (PD) for both the nonpsychopathic and the psychopathic group was Antisocial PD (44.7% versus 81.5%; see Table 2). Group differences on DSM-IV Axis I and Axis II diagnosis were found in the prevalence of Antisocial personality disorder [χ^2 (df = 1, N = 74) = 9.54; $p = .003$] and Narcissistic personality disorder [χ^2 (df = 1, N = 74) = 13.46; $p = .000$]. Tests of between group differences

Table 2. DSM-IV diagnoses.

	PCL-R < 26 ($n = 47$)		PCL-R ≥ 26 ($n = 27$)	
	n	$\%$	n	$\%$
DSM-IV axis I disorder				
Substance related disorders	37	78.7	25	92.6
Attention deficit and disruptive behavior disorders	10	21.3	7	25.9
Impulse control disorders	7	14.9	2	7.4
Mood disorders	5	10.6	0	0
Anxiety disorders	4	8.5	1	3.7
Learning disorders	0	0	1	3.7
Any axis I disorder	40	85.1	26	96.3
Any axis I disorder without substance abuse	22	46.8	10	37.0
DSM-IV axis II disorder				
Antisocial PD	21	44.7	22	81.5**
Borderline PD	6	12.8	7	25.9
Narcissistic PD	0	0	7	25.9**
Paranoid PD	4	8.5	2	7.4
Schizoid PD	4	8.5	2	7.4
Obsessive compulsive PD	2	4.3	1	3.7
Avoidant PD	2	4.3	0	0
Dependent PD	1	2.1	0	0
Any PD	23	48.9	24	88.9

PD = personality disorder. **significant difference between the groups at $p < 0.01$, χ^2-test (two-tailed).

on demographic and judicial variables shows that the psychopathic group was significantly younger at the time of their first conviction [$F(1,73) = 5.44$, $p = .022$] and they had a significantly higher number of prior TBS-orders [χ^2 (df = 1, N = 74) = 9.35; $p = .005$].

The mean PCL-R score was 23.10 (SD = 6.40). For the PCL-R facets as defined by Hare (2003), the mean score for the Interpersonal facet was 4.0 (SD = 2.24), 5.73 (SD = 1.82) for the Affective facet, 5.65 (SD = 2.15) for the Impulsive lifestyle facet and 5.85 (SD = 2.70) for the Antisocial facet (see Table 3). The mean scores on the BEST-Index total was 234.74 (SD = 32.50). The mean scores for the BEST-Index factors at T1 were 91.26 (SD = 14.56) for the Social skills factor, 62.69 (SD = 14.89) for the Insight factor, 47.48 (SD = 7.48) for the Interpersonal hostility factor and 33.30 (SD = 2.70) for the Physical violence factor (see Table 4).

Table 3.   PCL-R mean and facet scores for the non-psychopathic and psychopathic patients.

Scale	PCL-R < 26 (n = 47) Mean (SD)	PCL-R ≥ 26 (n = 27) Mean (SD)
PCL-R total	19.35 (4.32)	29.61 (3.51)
Facet 1: interpersonal	3.34 (2.14)	5.15 (1.97)
Facet 2: affective	5.19 (1.78)	6.67 (1.52)
Facet 3: lifestyle	4.83 (1.86)	7.07 (1.88)
Facet 4: antisocial	4.74 (2.48)	7.78 (1.89)

All differences between psychopaths and nonpsychopaths in PCL-R mean scores are significant at $p < 0.001$ (two-tailed, *t*-test).

Table 4.   Mean scores and standard deviations on the different time points for the psychopathic and nonpsychopathic patients.

Scales	PCL-R score	Mean T1 (SD)	Mean T2 (SD)	Mean T3 (SD)	Mean T4 (SD)
BEST total	<26	232.77 (34.52)	245.54 (34.09)	243.92 (32.11)	256.12 (33.53)
	≥26	238.17 (28.95)	238.02 (26.26)	245.28 (34.16)	245.82 (37.55)
Social skills	<26	89.60 (15.97)	94.76 (13.27)	94.19 (13.00)	97.84 (14.67)
	≥26	94.17 (11.40)	93.76 (9.83)	96.76 (11.53)	96.98 (13.20)
Insight	<26	61.53 (15.51)	69.97 (15.50)	68.30 (15.42)	74.12 (16.29)
	≥26	64.70 (13.80)	66.48 (12.46)	71.46 (16.90)	70.81 (17.13)
Interpersonal hostility	<26	48.26 (7.15)	47.80 (7.44)	48.01 (6.92)	50.30 (5.70)
	≥26	46.11 (7.96)	45.39 (7.14)	45.37 (8.16)	46.46 (8.15)
Physical violence	<26	33.37 (2.56)	33.02 (2.87)	33.43 (2.73)	33.86 (1.86)
	≥26	33.18 (2.97)	32.39 (3.01)	31.69 (4.11)	31.56 (3.60)

The higher the BEST-index score, the more favorable patients performed on the scale. The mean length of time between T1 and T2 was 6.2 months (SD = 1.9), between T2 and T3 was 5.8 months (SD = 2.1) and between T3 and T4 was 5.9 months (SD = 1.9).

The most prevalent personality disorder (PD) for both the non-psychopathic and the psychopathic group was antisocial PD (44.7% *versus* 81.5%; see Table 2). Group differences on DSM-IV Axis I and Axis II diagnosis were found in the prevalence of Antisocial personality disorder [χ^2(df = 1,   N = 74) = 9.54;   p = 0.003] and Narcissistic personality disorder [(χ^2(df = 1, N = 74) = 13.46; p = 0.000]. Tests of between group differences on demographic and judicial variables show that the psychopathic group was significantly younger at the time of their first conviction [F(1, 73) = 5.44, p = 0.022] and they had a significantly higher number of prior TBS-orders [χ^2(df = 1,   N = 74) = 9.35; p = 0.005].

Change during treatment at the group level

The results of the repeated measures ANOVA are presented in Table 4. The means and standard deviations at the four time points are presented for the psychopaths (PCL-R $\geq$ 26) and nonpsychopaths (PCL-R < 26) on the BEST-Index factors and total score. On the BEST-Index total score, there was no difference between the scores of psychopaths and nonpsychopaths (no main effect for Group: F(1, 72) = 0.19, p = 0.668). The BEST-Index total score improved during treatment for both the psychopaths and nonpsychopaths (main effect for Time: F(3, 216) = 5.54, p = 0.001). The interaction between Group and Time did not reach statistical significance, F(3, 216) = 1.82, p = 1.44. For the BEST-Index Social skills factor, the sphericity assumption was not met and the degrees of freedom were corrected using the Huynh-Feldt correction (ϵ = 0.93). The main effect of time for the social skills factor was statistically significant, F(2.80, 201.45) = 4.13, p = 0.009 (with Huynh-Feldt correction). The main effect of Group was not significant, F(1, 72) = 0.27, p = 0.606, and the Group $\times$ Time interaction for the Social skills factor was also not significant, F(2.80, 201.45) = 1.43, p = 0.238 (with Huynh-Feldt correction). For the BEST-Index Insight factor, the main effect of time was significant, F(3, 216) = 9.43, p = 0.000. Again, the main effect of Group was not significant, F(1, 72) = 0.00, p = 0.970, and the Group $\times$ Time interaction for the Insight factor was also not significant, F(3, 216) = 2.18, p = 0.096.

For the BEST-Index Interpersonal hostility factor, the psychopaths scored lower than the nonpsychopaths (main effect of Group: F (1, 72) = 4.30, p = 0.042), indicating that psychopaths showed more interpersonally hostile behavior. Follow-up t tests revealed that psychopaths scored lower than nonpsychopaths at Time 4, t(40.84) = 2.161, p = 0.037. There were no differences between the mean scores of psychopaths on non-psychopaths at the other time points [Time 1, t(72) = 1.197, p = 0.235; Time 2, t(72) = 1.360, p = 0.178; Time 3,

$t(72) = 1.479$, $p = 0.144$]. There was no main effect of Time [$F(3, 216) = 1.59$, $p = 0.193$], and no significant Group × Time interaction [$F(3, 216) = 0.33$, $p = 0.806$] for the interpersonal hostility factor. For the physical violence factor, the main effect of Group was significant, $F(1, 72) = 4.78$, $p = 0.032$. Follow-up t tests revealed that psychopaths scored lower than nonpsychopaths at Time 4, t $(34.11) = 3.104$, $p = 0.004$, meaning that the psychopaths showed more physical violence at Time 4. And again, there were no differences at the other time points [Time 1, $t(72) = 0.286$, $p = 0.776$; Time 2, $t(72) = 0.895$, $p = 0.374$; Time 3, $t(39.43) = 1.96$, $p = 0.057$]. The main effect of time was not significant for the physical violence factor, $F(3, 216) = 1.69$, $p = 0.170$, but the Time × Group interaction was significant, $F(3, 216) = 3.96$, $p = 0.009$. *Post hoc* analyses using repeated measures analysis of variance with polynomial contrasts for Group × Time suggest that this interaction is due to a significant linear trend ($F(1,72) = 9.546$, $p = 0.003$), indicating that the change slope over the four time points for the two groups are in opposite direction of each other. There was no main effect of time and no significant simple effects, for the Physical violence scores, but an inspection of the means shows a small improvement in Physical violence scores for the nonpsychopaths, as well as a slight deterioration in these scores for the psychopaths.

Table 5. Reliable change for the BEST total and subscales from T1 to T4 ($n = 74$).

Scales	Reliable deterioration		Uncertain change		Reliable improvement	
	n	%	n	%	n	%
BEST total						
PCL < 26	3	6.4	16	34.0	28	59.6
PCL ≥ 26	2	7.4	15	55.6	10	37.0
Social skills						
PCL < 26	3	6.4	24	51.1	20	42.6
PCL ≥ 26	3	11.1	15	55.6	9	33.3
Insight						
PCL < 26	4	8.5	14	29.8	29	61.7
PCL ≥ 26	2	7.4	15	55.6	10	37.0
Interpersonal hostility						
PCL < 26	3	6.4	36	76.6	8	17.0
PCL ≥ 26	3	11.1	18	66.7	6	22.2
Physical violence						
PCL < 26	0	0	43	91.5	4	8.5
PCL ≥ 26	6**	22.2	19	70.4	2	7.4

**$p < 0.01$ (fisher exact tests, two-tailed).

Change during treatment at the individual level

Table 5 presents the results of individual level change using the RCI, including the number and percentage of psychopaths and nonpsychopaths that reliably deteriorated, showed uncertain change or reliably improved. Differences in the observed and expected distribution of reliable change in psychopaths and nonpsychopaths were tested using the Fisher exact tests for association in the 3×2 contingency table. The Fisher exact tests was not statistically significant for the BEST-Index total score ($p = 0.166$), for the social skills factor ($p = 0.623$), for the insight factor ($p = 0.114$), and for the interpersonal hostility factor ($p = 0.683$). The Fisher exact tests was statistically significant for the physical violence factor ($p = 0.003$). A *post hoc* Fisher exact tests in the 2×2 contingency table (deterioration *versus* uncertain change/improvement) revealed that a greater proportion of psychopaths reliably deteriorated on the Physical violence factor during treatment ($p = 0.002$). No association was found between psychopathy and reliable improvement in the *post hoc* Fisher exact tests (deterioration/ uncertain change *versus* improvement), $p = 1.000$. *Post hoc* analyses were performed to assess whether the difference between the psychopaths who reliably improved and the psychopaths who showed no reliable change and/ or deteriorated could be explained by different scores on the four PCL-R facets. No significant differences within the psychopathic group were found on the PCL-R facet mean scores (deterioration/uncertain change *versus* improvement). Also no significant differences within the psychopathic were found on the sample characteristics (see Table 1) when we compared the psychopathic patients who improved *versus* psychopathic patients who showed no change or showed deterioration during treatment on the BEST-index total and factors.

Discussion

In the present study we examined treatment responsiveness in a sample of psychopathic and nonpsychopathic offenders in an inpatient forensic psychiatric hospital. We hypothesized that there would be no group differences in treatment responsiveness, and that psychopaths and non-psychopaths would show similar percentages of reliable individual change. To the best of our knowledge, no earlier studies have examined individual change in forensic patients, where the level of psychopathy was assessed with Hare's PCL-R.

Does treatment makes psychopath worse?

In contrast to clinical lore, treatment in this study did not make all or a majority of psychopaths worse. In terms of individual change, psychopaths

and nonpsychopaths showed overall the same pattern of reliable deterioration on the BEST total score. Acts of physical violence were an exception, however, nearly a quarter (22%) of psychopaths showed reliable deterioration, whereas nonpsychopaths showed no reliable deterioration in physical violence during treatment. This finding is in line with previous studies demonstrating that psychopaths are more likely to show inpatient aggression during hospitalization (e.g. Hildebrand et al., 2004) and that psychopaths are more likely to show negative behaviors during treatment than nonpsychopaths (Hobson, Shine, & Roberts, 2000). Hobson et al. examined 104 inmates with the PCL-R and a negative behavior checklist during 6 months in a prison-based therapeutic community. They found that psychopathy was significantly associated with poor adjustment to the therapeutic community as demonstrated by disruptive behaviors during therapy sessions, on the ward and during other activities (e.g. 'manipulates others'; 'tells lies'; 'inflated sense of self-importance'). On the basis of their findings, the authors suggest that the problematic behaviors of psychopaths could be explained by a lower level of treatment involvement due to the interpersonal and affective features of psychopathy. Similar to these findings, Morrissey, Mooney, Hogue, Lindsay, and Taylor (2007) found that psychopathy, and especially the interpersonal and affective facets, predicted inpatient movements from medium to high hospital security conditions in a sample of intellectual disabled offenders ($N = 73$). In our sample – which was rather small to conduct *post hoc* analyses – we did not find differences on the four facet scores between the psychopaths who deteriorated and those who improved. Given the findings of brain imaging studies in psychopaths (Raine, Lencz, Bihrle, Lacasse, & Coletti, 2000), an additional explanation could be that these – deteriorating – psychopaths may have poorer functioning in brain regions (e.g. prefrontal cortex) involved in regulating and controlling aggressive behavior. Overall, there were significant differences on physical violence (in group and individual level analyses), and on interpersonal hostility (in group level analyses only).

Do psychopaths improve with treatment?

First, we found no differences between psychopaths and nonpsychopaths on the BEST-Index scales at baseline assessment. Whilst differences between psychopathic and nonpsychopathic patients on the baseline measurement BEST-Index scales are to be expected, few studies have addressed pre-treatment differences between PCL-R psychopaths and nonpsychopaths. We could retrieve only one PCL-R based study in a comparable forensic psychiatric sample (Hildebrand & de Ruiter, submitted for publication) addressing pre-treatment differences between psychopaths ($n = 27$) and nonpsychopaths ($n = 70$) on several indirect and self-report measures of dynamic risk, including insight and aggression. Hildebrand and de Ruiter found no pre-treatment

differences between psychopathic and nonpsychopathic patients on measures of negative attitudes, egocentrism, impulsivity and lack of insight. They did, however, find pre-treatment differences between psychopaths and non-psychopaths on two out of six indicators of anger/hostility. Second, in line with previous studies our results suggest that some psychopaths do improve with treatment (Olver & Wong, 2009; Skeem, 2008; Skeem et al., 2002). The group and individual level analyses suggest that both psychopaths and nonpsychopaths are responsive to inpatient forensic treatment, demonstrated by significant changes on the BEST-Index total score, and on the BEST-Index factors measuring social skills and insight. Psychopaths, as well as non-psychopaths, seem to change on BEST-Index scales that measure adaptive social behavior, communication skills, level of insight and attribution of responsibility. More specifically, approximately one third of the psychopaths showed reliable improvement on the BEST-index total score, and the BEST-index factors measuring social skills and insight. Again, these findings contradict the widely held belief that psychopaths cannot change (Cleckley, 1988). On the other hand, there is some indication that the non-psychopathic patients improved somewhat more than the psychopathic patients. Overall, both the group and individual level analyses showed few significant differences between psychopaths and non-psychopaths in terms of change.

Overall, our findings confirm our hypothesis that within the group of psychopaths treatment responsiveness varies and underlines our notion that individual variability within psychopaths is masked in group level analyses. Consistent with some previous research (Olver & Wong, 2009; Skeem, 2008; Skeem et al., 2002) our findings demonstrate that treatment does not make all or most psychopaths worse, but also that there are differences between psychopaths and non-psychopaths in terms of change during treatment.

Limitations of the study

The findings of the current study should be considered with several limitations in mind. The generalizability of the findings is limited to inpatient male forensic psychiatric patients with personality disorders. The study should be replicated in larger and different samples (e.g. forensic psychiatric patients with psychotic disorders) to assess the strength of our findings. In addition, the observation period in this study ($M = 1.67$ years) is roughly one third of the regular treatment duration under the Dutch TBS order ($M = 5.9$ years; Wartna, el Harbachi, & Essers, 2006). Research with longer observation periods should be performed to determine if psychopathic patients remain responsive to forensic psychiatric treatment over time. One limitation of the current study is the lack of a no-treatment control group, inherent to forensic clinical outcome studies where withholding treatment is seen as unwarranted and/or unethical. This methodological limitation indicates that any observed differences, or in this study the lack of differences, in change between the

psychopathic and non-psychopathic offenders could be due to natural changes in the subjects over time rather than to effects of treatment. However, giving the documented behaviors in psychopaths of poor adjustment to therapeutic communities (e.g. Hobson, Shine, & Roberts, 2000; Ogloff, Wong, & Greenwood, 1990), increased aggressive behaviors during treatment (e.g. Hildebrand et al., 2004) and higher attrition rates (e.g. Ogloff et al., 1990; Olver & Wong, 2009), we believe that improvements shown by psychopathic patients in our study are unlikely to be due to solely to the passage of time. However, firm conclusions on the observed change in psychopathic as well non-psychopathic offenders can only be drawn in relation to a no-treatment control group. Another limitation was that we could not compare the psychopathic and nonpsychopathic patients on actual therapy involvement and intensity of treatment. No reliable data on treatment intensity and involvement were available for the observation period. However, the treatment programs for personality disordered offenders in the dRW consist generally of the same treatment modalities (see Method section) and there is no separate treatment policy for psychopathic patients. Future research should take specific treatment objectives and strategies into account to examine the differences in responsiveness between psychopathic and nonpsychopathic patients. Future research should also aim at examining how treatment responsiveness of psychopathic patients is related to change in future violence risk. At this time, we cannot say what bearing these findings have on future offending behavior.

Furthermore, we used a behaviorally-based measure of forensic risk, the BEST index. Although, we had a large sample of raters ($n = 206$) in this study with satisfactory interrater reliability, we cannot be certain if and how habituation of the staff to patients' behavior effected the observed change. However, given the large sample of raters and different raters over time who rated each patient, we believe that possible habituation effects are minimal. Also, we do not know to what extent inferences based on observable behaviors correspond to changes in underlying personality constructs such as lack of empathy, lack of remorse, or impulsivity (Blair, 2003; Cooke & Michie, 2001). There is considerable evidence suggesting that callous-unemotional traits form the temperamental core of psychopathy, that these traits are under strong genetic control, and that they are rooted in neurobiological deficits in limbic (e.g. amygdale) and other brain regions (Blair, 2003; Raine et al., 2000). We do not know whether behavioral improvements exhibited by some psychopaths during treatment are associated with corresponding changes in cognitive, affective, or neurobiological mechanisms that are thought to underlie psychopathy. Future studies should attempt to measure changes in these domains during forensic treatment, using experimental paradigms such as tests of implicit cognition (Gray, MacCulloch, Smith, Morris, & Snowden, 2003), psychophysiological responding, and brain imaging procedures.

We also cannot say with certainty that some of the improvement seen in psychopathic patients was not 'faked'. Psychopaths are known to be extremely adept at impression management. For this reason, we avoided using self-report measures of treatment responsiveness that are known to be susceptible to this kind of manipulation. However, it is still conceivable that some of the apparent improvements in areas such as insight and social skills were based partly or entirely on the pseudo-adaptation of psychopaths who had learned to produce socially desirable behaviors to meet the expectations of their treatment providers. On the other hand, the fact that the vast majority of psychopathic patients (>77.8%) in the study showed little or no physical aggression or interpersonal hostility during the nearly 2-year course of treatment suggests that many of these patients are able to exert genuine self-control over their behavior, at least within a structured forensic psychiatric setting. Nevertheless, we will only be able to make more definitive statements about the genuineness of these changes when we examine the long-term outcomes of these patients with respect to recidivism.

Clinical implications

Our findings, as well as those of some previous studies (e.g. Barbaree, 2005; D'Silva et al., 2004; Olver & Wong, 2009; Salekin, 2002; Skeem, 2008; Skeem et al., 2002), suggest that treatments should not be routinely withheld from forensic patients based on their high PCL-R scores. Although psychopaths do show higher rates of recidivism than other forensic patients, this does not imply that they are untreatable. Indeed, a growing body of research suggests that psychological services, such as psychotherapy and addiction treatment programs, can be beneficial for many forensic patients, including psychopathic ones. In the future, it will be important to predict which psychopathic patients are able to benefit from forensic treatments, *versus* those who deteriorate or remain unchanged, so that treatment decisions can be made on an empirically supported and cost-effective basis. However, at this time, the blanket decision to withhold treatment from psychopathic patients in general does not seem to be supported by research findings.

References

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

Atkins, D.C., Bedics, J., McGlinchey, J.B., & Beauchaine, T.P. (2005). Assessing clinical significance: Does it matter which method we use? *Journal of Consulting and Clinical Psychology*, 73, 982–989.

Barbaree, H.E. (2005). Psychopathy, Treatment behavior, and recidivism: An extended follow-up of Seto and Barbaree. *Journal of Interpersonal Violence*, 20, 1115–1131.

Blair, R.J.R. (2003). Neurobiological basis of psychopathy. *British Journal of Psychiatry*, *182*, 5–7.

Brinkley, C.A., Newman, J.P., Widiger, T.A., & Lynam, D.R. (2004). Two approaches to parsing the heterogeneity of psychopathy. *Clinical Psychology: Science and Practice*, *11*, 69–94.

Chakhssi, F., de Ruiter, C., & Bernstein, D. (2009). Reliability and validity of the Dutch version of the behavioural status index: A nurse-rated forensic assessment tool. *Assessment*, Prepublished online August 21, 2009; doi: 10.1177/1073191109338815.

Christensen, L., & Mendoza, J.L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy*, *15*, 305–308.

Cleckley, H.M. (1988). *The mask of sanity: An attempt to clarify some issues about the so called psychopathic personality* (5th ed.). Augusta: Emily S. Cleckley.

Cooke, D.J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment*, *13*, 171–188.

de Jong, C.A.J., Derks, F.C.H., van Oel, C.J., & Rinne, T. (1996). *Gestructureerd interview voor de DSM-IV Persoonlijkheidsstoornissen. [Structured interview for DSM-IV personality disorders]*. Sint Oedenrode, the Netherlands: Stichting Verslavingszorg Oost Brabant.

Douglas, K.S., Guy, L.S., & Weir, J. (2006). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. Retrieved May 26, 2008, from http://kdouglas.wordpress.com/hcr-20/

D'Silva, K., Duggan, C., & McCarthy, L. (2004). Does treatment really make psychopaths worse? A review of the evidence. *Journal of Personality Disorders*, *18*, 163–177.

Gray, N.S., McCulloch, M.J., Smith, J., Morris, M., & Snowden, R.J. (2003). Violence viewed by psychopathic murderers: Adapting a revealing test may expose those psychopaths who are most likely to kill. *Nature*, *423*, 497–498.

Goldstein, A.P., Glick, B., Reiner, S., Zimmerman, D., & Coultry, T. (1987). *Aggression replacement training*. Illinois: Research Press.

Hare, R.D. (1991). *Hare psychopathy checklist-revised (PCL-R)*. Toronto, Canada: Multi-Health Systems.

Hare, R.D. (2003). *Hare psychopathy checklist-revised (PCL-R)* (2nd ed). Toronto, Canada: Multi-Health Systems.

Hare, R.D., Clark, D., Grann, M., & Thornton, D. (2000). Psychopathy and the predictive validity of the PCL-R: An international perspective. *Behavioral Sciences and the Law*, *18*, 623–645.

Harris, G.T., & Rice, M.E. (2006). Treatment of psychopathy: A review of empirical findings. In C.J. Patrick (Ed.), *Handbook of psychopathy* (pp. 555–572). New York: Guilford Press.

Hemphill, J., Hare, R., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal Criminological Psychology*, *3*, 141–172.

Hildebrand, M., & de Ruiter, C. (2004). PCL-R psychopathy and its relation to DSM-IV Axis I and Axis II disorders in a sample of male forensic psychiatric patients in the Netherlands. *International Journal of Law and Psychiatry*, *27*, 233–248.

Hildebrand, M., & de Ruiter, C. PCL-R psychopathy and change on indicators of dynamic risk factors during inpatient forensic psychiatric treatment. Manuscript submitted for publication.

Hildebrand, M., de Ruiter, C., & de Vogel, V. (2004). Psychopathy and sexual deviance in treated rapists: Association with sexual and nonsexual recidivism. *Sexual Abuse: A Journal of Research and Treatment*, *16*, 1–24.

Hildebrand, M., de Ruiter, C., & Nijman, H. (2004). PCL-R psychopathy predicts disruptive behavior among male offenders in a Dutch forensic psychiatric hospital. *Journal of Interpersonal Violence*, *19*, 13–29.

Hildebrand, M., Hesper, B.L., Spreen, M., & Nijman, H.L.I. (2005). *De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie [The value of structured risk assessment and the diagnosis of psychopathy]*. Utrecht: EFP.

Hobson, J., Shine, J., & Roberts, R. (2000). How do psychopaths behave in a prison therapeutic community? *Psychology, Crime and Law*, *6*, 139–154.

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.

Laws, D.R., Hudson, S.M., & Ward, T. (2000). *Remaking relapse prevention with sex offenders: A sourcebook*. Thousand Oaks: Sage.

Leistico, A.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.

Liberman, R.P., DeRisi, W.J., & Mueser, K.T. (1989). *Social skills training for psychiatric patients*. New York: Pergamon Press.

McGlinchey, J.B., Atkins, D.C., & Jacobson, N.S. (2002). Clinical significance methods: Which one to use and how useful are they? *Behavior Therapy*, *33*, 529–550.

Miller, W.R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people to change*. New York: Guilford Press.

Morrissey, C., Mooney, P., Hogue, T.E., Lindsay, W.R., & Taylor, J.L. (2007). Predictive validity of the PCL-R for offenders with intellectual disability in a high security hospital: Treatment progress. *Journal of Intellectual and Developmental Disability*, *32*, 125–133.

Ogles, B.M., Lunnen, K.M., & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review*, *21*, 421–446.

Ogloff, J.R.D., Wong, S., & Greenwood, A. (1990). Treating criminal psychopaths in a therapeutic community program. *Behavioral Sciences and the Law*, *8*, 181–190.

Olver, M.E., & Wong, S. (2009). Therapeutic responses of psychopathic sexual offenders: Treatment attrition, therapeutic change, and long term recidivism. *Journal of Consulting and Clinical Psychology*, *77*, 328–336.

Patrick, C.J. (Ed.). (2006). *Handbook of psychopathy*. New York: Guilford Press.

Pfohl, B., Blum, N., & Zimmerman, M. (1997). *Structured interview for DSM-IV personality: SIDP-IV*. Washington, DC: American Psychiatric Press.

Raine, A., Lencz, T., Bihrle, S., LaCasse, L., & Colletti, P. (2000). Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Archives of General Psychiatry*, *57*, 119–127.

Reed, V., Woods, P., & Robinson, D. (2000). *Behavioural status index (BEST-Index): A 'life skills' assessment for selecting and monitoring therapy in mental health care*. UK: Psychometric Press.

Rice, M.E., Harris, G.T., & Cormier, C.A. (1992). An evaluation of a maximum-security therapeutic community for psychopaths and other mentally disordered offenders. *Law and Human Behavior*, *16*, 399–412.

Ross, T., Woods, P., Reed, V., Sookoo, S., Dean, A., Kettles, A.M. ..., Pfäfflin, F. (2008). Selecting and monitoring living skills in forensic mental health care: Cross-border validation of the BEST-index. *International Journal of Mental Health*, *36*, 3–16.

Salekin, R.T. (2002). Psychopathy and therapeutic pessimism: Clinical lore or clinical reality? *Clinical Psychology Review*, *22*, 79–112.

Seto, M.C., & Barbaree, H.E. (1999). Psychopathy, treatment behavior, and sex offenders recidivism. *Journal of Interpersonal Violence*, *14*, 1235–1248.

Skeem, J. (2008, July). *High risk, not hopeless: Recent research on treating individuals with psychopathy*. Paper presented at the annual meeting of the International Association of Forensic Mental Health, Vienna, Austria.

Skeem, J., Monahan, J., & Mulvey, E. (2002). Psychopathy, treatment involvement, and subsequent violence among civil psychiatric patients. *Law and Human Behavior*, *26*, 577–603.

Smeijsters, H., & Cleven, G. (2006). The treatment of aggression using arts therapies in forensic psychiatry: Results of a qualitative inquiry. *The Arts in Psychotherapy*, *33*, 37–58.

SPSS Inc. (2005). *SPSS 13.0 for windows*. [Computer software]. Chicago, IL: SPSS Inc.

Van Erven, A. (1999). *Behavioural status index, Dutch version, (BSI-D). A broad risk assessment for mental health workers*. Eindhoven: GGzE.

Vertommen, H., Verheul, R., de Ruiter, C., & Hildebrand, M. (2002). *Hare's psychopathie checklist: Handleiding [Dutch version of the Hare Psychopathy Checklist-Revised- Manual]*. Lisse: Swets Test Publishers.

Walters, G.D. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factor scores: A meta-analysis. *Law and Human Behavior*, *27*, 541–558.

Wartna, B.S.J., Harbachi, S. el., & Essers, A.A.M. (2006). *Strafrechtelijke recidive ex-terbeschikkinggestelden: Fact sheet 2006-8 [Criminal recidivism of ex-forensic psychiatric patients detained under the Dutch TBS-order]*. Den Haag: WODC.

Webster, C.D., Douglas, K.S., Eaves, D., & Hart, S.D. (1997). *HCR-20. Assessing risk for violence, version 2*. Burnaby, BC: Simon Fraser University, Mental Health, Law and Policy Institute.

Woods, P. (2000). *The behavioural status index: Descriptive studies within a forensic context* (PhD thesis). Faculty of Health and Social Studies, Anglia Polytechnic University.

Young, J.E. (1994). *Cognitive therapy for personality disorders: A schema-focused approach* (2nd ed.). Sarasota, FL: Professional Resource Press.

Appendix: Calculation of the reliable change index (Jacobson & Truax, 1991), and an example

To address whether patients' change during treatment exceeds the change that would be expected based on measurement error alone, Jacobson and Truax (1991) developed the RCI:

$$RCI = \frac{X_2 - X_1}{\sqrt{2(SE)^2}}$$

where $X_1 =$ subject's pretest score; $X_2 =$ subject's post-test score and $SE = s_1\sqrt{1 - r_{xx}}$; $s_1 =$ standard deviation of the sample at pretreatment; $r_{xx} =$ reliability of the measure. With a subject's BEST-Index total score of 232 at T1, a total score of 256 at T4, standard deviation of the total sample of 32.50 at pretreatment and the internal consistency of 0.97 (Cronbach's α) for the BEST-Index total score, as reported by Chakhssi et al. (2009), the RCI is calculated as:

$$\text{SE} = s_1\sqrt{1 - r_{xx}} = 32.50\sqrt{1 - 0.97} = 5.63$$

$$\sqrt{2(5.63)^2} = 7.96$$

$$\text{RCI} = 256 - 232/7.96 = 3.02$$

In this example, the subject has changed beyond the range that would be explained by measurement error alone. The RCI is 3.02, which is greater than 1.96, and this change is reliable at $p < 0.05$.