Psychological Medicine

cambridge.org/psm

Schema therapy for violent PD offenders: a randomized clinical trial

David P. Bernstein[1], Marije Keulen-de Vos[2], Maartje Clercx[2], Vivienne de Vogel[3], Gertruda C. M. Kersten[2,4], Marike Lancel[5,6], Philip P. Jonkers[7], Stefan Bogaerts[8,9], Mariëtte Slaats[5,10,11], Nick J. Broers[12], Thomas A. M. Deenen[11] and Arnoud Arntz[1,13]

[1]Department of Clinical Psychological Science, Maastricht University, Maastricht, the Netherlands; [2]FPC de Rooyse Wissel, Venray, the Netherlands; [3]Van der Hoeven kliniek, Utrecht, the Netherlands; [4]De Akkerdistel, Beuningen, the Netherlands; [5]FPK Assen, Assen, the Netherlands; [6]Rijksuniversiteit Groningen, Groningen, the Netherlands; [7]FPC Oostvaarderskliniek, Almere, the Netherlands; [8]Social and Behavioral Sciences, Tilburg University, Tilburg, the Netherlands; [9]FPC De Kijvelanden, Poortugaal, the Netherlands; [10]FPC Dr S. Van Mesdag, Groningen, the Netherlands; [11]FPC Veldzicht, Balkbrug, the Netherlands; [12]Department of Methodology and Statistics, Maastricht University, the Netherlands and [13]Department of Clinical Psychology, University of Amsterdam, Amsterdam, the Netherlands

Original Article

Cite this article: Bernstein DP *et al* (2023). Schema therapy for violent PD offenders: a randomized clinical trial. *Psychological Medicine* **53**, 88–102. https://doi.org/10.1017/S0033291721001161

Received: 24 September 2020
Revised: 12 March 2021
Accepted: 15 March 2021
First published online: 15 June 2021

Key words:
Schema Therapy; psychotherapy; treatment; forensic; Antisocial Personality Disorder; personality disorders; psychopathy; violence; aggression; rehabilitation; randomized clinical trial; treatment outcomes; recidivism

Author for correspondence:
David P. Bernstein,
E-mail: d.bernstein@maastrichtuniversity.nl

Abstract

Background. Violent criminal offenders with personality disorders (PD's) can cause immense harm, but are often deemed untreatable. This study aimed to conduct a randomized clinical trial to test the effectiveness of long-term psychotherapy for rehabilitating offenders with PDs.
Methods. We compared schema therapy (ST), an evidence-based psychotherapy for PDs, to treatment-as-usual (TAU) at eight high-security forensic hospitals in the Netherlands. Patients in both conditions received multiple treatment modalities and differed only in the individual, study-specific therapy they received. One-hundred-three male offenders with antisocial, narcissistic, borderline, or paranoid PDs, or Cluster B PD-not-otherwise-specified, were assigned to 3 years of ST or TAU and assessed every 6 months. Primary outcomes were rehabilitation, involving gradual reintegration into the community, and PD symptoms.
Results. Patients in both conditions showed moderate to large improvements in outcomes. ST was superior to TAU on both primary outcomes – rehabilitation (i.e. attaining supervised and unsupervised leave) and PD symptoms – and six of nine secondary outcomes, with small to moderate advantages over TAU. ST patients moved more rapidly through rehabilitation (supervised leave, treatment*time: $F(5308) = 9.40$, $p < 0.001$; unsupervised leave, treatment*-time: $F(5472) = 3.45$, $p = 0.004$), and showed faster improvements on PD scales (treatment*-time: $t(1387) = -2.85$, $p = 0.005$).
Conclusions. These findings contradict pessimistic views on the treatability of violent offenders with PDs, and support the effectiveness of long-term psychotherapy for rehabilitating these patients, facilitating their re-entry into the community.

Violent offenders with personality disorders (PDs) can cause immense harm to victims and society. They commit more serious and violent crimes (Blackburn, 1998; Johnson et al., 2000), more repeat offenses, and receive longer sentences (Hart, Webster, & Menzies, 1993) than other offenders. Antisocial, narcissistic, borderline, and paranoid PDs are the most prevalent PDs in correctional settings (Blackburn, Logan, Donnelly, & Renwick, 2003; Lindsay et al., 2006; Nijman, Cima, & Merckelbach, 2003), and are associated with high recidivism. Offenders with psychopathy, the severest subgroup of antisocial PD, are at particularly high risk for reoffending (Serin, 1996). Incarceration is estimated to cost $80 billion per year in the US (Lewis & Lockwood, 2019), where per capita incarceration is nine times higher than in the Netherlands (Walmsley, 2013). Reducing recidivism in offenders with PDs and aggression could therefore have significant economic as well as societal benefits (Settumba, Chambers, Shanahan, Schofield, & Butler, 2018).

Treatment is used increasingly to lower recidivism risk (Larsen, Jalava, & Griffiths, 2020; Papalia, Spivak, Daffern, & Ogloff, 2019; Wilson, 2014). However, standard treatments for offenders with PD and aggression, usually based on cognitive-behavioral therapy (CBT), only have modest effects (Ross, Quayle, Newman, & Tansey, 2013; Wilson, 2014; Wong, Gordon, Gu, Lewis, & Olver, 2012). Moreover, their focus is on controlling aggressive behavior (Timmerman & Emmelkamp, 2005), but not ameliorating PDs associated with (violent) recidivism. In fact, offenders with severe PDs, especially psychopathic ones, are often assumed

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (http://creativecommons.org/licenses/by-nc-sa/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

CAMBRIDGE UNIVERSITY PRESS

impossible to treat (Rice, Harris, & Cormier, 1992), despite the lack of methodologically sound randomized clinical trials (RCTs) to test this hypothesis (D'Silva, Duggan, & McCarthy, 2004). Recent meta-analyses and reviews suggest that violent offenders (Papalia et al., 2019), patients with antisocial PD (Wilson, 2014), and patients with psychopathy (Larsen et al., 2020), may be treatable. A few pilot studies of evidence-based PD treatments, such as mentalization-based therapy (Bateman, Fonagy, & Campbell, 2018), dialectical behavior therapy (Linehan et al., 2006), and schema therapy (ST) (Young, Arntz, & Giesen-Bloo, 2006), have shown promising results in forensic patients (Berzins & Trestman, 2004; Chakhssi, Kersten, de Ruiter, & Bernstein, 2014; Moulden, Mamak, & Chaimowitz, 2020; Ware, Wilson, Tapp, & Moore, 2016). However, rigorous RCTs of these treatments in forensic populations are lacking.

Furthermore, questions have been raised about the effectiveness of mandated (i.e. coerced) forensic treatment. In a meta-analysis of 129 diverse offender treatment studies (Parhar, Wormith, Derkzen, & Beauregard, 2008), the mean treatment effect was small, $d = 0.15$, with more coercive treatments being less effective. For the highest levels of coercion, where, as in our RCT, institutional treatment was mandated as a condition of offenders' sentences, there was no mean effect of treatment, $d = 0$ (Parhar et al., 2008).

In the Netherlands, individuals convicted of serious offenses which are (partly) attributable to mental disorders, including PDs, can be sentenced to mandated treatment in high-security hospitals, known as 'TBS clinics' ('Terbeschikkingstelling,' abbreviated as TBS, means 'at the discretion of the state'). The TBS sentence protects the public while providing treatment and lowering recidivism risk. This sentence is re-evaluated every 1–2 years and can be indefinite. Patients must show evidence of lowered risk before beginning 'rehabilitation', involving gradual re-entry into the community, first supervised leave, and then unsupervised leave. Leave can be withdrawn if patients exhibit high-risk behaviors (e.g. addiction, aggression). TBS patients may refuse or withdraw from therapy; however, this may have consequences (e.g. losing privileges or transfer to another clinic). In non-randomized studies, TBS treatment was superior to incarceration for reducing recidivism (Bregman & Wartna, 2010; Hildebrand, Hesper, Spreen, & Nijman, 2005).

Schema Therapy (ST; Young, Klosko, and Weishaar, 2003) is an integrative therapy for PDs that has demonstrated effectiveness in non-forensic patients with borderline PD (Giesen-Bloo et al., 2006; Nadort et al., 2009) and cluster C PDs, narcissistic, histrionic, and paranoid PDs (Bamelis, Evers, Spinhoven, & Arntz, 2014). We adapted ST for forensic PD patients (Bernstein, Arntz, & de Vos, 2007), with the aims of motivating and engaging patients, ameliorating PD symptoms, lowering risks and building strengths (i.e. protective factors), and facilitating re-entry into the community (Bernstein, Clercx, & Keulen-De Vos, 2019; Chakhssi, Bernstein, & De Ruiter, 2014). ST focuses on changing ingrained patterns of cognition and emotion (*early maladaptive schemas*; EMS) and maladaptive emotional states (*schema modes*) (Young et al., 2003), constructs associated with personality disorders in forensic (Keulen-de Vos et al., 2017) and non-forensic (Bamelis et al., 2014) populations. Consistent with the general aggression model (Gilbert & Daffern, 2011), EMS and schema modes involving attachment insecurity, insufficient self-control, and over-compensatory coping are associated with increased risk for aggression and criminal behavior (Keulen-de Vos et al., 2016), Cluster B personality disorders and recidivism (Keulen-de Vos et al., 2017), sex offending (Chakhssi, De Ruiter, & Bernstein, 2013), and psychopathy (Chakhssi et al., 2014; Keulen-de Vos et al., 2016).

We conducted a 3-year RCT to test the effectiveness of ST *v.* treatment-as-usual (TAU) for offenders with PD and aggression in 8 TBS clinics. To our knowledge, this is the first RCT of long-term, intensive therapy for violent offenders with PDs. Patients at each clinic were randomly assigned to ST or another form of individual therapy (TAU), typically CBT, eclectic/integrative therapy, or systemic therapy, which constituted usual therapy at the institution. Patients in both treatment conditions received the full range of other therapeutic modalities (e.g. group therapy, creative arts therapies) that they would normally receive. The primary aims of treatment were to reduce PD symptoms and facilitate rehabilitation into the community. We also assessed several secondary outcomes, including recidivism risks, strengths, institutional violence, EMS, and schema modes.

Methods

Patients

One-hundred-three male TBS patients with *DSM-IV* (American Psychiatric Association, 2000) antisocial, borderline, narcissistic, or paranoid PD or Cluster B PD-not-otherwise-specified (PD-NOS, meeting 5 or more Cluster B PD criteria) were randomized to receive ST ($N = 54$) or TAU ($N = 49$). Eight of the 12 TBS facilities in the Netherlands participated: de Rooyse Wissel, Venray ($N = 19$), de Rooyse Wissel, Maastricht ($N = 6$), van der Hoeven ($N = 17$), Oostvaarders ($N = 7$), Mesdag ($N = 16$), Veldzicht ($N = 11$), Kijvelanden ($N = 18$), and FPK Assen ($N = 9$). We only recruited male patients, because they constitute the vast majority in TBS (Van Gemmert, Van Schijndel, Gordeau, & Casanova, 2015). Exclusion criteria were (a) current psychotic symptoms, (b) schizophrenia or bipolar disorder, (c) current drug or alcohol dependence (but not abuse), (d) full-scale IQ <80, (e) serious neurological impairment, (f) autistic spectrum disorder, and (g) exclusive pedophilia (excluded because these problems suggested the need for other treatment methods).

The characteristics of the patients are given in Table 1. Nearly all were sentenced for physically violent offenses (54%), sexually violent offenses (26.2%), threats/coercion (16.5%), or arson (2.9%). Eighty-three percent ($N = 88$) were at high risk for (violent) recidivism (*Historic-Clinical-Risk Management Scheme-20-Version 2* ($HCR\text{-}20^{V2}$); Webster, Douglas, Eaves, & Hart, 1997). Their most prevalent PDs were antisocial, 60%, narcissistic, 21%, and borderline, 17.1% (*Structured Interview for DSM-IV Personality* Disorders (SIDP-IV, Pfohl, Blum, & Zimmerman, 1997). Fifty-four percent had significant traits of psychopathy (PCL-R $\geqslant 25$; *Psychopathy Checklist-Revised*, Hare & Hart, 1993), and 22% were highly psychopathic (PCL-R $\geqslant 30$).

Procedure

This study was approved by Maastricht University's Medical Ethical Committee (MEC 06-3-066). Patients were screened and recruited by research assistants at each hospital between January 2007 and March 2012. Patients were free to refuse the study or to withdraw at any time, without consequences. Dropout was defined as withdrawing from the study at the patient's request, and pushout (De Boer, Boon, De Haan, & Vermeiren, 2016) as withdrawing at treatment or research staff's request (Fig. 1). No

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

Table 1. Demographic, clinical, and forensic background of the sample

Baseline demographics					
		ST $N = 54$		TAU $N = 49$	
		Mean	S.D.	Mean	S.D.
Age	Age in years at baseline	38.81	8.92	36.45	10.77
IQ	Total IQ	93.80	9.98	94.00	10.83
Time in hospital at baseline	Number of days in hospital prior to starting ST or TAU	583.93	398.84	542.37	437.34
		Frequency	%	Frequency	%
Nationality	Netherlands	48	88.9	39	79.6
	Other EU country, Dutch Antilles, Morocco or Surinam	6	11.1	6	12.2
	Africa/Asia/South America	0	0.0	4	8.1
Highest level of education	None	4	7.4	3	6.1
	Primary school or primary school for special needs	12	22.2	12	24.4
	Secondary school	34	62.9	32	65.3
	Post-secondary education (college)	1	1.9	0	0.0
	Unknown	3	5.6	2	4.1
Psychiatric disorders					
		Frequency ST	% ST	Frequency TAU	% TAU
Axis I disorders, lifetime (SCID-I)	Mood disorder	13	24.1	8	16.0
	ADHD	4	7.4	6	12.0
	Psychotic disorder	3	5.6	4	8.0
	Anxiety disorder	14	25.9	10	20.0
	Substance related disorder	36	66.7	33	66.0
	Paraphilia	7	13.0	7	14.3
	Other Axis I diagnosis	3	5.6	3	6.1
	>1 Axis I diagnosis	22	40.7	19	38.0
Axis II (SIDP-IV)	Antisocial PD	33	61.1	30	60.0
	Borderline PD	12	22.2	6	12.0
	Narcissistic PD	11	20.4	11	22.0
	Paranoid PD	2	3.7	2	4.0
	PD-NOS (⩾5 Cluster B PD criteria)	7	13.0	2	4.0
	PD other	7	13.0	5	10.2
	>1 PD diagnosis	40	74.1	40	80.0
Criminal background at baseline					
		Mean ST	S.D. ST	Mean TAU	S.D. TAU
Age at first offense	Years	21.87	8.75	21.22	8.41
PCL-R Psychopathy	PCL-R total score	24.74	7.07	23.75	7.01
		Frequency ST	% ST	Frequency TAU	% TAU
PCL-R Psychopathy cut-off	PCL-R cut-off 25 or higher	29	53.7	27	54.0
	PCL-R cut-off 30 or higher	13	24.1	10	20.0
		Mean	S.D.	Mean	S.D.
HCR-20 Total score	Total HCR-20 scored for risk outside of the hospital	26.83	5.78	26.94	5.23

(*Continued*)

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

Table 1. (*Continued.*)

Criminal background at baseline					
		Mean ST	S.D. ST	Mean TAU	S.D. TAU
		Frequency ST	% ST	Frequency TAU	% TAU
HCR-20 Final risk clinical judgement	Low risk	0	0.0	1	2.0
	Medium risk	7	13.0	5	10.2
	High risk	45	83.3	43	87.8
		Frequency ST	% ST	Frequency TAU	% TAU
START	Total strengths score baseline	22.19	7.47	22.06	6.48
	Total vulnerability score baseline	15.69	8.75	16.53	7.44
		Mean ST	S.D. ST	Mean TAU	S.D. TAU
Incidents in forensic hospital	Weighed total baseline incidents	1.78	2.23	1.56	1.97
Criminal background at baseline					
		Frequency ST	% ST	Frequency TAU	% TAU
Baseline leave permission	Supervised leave at baseline	12	22.2	11	22.4
	Unsupervised leave at baseline	5	9.3	3	6.1
		Frequency ST	% ST	Frequency TAU	% TAU
Index crime	Murder/attempted murder	7	13.0	4	8.2
	Manslaughter/attempted manslaughter	14	25.9	14	28.5
	Rape/attempted rape	10	18.5	9	18.4
	Assault/attempted assault/(attempted) violent abuse	10	18.6	7	14.3
	Pedosexual crime	4	7.4	4	8.2
	Property crime with threat or violence	6	11.1	8	16.3
	Threatening or coercion, arson	3	5.6	3	6.0

incentives for participation were provided. After giving written informed consent and completing baseline assessments, patients were randomized to ST or TAU by a central research assistant with no access to patient information, using an adapted biased urn algorithm (Schouten, 1995), stratified by site. The algorithm assigned patients at random to the two conditions at each site, while ensuring that the overall proportion between conditions was balanced.

Patients were assessed by trained research assistants or diagnostic specialists at baseline and every 6 months for 3 years. In a 3-year study, we could not keep raters, staff, or patients blind to treatment conditions. However, we measured most outcomes using methods (e.g. decisions of leave advisory boards, informant ratings, risk assessments, incident registers) that were independent of patients' self-reports to mitigate response biases (Schretlen & Arkowitz, 1990). At the time that we initiated our study, the Dutch RCT register, which is recognized by the World Health Organization, was new and used mainly for medical RCTs, and not psychotherapy RCTs. When we registered our study on 16 January 2008, we had already enrolled the first 28 patients, for a duration of 1–11 months.

Treatment conditions

ST was initially given twice per week, the recommended frequency for severe PDs (Young et al., 2003), and found effective in borderline PD (Giesen-Bloo et al., 2006; Nadort et al., 2009). ST sessions were usually reduced to once per week after patients attained leave, mainly for practical reasons, as they spent more time outside of the hospital. The TAU patients at seven of the eight hospitals received individual therapy once per week as their primary therapy, which is usual practice at most TBS clinics. The TAU patients at one hospital (Kijvelanden) received group therapy as the primary therapy because this was the usual practice at this hospital.

The hours of primary, individual therapy per patient differed between the treatment conditions at each time-point between baseline and 18 months, and overall (ST 149.37 [S.D. = 73.66], TAU 101.33 [S.D. = 72.38] $U = 770.0$, $p < 0.001$). See online Supplementary Fig. 1 for therapy hours per time-point. However, TAU patients received more hours of auxiliary individual therapies (e.g. creative arts therapies) than ST patients from 12 to 18 months (ST = 3.96 [S.D. = 7.75], TAU = 11.09 [S.D. = 24.66], $U = 797.5$, $p = 0.047$). Combining all therapy hours, there were no differences between the treatment conditions at any time-point nor overall (ST = 260.51 [S.D. = 133.97], TAU = 230.35 [S.D. = 157.69], $U = 1083.5$, $p = 0.114$ N.S.).

There were no treatment condition differences in medication use. Seventy-six percent ($N = 78$) of patients were given psychotropic medications, with antipsychotics ($N = 41$, 39.8%) and antidepressants ($N = 34$, 33.0%) most commonly prescribed. Thirty-three patients received other medications, including

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press



A Dropout is defined as the termination of treatment or study participation on the participants' request.

B Pushout is defined as the termination of treatment or study participation on the request of treatment or research staff.

Fig. 1. CONSORT flowchart – overview of patient flow throughout the study.

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

anxiolytics, mood stabilizers, libido-reducing medications, and methadone.

Therapists

Twenty-two ST therapists and 19 TAU therapists participated, with separate therapists per condition. All therapists held Master's degrees in Psychology; 61% ($N = 25$) were female, and 36 (87.8%) completed a post-masters clinical specialization. Most TAU therapists were cognitive-behavioral (57.9%, $N = 11$), integrative/eclectic (26.3%, $N = 5$), or systemic in orientation (10.5%, $N = 2$). Therapists reported equivalent years of previous clinical experience (TAU, 13.95 [S.D. = 9.92], ST 11.82 [S.D. = 8.15], $t[39] = .755$, $p = 0.455$ N.S.); years of experience with PDs (TAU, 10.87 [S.D. = 8.44], ST 9.80 [S.D. = 7.37], $t[39] = .434$, $p = 0.666$.N.S.), number of patients treated (TAU, 160.44 [S.D. = 171.24] *v.* ST, 71.76 [S.D. = 84.08], $U = 138.5$, $p = 0.154$ N.S.), and number of patients with PD treated (TAU, 61.72 [S.D. = 63.52], ST 60.90 [S.D. = 72.23], $U = 178.0$, $p = 0.953$ N.S.).

ST therapists received 6–8 days of training by Bernstein and Kersten; attended biweekly supervision groups for 2 h, and practiced with one or two patients before submitting two randomly selected videotaped sessions to an independent rater. Therapists who received mean scores of four or higher on the six-point *Schema Therapy Rating Scale* (Young et al., 2006) were deemed competent to treat ST patients for the study. TAU therapists usually also participated in some form of supervision or peer-supervision group, typically for 1 h weekly or biweekly. To check treatment fidelity, we randomly selected 31 tapes, scored blind to treatment condition by two masters-level psychologists, using the *Schema Therapy Integrity Scale* (Bernstein, 2007). The interrater reliability for 15 double-scored tapes was ICC = 0.86. ST therapists scored significantly higher than the TAU therapists ($t[44] = 7.036$, $p < 0.001$).

Measures

Baseline diagnostic assessment included the *Structured Clinical Interview for DSM-IV Axis I* (SCID-1, First et al., 1998; Schouten, 1995) and *Structured Interview for DSM-IV Personality* Disorders (SIDP-IV, Pfohl et al., 1997) to establish DSM-IV-TR Axis I and II disorders, respectively, and the PCL-R (Hare, 2003), to assess psychopathy. Mean inter-rater reliability for the SCID-I diagnoses in 20 randomly selected patients was $\kappa = 0.93$ with no diagnosis lower than $\kappa = 0.77$. In 32 randomly selected patients, the inter-rater reliabilities (ICCs) for the SIDP-IV main diagnoses were antisocial PD, 0.91, borderline PD, 0.79, narcissistic PD, 0.93, paranoid PD, 0.96, and Cluster B PD-NOS, 0.79. Inter-reliability for the PCL-R total scores was ICC = 0.91 in 29 patients.

Primary outcome variables

Rehabilitation was operationalized as obtained permission for supervised or unsupervised leave. By law, patients applying for leave were evaluated with the help of structured risk assessments, usually the HCR-20^{V2} (Webster et al., 1997) or a similar Dutch instrument (Hildebrand et al., 2005). Leave requests must be approved by a 'Leave Advisory Board' within each hospital, with the final decision being made by a committee at the Ministry of Security and Justice, which usually follows the clinic's advice.

PD symptoms were assessed with the patient self-report and informant-report of the *Schedule for Nonadaptive and Adaptive Personality* (Clark & Vanderbleek, 2017; Melley, Oltmanns, and Turkheimer, 2002), adapted and validated for forensic patients (SNAP-FV, Keulen-de-Vos et al., 2011). The SNAP-FV consists of four PD scales (antisocial, narcissistic, borderline, and paranoid PD) and three temperament scales (positive and negative temperament, disinhibition). Patients completed the self-report version of the SNAP-FV. Three staff members per patient completed the informant version at each time point, typically the patients' primary psychotherapists, art therapists, psychiatric nurses, or treatment managers. Cronbach's α ranged from 0.91 to 0.93 for the PD scales and 0.82 to 0.89 for the temperament scales for the self-report SNAP-FV. ICCs ranged from 0.68 to 0.74 for the PD scales and 0.56 to 0.64 for the temperament scales for the informant SNAP-FV.

Secondary outcome variables

Violence risk was assessed with the *Historic-Clinical-Risk management scheme-20* (HCR-20, 2nd Edition, Webster et al., 1997), completed for risk level outside of the hospital setting (De Vogel & De Ruiter, 2006; Douglas & Reeves, 2010; Douglas, Ogloff, Nicholls, & Grant, 1999), and *Short-Term Assessment of Risk and Treatability* (START, Braithwaite, Charette, Crocker, & Reyes, 2010), which measures short-term dynamic vulnerabilities as well as strengths inside the hospital. For this study, we used the HCR-20^{V2} total quantitative risk rating, and the START total strengths and vulnerabilities scores. Research-assistants double-scored 31 patient files, which had been blinded to treatment condition. Inter-rater reliabilities between the original (non-blinded) raters and the blinded raters were satisfactory to good: HCR-20^{V2} total risk score, ICC = 0.82, START strength score, ICC = 0.71, start risk score, ICC = 0.69.

Institutional incidents were registered according to four incident categories: verbal aggression, threats, physical aggression, and violation of hospital rules (i.e. drug use). We created a total incidents score, which was a weighted sum (by severity) of all incidents occurring in any given 6-month-period. Because hospital staff recorded incidents on a daily basis in an electronic database, it was not possible to determine the inter-rater reliabilities for these scores.

Schema modes and EMS were assessed with the *Schema Mode Inventory-Revised* (SMI-R, Lobbestael, van Vreeswijk, Spinhoven, Schouten, & Arntz, 2010) and the *Young Schema Questionnaire-Short version* (YSQ-S, Young, 1998), respectively. We used the total scores for maladaptive modes and healthy modes on the SMI-R, which had reliabilities of $\alpha = 0.94$ and 0.96, respectively, and the YSQ-S total score, with α of 0.97, in our sample.

General syndrome psychopathology was assessed with the total score of the *Symptom Checklist-90-Revised* (Derogatis & Unger, 2010), which had reliability of $\alpha = 0.96$.

Statistics

A-priori power analysis estimated an N of 114, for the power of 0.80 with a two-tailed $\alpha = 0.05$ and medium effect size (based on previous clinical trials of ST: Giesen-Bloo et al., 2006; Nadort et al., 2009). Due to a reduced rate of referrals to TBS clinics during the last 3 years of recruitment, the attained N was slightly less. Outcomes were analyzed with (generalized) linear mixed models [(G)LMM; also called multilevel analysis or mixed regression] with the site as random intercept (if this did

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

not create estimation problems) and the appropriate distributional model and link dependent on the dependent variable: binomial with complementary log-log link for survival analysis, negative binomial with log-link for counts (with over-dispersion), gamma with log-link for skewed dimensional variables, and normal with identity link for normally distributed variables. Intention-to-treat analyses were performed for all outcomes, applying (G)LMM on all available data. There was very little missing data for rehabilitation, START, HCR-20, and incidents, because these were independent of patients' self-reports, and staff scoring continued when patients left the study.

Results

Baseline characteristics of the patients are presented in Table 1. There were very few significant differences at baseline following randomization, except for a trend in differences between total baseline SCL-90 score, with the ST group scoring slightly higher on total symptoms than the TAU group ($t[91] = 1.687$, $p = 0.09$).

Primary outcomes

Findings on primary outcomes are presented in Table 2 and Fig. 2.

Supervised leave

ST was significantly superior to TAU (treatment condition $F[1308] = 3.90$, $p = 0.048$, treatment*time $F[5308] = 9.40$, $p < 0.001$, $r = 0.36$). The significant treatment by time interaction was related to a particularly strong difference during the first 6 months in year 3 ($p = 0.004$). Cumulative proportions of attained supervised leave permission were higher in ST than in TAU (year 1, ST 56.9%, TAU 51.0%; year 2, ST 85.0%, TAU 76.5%; year 3, ST 96.7%, TAU 90.5%).

Unsupervised leave

The significant treatment-by-time interaction, $F(5472) = 3.45$, $p = 0.004$, $r = 0.19$, was related to a significantly higher chance to get permission in ST than in TAU in the first half-year, $t(472) = 3.84$, $p < 0.001$. Cumulative proportions of attained unsupervised leave permission were higher in ST than in TAU (year 1, ST 15.2%, TAU, 6.9%; year 2, ST 42.7%, TAU 37.8%; year 3, ST 67.4%, TAU 59.4%).

SNAP-FV personality disorder scores (Keulen-de-Vos et al., 2011)

ST showed a significantly steeper decrease in SNAP-FV PD scores over time, compared to TAU, $t(1387) = -2.85$, $p = 0.005$. The within-group effect sizes were $d = 0.56$ (TAU) and $d = 0.78$ (ST); the difference between treatments was $d = 0.22$. Patients reported significantly lower SNAP-FV PD scores than informants, but the change over time was parallel over patients and informants, without interaction with treatment. Post-hoc tests showed that paranoid PD scores were significantly higher than the other three PD scales (main effect of PD scale). Moreover, borderline and paranoid PD scores showed a significantly steeper decrease over time than narcissistic and antisocial PD scores (PD scale by time interaction), with no interaction with treatment.

Secondary outcomes

Findings on secondary outcomes are presented in Table 2 and Fig. 3.

SNAP-FV temperament scales (Keulen-de-Vos et al., 2011)

The treatment-by-time interaction showed a steeper decrease in SNAP-FV temperament scores in ST than in TAU, $t(1014) = 2.60$, $p = 0.01$. The within effect sizes were $d = 0.38$ (TAU) and $d = 0.63$ (ST). Post-hoc tests showed that patients reported lower (i.e. less negative) temperament scores than informants, but the change over time was parallel over raters, without interaction with treatment. Reversed positive emotions scores were generally higher than for the other temperament scales, and showed a smaller change over time, whereas negative emotions showed the largest change.

HCR-20^{V2} total risk score (Webster et al., 1997)

The piecewise LMM analysis showed reduced scores over time, with a trend towards faster HCR-20^{V2} reduction in the first 1.5 years in ST than in TAU, treatment-by-time $t(446) = -1.91$, $p = 0.057$; $d = 0.40$ (ST) *v.* 0.20 (TAU) at 1.5 years.

START vulnerabilities and strength scores (Braithwaite et al., 2010)

The analysis showed significant treatment-by-time and treatment-by-time-squared interactions, with ST showing steeper and earlier improvements, compared to TAU. However, at 3 years the TAU patients had 'caught up' and there was no longer a difference between treatments. The within treatment effect sizes at 1 year were $d = 0.23$ (TAU) and $d = 0.41$ (ST), differential $d = 0.18$; at 1.5 years, $d = 0.37$ (TAU) and $d = 0.56$ (ST), differential $d = 0.19$. At 3 years, the change compared to baseline was $d = 0.85$.

Incidents

The effect of time on weighted incidents was highly significant with no interaction of time with treatment condition, indicating that both treatments showed a comparable reduction. At 3 years change compared to baseline was large, $d = 0.97$.

Early maladaptive schemas (Young & Brown, 1998)

The effect of time on the EMS total score was highly significant, and there was a significant treatment-by-time interaction, indicating a stronger reduction in ST than in TAU, $d = 0.28$. The within treatment effect sizes were $d = 0.33$ (TAU) and $d = 0.61$ (ST).

Schema modes (Lobbestael et al., 2010)

The effect of time on healthy and maladaptive schema mode scores was highly significant, as were the treatment-by-time, and treatment-by-time-squared, interactions. At 1 year, the change in healthy modes in ST was $d = 0.40$; in TAU, $d = 0.06$; at 1.5 years, in ST, $d = 0.50$, in TAU, $d = 0.15$. At 1 year, the change in maladaptive modes in ST was $d = 0.48$; in TAU, $d = 0.14$; at 1.5 years, in ST, $d = 0.64$; in TAU, $d = 0.28$. Differences disappeared at 3 years.

SCL-90 total score (Derogatis & Unger, 2010)

The effect of time on SCL-90 total scores was highly significant, but there was no interaction of time-by-treatment. At 3 years change from baseline was (ST $d = 0.44$; TAU $d = 0.30$; differential $d = 0.14$).

Treatment retention

Retention was high in both conditions and dropouts were rare (Fig. 1) with a significant interaction of time-by-treatment condition (online Supplementary Fig. 2). Follow-up analyses showed a significant difference in year 1 with higher retention in ST (93%)

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

Table 2. Results from (generalized) mixed models analysis for treatment retention, and primary and secondary outcomes

Primary outcomes											
						Estimated chance TAU			Estimated chance ST		
	F	*d.f.*	*p*	*r*[f]	Period	95% *CI*			95% *CI*		
Supervised leave[a]					Months	Estimate	Lower	Higher	Estimate	Lower	Higher
Time	1.958	5, 308	0.085	0.18	0–6	0.285	0.074	0.548	0.326	0.143	0.524
Treatment	3.899	1, 308	**0.048**	0.11	6–12	0.314	0.136	0.511	0.360	0.202	0.521
Treatment × time	9.401	5, 308	**<0.001**	0.36	12–18	0.305	0.122	0.512	0.359	0.157	0.569
					18–24	0.311	0.070	0.600	0.456	0.205	0.677
					24–30	0.308	0.077	0.582	0.550	0.319	0.731
					30–36	0.415	0.187	0.629	0.515	0.133	0.803
					Overall	0.323	0.173	0.483	0.429	0.271	0.579
Unsupervised leave[a]					Months	Estimate	Lower	Higher	Estimate	Lower	Higher
Time	4.968	5, 472	<0.001	0.22	0–6	0.019	0.001	0.116	0.125	0.024	0.311
Treatment	1.086	1, 472	0.298	0.05	6–12	0.051	0.005	0.194	0.031	0.002	0.142
Treatment × time	3.451	5, 472	**0.004**	0.19	12–18	0.156	0.055	0.305	0.181	0.100	0.282
					18–24	0.208	0.091	0.357	0.175	0.064	0.332
					24–30	0.181	0.060	0.356	0.285	0.106	0.495
					30–36	0.203	0.081	0.363	0.203	0.096	0.339
					Overall	0.119	0.050	0.220	0.154	0.075	0.258

Primary outcomes										
		95% *CI* (B)					Effect sizes			
	B	Lower	Upper	*t*	*d.f.*	*p*	Between-subjects *d*	Within subjects *d* total sample[b]	Within subjects *d* TAU[b]	Within subjects *d* ST[b]
SNAP PD scales										
Time	−0.046	−0.051	−0.041	−17.385	1393	<0.001	0.22	0.67	0.56	0.78
Treatment × time	−0.015	−0.025	−0.005	−2.845	1387	**0.005**				
SNAP temperament										
Time	−0.035	−0.041	−0.028	−10.545	1019	<0.001	0.25	0.50	0.38	0.63
Treatment × time	−0.017	−0.030	−0.004	−2.596	1014	**0.010**				
HCR-20 risk[c]										
Time	−0.493	−0.707	−0.279	−5.390	7	0.001	0.20 (1.5 years) 0.07 (3 years)	0.30 (1.5 years) 0.60 (3 years)	0.20 (1.5 years) 0.57 (3 years)	0.40 (1.5 years) 0.64 (3 years)

(*Continued*)

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

Table 2. (*Continued.*)

Primary outcomes										
		95% *CI* (B)					Effect sizes			
	B	Lower	Upper	*t*	*d.f.*	*p*	Between-subjects *d*	Within subjects *d* total sample[b]	Within subjects *d* TAU[b]	Within subjects *d* ST[b]
Treatment × time	−0.333	−0.676	0.010	−1.909	446	0.057				
Treatment × time after knot	0.551	−0.182	1.283	−1.479	295	0.140				
Secondary outcomes										
		95% *CI* (B)					Effect sizes			
	B	Lower	Upper	*t*	*d.f.*	*p*	Between-subjects *d*	Within subjects *d* total sample[b]	Within subjects *d* TAU[b]	Within subjects *d* ST[b]
START strengths (inversed)[d]										
Time	−0.089	−0.127	−0.051	−4.570	1078	<0.001	0.18 (1 year) 0.19 (1.5 years)	0.85 (3 years)	0.23 (1 year) 0.37 (1.5 years)	0.41 (1 year) 0.56 (1.5 years)
Treatment × time	−0.079	−0.155	−0.002	−2.025	1078	**0.043**				
Time-squared	0.002	−0.004	0.008	0.770	935	0.441				
Treatment × time-squared	0.015	0.003	0.027	2.469	935	**0.014**				
START vulnerabilities[d]										
Time	−0.089	−0.127	−0.051	−4.570	1078	<0.001	0.18 (1 year) 0.19 (1.5 years)	0.85 (3 years)	0.23 (1 year) 0.37 (1.5 years)	0.41 (1 year) 0.56 (1.5 years)
Treatment × time	−0.079	−0.155	−0.002	−2.025	1078	**0.043**				
Time-squared	0.002	−0.004	0.008	0.770	935	0.441				
Treatment × time-squared	0.015	0.003	0.027	2.469	935	**0.014**				
Incidents[e]										
Time	−0.226	−0.328	−0.124	−4.398	93	<0.001		0.97		
Treatment × time	0.168	−0.036	0.372	1.640	93	0.104				
YSQ schemas[d]										
Time	−0.024	−0.031	−0.017	−6.601	394	<0.001	0.28	0.47	0.33	0.61
Treatment × time	−0.015	−0.029	−0.001	−2.021	394	**0.044**				
Maladaptive schema modes[d]										
Time	−0.050	−0.076	−0.024	−3.780	389	<0.001	0.34 (1 year) 0.36 (1.5 years)	0.88	0.14 (1 year) 0.28 (1.5 years)	0.48 (1 year) 0.64 (1.5 years)
Treatment × time	−0.081	−0.133	−0.029	−3.063	389	**0.002**				
Time-squared	0.001	−0.004	0.005	0.298	373	0.766				
Treatment × time-squared	0.140	0.006	0.023	3.228	373	**0.001**				

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

Secondary outcomes										
		95% *CI* (*B*)					Effect sizes			
	B	Lower	Upper	*t*	*d.f.*	*p*	*B*	Lower	Upper	*t*
Healthy schema modes (inversed)[d]										
Time	−0.035	−0.064	−0.007	−2.449	389	0.015	0.34 (1 year) 0.35 (1.5 years)	0.57	0.06 (1 year) 0.15 (1.5 years)	0.40 (1 year) 0.50 (1.5 years)
Treatment × time	−0.079	−0.136	−0.023	−2.749	389	**0.006**				
Time-squared	0.001	−0.003	0.006	0.565	373	0.572				
Treatment × time-squared	0.015	0.006	0.025	3.120	373	**0.002**				
SCL-90										
Time	−0.025	−0.037	−0.012	−3.937	405	<0.001	0.14	0.37	0.30	0.44
Treatment × time	−0.009	−0.034	0.016	−0.717	405	0.474				
Treatment retention[a]										
Time	3.542	2.264	0.030	0.16	0–12	0.798	0.695	0.883	0.932	0.819
Treatment	0.230	1.264	0.632	0.03	12–24	0.901	0.783	0.969	0.862	0.747
Treatment × time	5.576	2.264	**0.004**	0.20	24–36	0.945	0.861	0.986	0.931	0.825
					Overall	0.890	0.809	0.947	0.912	0.830

*Notes with Table ST2.
Time was coded 0, 1, 2, 3, 4, 5, 6 thus 1 unit = 6 months. For GLMM survival analysis time was a factor with 6 months intervals (supervised and unsupervised leave) or 1-year intervals (treatment retention). Treatment was coded −0.5 (TAU) and 0.5 (ST). In case of log-link transformed scales, results are in transformed scale. Unless otherwise indicated, LMMs based on a normal distribution were used. Significance levels <0.05 of treatment or treatment × time effects are indicated in **bold**. Where the main effect of treatment is not reported, this is because it reflects the difference between treatments at time = 0 (i.e. at baseline) which is not relevant for the hypothesis and not of interest because participants were randomized over treatments.
[a]GLMM survival analysis with a complementary log-log link. Estimated chances are in the original scale (range 0–1).
[b]Effect size Cohen's d = change over 3 years (unless otherwise indicated) from fixed part divided by *s.d.* based on residual and random intercept variance from GLMM.
[c]LMM piecewise regression with the knot (the point where there is a change in the slope) at 1.5 years.
[d]GLMM gamma regression with log link. Results are in transformed scale.
[e]GLMM negative binomial regression for counts with log link. Results are in transformed scale.
[f]Effect size $r = \sqrt{(df_n * F / (df_n * F + df_d))}$.

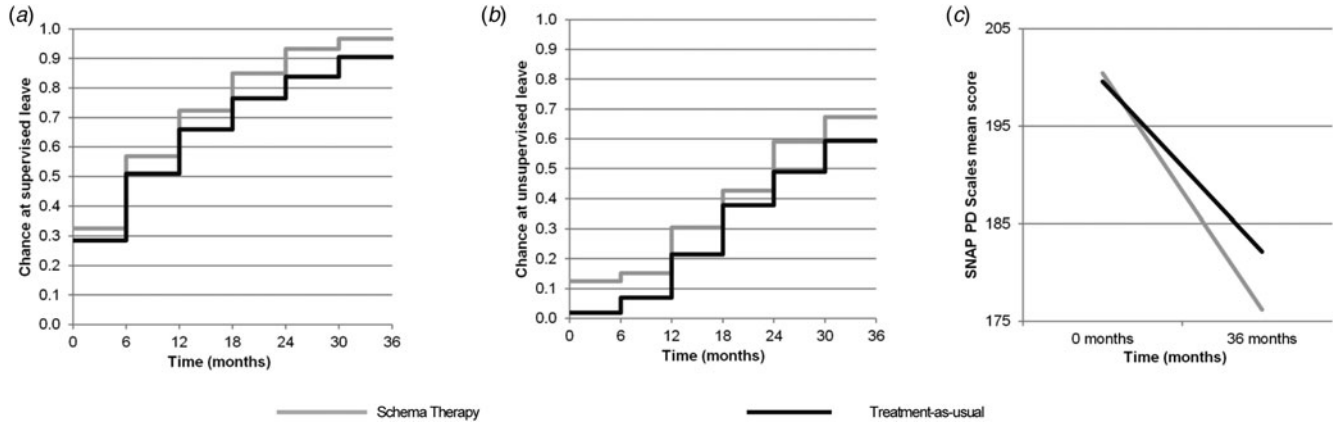


Fig. 2. Graphic representation of primary outcome measures. (*a*) Estimated cumulative chance of supervised leave as function of time and treament. (*b*) Estimated cumulative chance of unsupervised leave as function of time and treament. (*c*) Estimated means of SNAP-FV PD scales score by condition and time.

than in TAU (80%), $t(264) = 2.066$, $p = 0.004$. At 3 years, 75% of ST patients and 68% of TAU patients were retained.

Discussion

Our RCT is the first to demonstrate the effectiveness of long-term, intensive psychotherapy for rehabilitating violent offenders with PDs. Both ST and TAU were effective, producing moderate to large improvements in outcomes from baseline to 36 months, and contradicting the usually pessimistic views of the treatability of these patients. The vast majority in both treatment conditions eventually attained supervised leave (ST = 96.7%, TAU = 90.5%) and about two-thirds of unsupervised leave (ST = 67.4%, TAU = 59.4%). These results are particularly impressive given this high-risk sample. These findings are considerably better than in most studies of mandated treatments, which often failed to show effectiveness compared to no-treatment or wait-list control groups (Parhar et al., 2008). Consistent with the 'risk-need-responsivity' principles of forensic treatment (Andrews et al., 1990; Andrews, Bonta, & Wormith, 2006), our findings suggest that PD offenders need high intensity, long-term treatment that matches their severity.

ST showed significantly faster improvements than TAU on both of the primary outcomes – rehabilitation (i.e. supervised/unsupervised leave) and PD symptoms – and six of nine secondary outcomes. Differential (between-condition) effect sizes were small to medium. ST also retained significantly more patients than TAU, although retention in both conditions was high. ST moved patients through rehabilitation more rapidly than TAU, with medium effects for supervised leave ($r = 0.36$), and small to medium for unsupervised leave ($r = 0.19$), with significant differences in unsupervised leave only in the first 6 months. Thus, ST patients had more opportunity to practice coping with risky situations, and building strengths, in the community. ST was quicker than TAU to lower vulnerabilities and promote strengths on the START, which may have facilitated rehabilitation. Studies are needed to examine if ST also leads to greater cost-effectiveness, relative to TAU.

ST patients showed large improvements from baseline to 36 months in PD symptoms ($d = 0.78$), while TAU patients showed moderate ones ($d = 0.56$). A previous study of standard CBT for PD patients in TBS showed less personality improvement relative to baseline ($d = 0.23$–0.44) over a similar duration (Timmerman & Emmelkamp, 2005). Both ST and TAU groups showed reductions in all SNAP-FV PD and temperament scales, with the strongest reductions for borderline and paranoid PD, and negative temperament. ST had modest advantages over TAU ($d = 0.22$) in improving traits such as self-control and self-regulation, which are protective factors against recidivism (DeLisi, Hochstetler, Higgins, Beaver, & Graeve, 2008; Malouf et al., 2014). Both ST and TAU were highly effective in reducing incidents, with no significant differences between them.

ST's effectiveness in our study was lower than in previous RCTs of ST in (mostly female) outpatients with borderline (Giesen-Bloo et al., 2006; Nadort et al., 2009) or Cluster C and other PDs (Bamelis et al., 2014). This could be due to the institutional, mandated treatments they received (Parhar et al., 2008), and the severity of emotion regulation and externalizing behavior problems in male, antisocial patients (Berke, Reidy, & Zeichner, 2018). Moreover, the greatest effects of ST were in the first 2 years, perhaps due to ST's ability to facilitate a therapeutic bond (Giesen-Bloo et al., 2006). ST patients showed rapid, curvilinear improvements in START vulnerabilities and strengths scores and schema modes in the first two years, after which their scores plateaued, while the TAU group 'caught up'. A possible explanation is that, because the ST patients attained leave more quickly, they experienced more setbacks, as they were exposed to more risks (e.g. drugs and alcohol, antisocial peers, family and work stress), outside the hospital. Furthermore, ST patients initially received twice per week individual sessions, which may have boosted ST's early effectiveness.

We went to considerable lengths to ensure high methodological quality. However, there were some limitations. First, as is the case in nearly all studies of long-term psychotherapy, we cannot rule out halo effects because of the lack of blindness to treatment allocation. Although we used objective measures and multiple raters whenever possible, the lack of blindness could have biased the findings in favor of our hypotheses. For example, the Leave Advisory Boards were not blind to patients' treatment status and occasionally included study personnel. On the other hand, the results were consistent across many different outcome variables, measured in different ways, and the study was implemented across many centers, patients, and therapists.

Next, we chose not to equate the intensity of the individual therapies in this study, because our goal was to compare ST to usual TBS treatment, where individual therapy is typically given

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

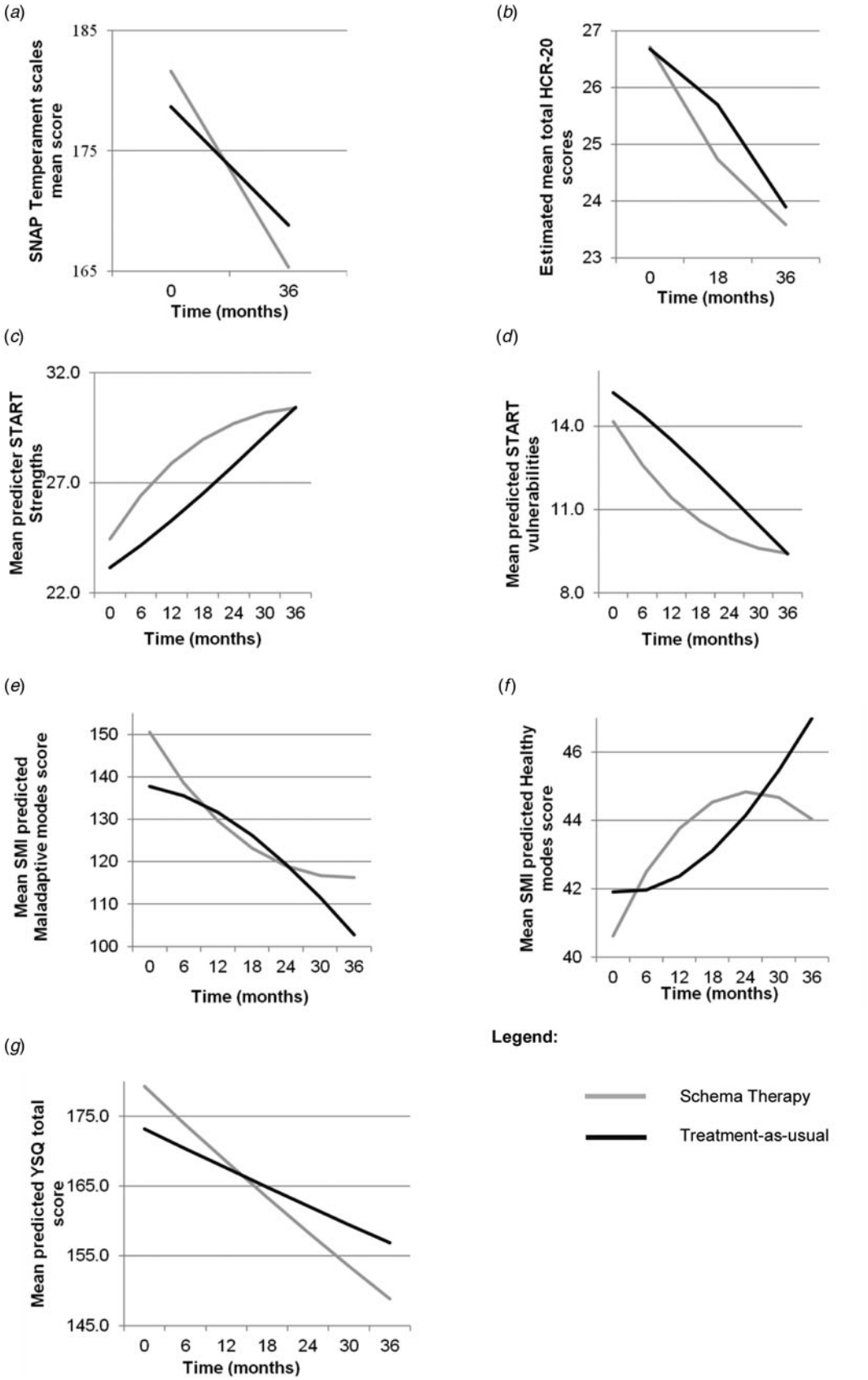


Fig. 3. Graphic representation of secondary outcome measures. (*a*) Estimated means of SNAP-FV Temperament scales score by condition and time. (*b*) Estimated means of total HCR-20^{V2} scores over time. (*c*) Estimated means of START Strength score over time. (*d*) Estimated means of START Vulnerabilities score over time. (*e*) Estimated means of SMI Maladaptive modes score over time. (*f*) Estimated means of SMI Health modes score over time. (*g*) Estimated means of YSQ score over time.

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

once per week. However, while ST patients received more individual therapy hours than TAU patients, there were no significant differences in the total therapy hours received, combining individual and auxiliary therapies, at any time point, or overall. Nevertheless, we cannot rule out that differences in effectiveness were due to the intensity of the individual therapy. The first step in validating ST was to compare it to high quality, usual TBS practice; the next step is to compare it to another specific, evidence-based treatment for PDs (e.g. Berzins & Trestman, 2004; Ware et al., 2016) while equating for individual therapy intensity. We will also augment individual ST with milieu-based ST approaches (van Wijk-Herbrink, Arntz, Broers, Roelofs, & Bernstein, 2019), to see if it enhances ST's effectiveness. Further, we did not compare the two conditions on recidivism, because arrests were uncommon during the study. We will examine recidivism at post-treatment follow-up. Finally, for ethical reasons, we did not use a no-treatment control group; thus, some of the improvements in both conditions could have been unrelated to treatment.

Finally, our findings can be most conservatively generalized to forensic hospital settings. We do not know if treatment retention and other outcomes would have differed in outpatient settings or in patients selected for specific diagnoses or offenses. Implementation studies are needed in other settings (e.g. outpatient clinics and prisons); populations (e.g. female offenders, youth offenders); and levels of coercion (e.g. voluntary treatments). Such studies would help in developing indications for treatment and adapting ST to patients with specific characteristics.

Conclusions

Our findings contradict typical notions regarding the treatability of offenders with PD and aggression. Intensive psychotherapy promoted rehabilitation and reduced PD symptoms in these hospitalized patients, with both ST and TAU showing evidence of effectiveness. ST produced more rapid improvements than TAU with modest but consistent advantages across most outcomes. These findings need replication in studies that equate for individual treatment intensity. Recidivism, the ultimate criterion in forensic populations, must also be investigated, before drawing any definite conclusions.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291721001161.

Acknowledgements. The authors wish to express their appreciation to the directors, staff, therapists, and patients of the participating forensic hospitals – Forensic Psychiatric Centers de Rooyse Wissel (Venray and Maastricht locations), van der Hoeven, Oostvaarders, Mesdag, Veldzicht, Kijvelanden, and Forensic Psychiatric Clinic Assen – whose dedication and financial and other support made this project possible. We thank past project coordinators and database managers funded by the EFP – Annette Löbbes, Lieke H.C. Bouts, Lotte C. de Geus, and Jacomina Gerbrandij – and former research assistants Eva de Spa, Ellen de Jonge, Merel van Vliet, Antoine Sint Fiet, Thijs Kanters, and Marloes Hartkoorn.

Financial support. We gratefully acknowledge the financial and other support of the Expertise Center for Forensic Psychiatry (EFP), and its past and current directors and staff; the program 'Quality in Forensic Care' ('*Kwaliteit Forensische Zorg*'; KFZ); and grant support (D. Bernstein, Principal Investigator) from the Netherlands Ministry of Security and Justice, the 'Scientific Research and Documentation Center' ('*Wetenschappelijk Onderzoek en Documentatiecentrum*'; WODC), and Maastricht University's Faculty of Psychology and Neuroscience.

Financial and other conflict of interest disclosure. The authors D. Bernstein, G. Kersten, and A. Arntz give trainings and supervision in Schema Therapy, for which they often receive financial compensation. For A. Arntz, the compensation goes to the university to support research. They have no other financial or other conflicts of interest. None of the other authors have financial or other conflicts of interest. In conducting this study, we have adhered to all pertinent ethical standards for maintaining anonymity and confidentiality, and respecting the rights of our patients.

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders – text revision (DSM-IV-TR)*. Washington, DC: American Psychiatric Association. doi:10.1176/appi.books.9780890423349.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, *52*(1), 7–27. doi:10.1177/0011128705281756.

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology: An interdisciplinary Journal*, *28*(3), 369–404. doi:10.1111/j.1745-9125.1990.tb01330.x.

Bamelis, L. L., Evers, S. M., Spinhoven, P., & Arntz, A. (2014). Results of a multicenter randomized controlled trial of the clinical effectiveness of schema therapy for personality disorders. *American Journal of Psychiatry*, *171*(3), 305–322. doi:10.1176/appi.ajp.2013.12040518.

Bateman, A W, Fonagy, P, & Campbell, C. (2018). Mentalization-based treatment. In W. J. Livesley, & R Larstone (Eds.), *Handbook of personality disorders: Theory, research, and treatment* (2nd edition, pp. 541–554). The Guilford Press.

Berke, D. S., Reidy, D., & Zeichner, A. (2018). Masculinity, emotion regulation, and psychopathology: A critical review and integrated model. *Clinical Psychology Review*, *66*, 106–116. doi:10.1016/j.cpr.2018.01.004$.

Bernstein, D. P. (2007). *Schema therapy integrity scale*. Maastricht: Maastricht University.

Bernstein, D. P., Arntz, A., & Vos, M. D. (2007). Schema focused therapy in forensic settings: Theoretical model and recommendations for best clinical practice. *International Journal of Forensic Mental Health*, *6*(2), 169–183. doi:10.1080/14999013.2007.10471261.

Bernstein, D. P., Clercx, M., & Keulen-De Vos, M. (2019). Schema therapy in forensic settings. In *The Wiley international handbook of correctional psychology* (pp. 654–668). Wiley Blackwell. doi:10.1002/9781119139980.ch41.

Berzins, L. G., & Trestman, R. L. (2004). The development and implementation of dialectical behavior therapy in forensic settings. *International Journal of Forensic Mental Health*, *3*(1), 93–103. doi:10.1080/14999013.2004.10471199.

Blackburn, R. (1998). Psychopathy and personality disorder: Implications of interpersonal theory. In D. J. Cooke, A. E. Forth, & R. D. Hare (Eds.), *Psychopathy: Theory, research and implications for society* (pp. 269–301). Dordrecht: Springer. doi:10.1007/978-94-011-3965-6_13.

Blackburn, R., Logan, C., Donnelly, J., & Renwick, S. (2003). Personality disorders, psychopathy and other mental disorders: Co-morbidity among patients at English and Scottish high-security hospitals. *The Journal of Forensic Psychiatry & Psychology*, *14*(1), 111–137. doi:10.1080/1478994031000077925.

Braithwaite, E., Charette, Y., Crocker, A. G., & Reyes, A. (2010). The predictive validity of clinical ratings of the short-term assessment of risk and treatability (START). *International Journal of Forensic Mental Health*, *9*(4), 271–281. doi:10.1080/14999013.2010.534378.

Bregman, I. M., & Wartna, B. S. J. (2010). Recidive TBS 1974–2006: Ontwikkelingen in de strafrechtelijke recidive van ex-terbeschikkinggestelden: Een tussenverslag [Recidivism TBS [Forensic hospitals] 1974–2006: Developments in criminal recidivism of ex-forensic patients: An interim report]. Wetenschappelijk Onderzoek en Documentatiecentrum (WODC).

Chakhssi, F., Bernstein, D., & De Ruiter, C. (2014). Early maladaptive schemas in relation to facets of psychopathy and institutional violence in offenders with personality disorders. *Legal and Criminological Psychology*, *19*(2), 356–372. doi:10.1111/lcrp.12002.

Chakhssi, F., De Ruiter, C., & Bernstein, D. P. (2013). Early maladaptive cognitive schemas in child sexual offenders compared with sexual offenders

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

against adults and nonsexual violent offenders: An exploratory study. *The Journal of Sexual Medicine*, *10*(9), 2201–2210. doi:10.1111/jsm.12171.

Chakhssi, F., Kersten, T., de Ruiter, C., & Bernstein, D. P. (2014). Treating the untreatable: A single case study of a psychopathic inpatient treated with schema therapy. *Psychotherapy*, *51*(3), 447. doi:10.1037/a0035773.

Clark L.A., & Vanderbleek E. (2017). Schedule for nonadaptive and adaptive personality. In D. Saklofske (Ed.), *Encyclopedia of personality and individual differences* (2nd edition, pp. 1–4). Cham: Springer. doi:10.1007/978-3-319-28099-8_942-2.

De Boer, S., Boon, A., De Haan, A., & Vermeiren, R. (2016). Treatment adherence in adolescent psychiatric inpatients with severe disruptive behaviour. *Clinical Psychology*, *22*, 55–62. doi:10.1111/cp.12111.

DeLisi, M., Hochstetler, A., Higgins, G. E., Beaver, K. M., & Graeve, C. M. (2008). Toward a general theory of criminal justice: Low self-control and offender noncompliance. *Criminal Justice Review*, *33*(2), 141–158. doi:10.1177/0734016808316781.

Derogatis, L. R., & Unger, R. (2010). Symptom checklist-90-revised. *The Corsini Encyclopedia of Psychology*, 1–2. doi:10.1037/t01210-000.

De Vogel, V., & De Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime & Law*, *12*(3), 321–336. doi:10.1080/10683160600569029.

Douglas, K. S., Ogloff, J. R., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the psychopathy checklist: Screening version. *Journal of Consulting and Clinical Psychology*, *67*(6), 917. doi:10.1037/0022-006x.67.6.917.

Douglas, K. S., & Reeves, K. A. (2010). Historical-Clinical-Risk management-20 (HCR-20) violence risk assessment scheme: Rationale, application, and empirical overview. In R. K. Otto, & K. S. Douglas (Eds.), *Handbook of violence and risk assessment* (pp. 147–186). New York: Taylor & Francis. doi:10.4135/9781412959537.n137.

D'Silva, K., Duggan, C., & McCarthy, L. (2004). Does treatment really make psychopaths worse? A review of the evidence. *Journal of Personality Disorders*, *18*(2), 163–177. doi:10.1521/pedi.18.2.163.32775.

First, M. B., Spitzer, R. L., Gibbon, M. (1998). In Patient (Ed.). *Structured Clinical Interview for DSM–IV Axis I disorders*New York: Biometrics Research, New York State Psychiatric Institute.

Giesen-Bloo, J., Van Dyck, R., Spinhoven, P., Van Tilburg, W., Dirksen, C., Van Asselt, T., … Arntz, A. (2006). Outpatient psychotherapy for borderline personality disorder: Randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, *63*(6), 649–658. doi:10.1001/archpsyc.63.6.649.

Gilbert, F., & Daffern, M. (2011). Illuminating the relationship between personality disorder and violence: Contributions of the general aggression model. *Psychology of Violence*, *1*(3), 230. doi:10.1037/a0024089.

Hare, R. D. (2003). *The hare psychopathy checklist – revised*. Toronto: Multi-Health Systems. doi:10.1037/t04993-000.

Hare, R. D., & Hart, S. D. (1993). Psychopathy, mental disorder, and crime. In *Mental disorder and crime* (pp. 104–115). Newbury Park: Sage. doi:10.1080/10683169508409761.

Hart, S. D., Webster, C. D., & Menzies, R. J. (1993). A note on portraying the accuracy of violence predictions. *Law and Human Behavior*, *17*(6), 695–700. doi:10.1007/bf01044690.

Hildebrand, M., Hesper, B. L., Spreen, M., & Nijman, H. L. I. (2005). De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie: Een onderzoek naar de betrouwbaarheid en predictieve validiteit van de HCR-20, HKT-30 en PCL-R [The value of structured risk taxation and psychopathy diagnosis: A study into the reliability and predictive validity of the HCR-20, HKT-30 and PCL-R]. Expertisecentrum Forensische Psychiatrie.

Johnson, J. G., Cohen, P., Smailes, E., Kasen, S., Oldham, J. M., Skodol, A. E., & Brook, J. S. (2000). Adolescent personality disorders associated with violence and criminal behavior during adolescence and early adulthood. *American Journal of Psychiatry*, *157*(9), 1406–1412. doi:10.1176/appi.ajp.157.9.1406.

Keulen-de Vos, M., Bernstein, D. P., Clark, L. A., Arntz, A., Lucker, T. P., & de Spa, E. (2011). Patient versus informant reports of personality disorders in forensic patients. *The Journal of Forensic Psychiatry & Psychology*, *22*(1), 52–71. doi:10.1080/14789949.2010.511242.

Keulen-de Vos, M., Bernstein, D. P., Clark, L. A., de Vogel, V., Bogaerts, S., Slaats, M., & Arntz, A. (2017). Validation of the schema mode concept in personality disordered offenders. *Legal and Criminological Psychology*, *22*(2), 420–441. doi:10.1111/lcrp.12109.

Keulen-de Vos, M. E., Bernstein, D. P., Vanstipelen, S., de Vogel, V., Lucker, T. P., Slaats, M., … Arntz, A. (2016). Schema modes in criminal and violent behaviour of forensic cluster B PD patients: A retrospective and prospective study. *Legal and Criminological Psychology*, *21*(1), 56–76. doi:10.1111/lcrp.12047.

Larsen, R. R., Jalava, J., & Griffiths, S. (2020). Are Psychopathy Checklist (PCL) psychopaths dangerous, untreatable, and without conscience? A systematic review of the empirical evidence. *Psychology, Public Policy, and Law*, *26*(3), 297. doi:10.1037/law0000239.

Lewis, N., & Lockwood, B. (2019). The Hidden Cost of Incarceration. Retrieved from https://www.themarshallproject.org/2019/12/17/the-hidden-cost-of-incarceration.

Lindsay, W. R., Hogue, T., Taylor, J. L., Mooney, P., Steptoe, L., Johnston, S., … Smith, A. H. (2006). Two studies on the prevalence and validity of personality disorder in three forensic intellectual disability samples. *The Journal of Forensic Psychiatry & Psychology*, *17*(3), 485–506. doi:10.1080/14789940600821719.

Linehan, M. M., Comtois, K. A., Murray, A. M., Brown, M. Z., Gallop, R. J., Heard, H. L., … Lindenboim, N. (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*, *63*(7), 757–766. https://doi.org/10.1001/archpsyc.63.7.757.

Lobbestael, J., van Vreeswijk, M., Spinhoven, P., Schouten, E., & Arntz, A. (2010). Reliability and validity of the short schema mode inventory (SMI). *Behavioural and Cognitive Psychotherapy*, *38*(4), 437–458. doi:10.1017/s1352465810000226.

Malouf, E. T., Schaefer, K. E., Witt, E. A., Moore, K. E., Stuewig, J., & Tangney, J. P. (2014). The brief self-control scale predicts jail inmates' recidivism, substance dependence, and post-release adjustment. *Personality and Social Psychology Bulletin*, *40*(3), 334–347. doi:10.1177/0146167213511666.

Melley, A. H., Oltmanns, T. F., & Turkheimer, E. (2002). The schedule for nonadaptive and adaptive personality (SNAP) temporal stability and predictive validity of the diagnostic scales. *Assessment*, *9*(2), 181–187. doi:10.1177/10791102009002009.

Moulden, H. M., Mamak, M., & Chaimowitz, G. (2020). A preliminary evaluation of the effectiveness of dialectical behaviour therapy in a forensic psychiatric setting. *Criminal Behaviour and Mental Health*, *30*(2-3), 141–150. doi:10.1002/cbm.2158.

Nadort, M., Arntz, A., Smit, J. H., Giesen-Bloo, J., Eikelenboom, M., Spinhoven, P., … van Dyck, R. (2009). Implementation of outpatient schema therapy for borderline personality disorder with versus without crisis support by the therapist outside office hours: A randomized trial. *Behaviour Research and Therapy*, *47*(11), 961–973. doi:10.1016/j.brat.2009.07.013.

Nijman, H., Cima, M., & Merckelbach, H. (2003). Nature and antecedents of psychotic patients' crimes. *The Journal of Forensic Psychiatry*, *14*(3), 542–553. doi:10.1080/1478994031000152754.

Papalia, N., Spivak, B., Daffern, M., & Ogloff, J. R. (2019). A meta-analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clinical Psychology: Science and Practice*, *26*(2), e12282. doi:10.1111/cpsp.12282.

Parhar, K. K., Wormith, J. S., Derkzen, D. M., & Beauregard, A. M. (2008). Offender coercion in treatment: A meta-analysis of effectiveness. *Criminal Justice and Behavior*, *35*(9), 1109–1135. doi:10.1177/0093854808320169.

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

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*(4), 399–412. doi:10.1007/BF02352266.

Ross, J., Quayle, E., Newman, E., & Tansey, L. (2013). The impact of psychological therapies on violent behaviour in clinical and forensic settings: A systematic review. *Aggression and Violent Behavior*, *18*(6), 761–773. doi:10.1016/j.avb.2013.09.001.

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press

Schouten, H. J. (1995). Adaptive biased urn randomization in small strata when blinding is impossible. *International Biometric Society*, *51*, 1529–1535. doi:10.2307/2533283.

Schretlen, D., & Arkowitz, H. (1990). A psychological test battery to detect prison inmates who fake insanity or mental retardation. *Behavioral Sciences & the Law*, *8*(1), 75–84. doi:10.1002/bsl.2370080109.

Serin, R. C. (1996). Violent recidivism in criminal psychopaths. *Law and Human Behavior*, *20*(2), 207–217. doi:10.1007/bf01499355.

Settumba, S. N., Chambers, G. M., Shanahan, M., Schofield, P., & Butler, T. (2018). Are we getting value for money from behavioral interventions for offenders? A research note reviewing the economic evaluation literature. *American Journal of Criminal Justice*, *43*(2), 411–431. doi:10.1007/s12103-017-9399-1.

Timmerman, I. G., & Emmelkamp, P. M. (2005). The effects of cognitive-behavioral treatment for forensic inpatients. *International Journal of Offender Therapy and Comparative Criminology*, *49*(5), 590–606. doi:10.1177/0306624×05277661.

Van Gemmert, N., Van Schijndel, C., Gordeau, W., & Casanova, L. (2015). Forensische Zorg in getal. 2010– 2014 [Forensic Care in Numbers 2010–2014]. DJI. Ministerie van Justitie [Judicial Institutions Agency].

van Wijk-Herbrink, M. F., Arntz, A., Broers, N. J., Roelofs, J., & Bernstein, D. P. (2019). A schema therapy based milieu in secure residential youth care: Effects on aggression, group climate, repressive staff interventions, and team functioning. *Residential Treatment for Children & Youth*, 1–18. doi:10.1080/0886571X.2019.1692758.

Walmsley, R. (2013). *World prison population list* (10th ed.). International Centre for Prison Studies, 1–6. https://www.prisonstudies.org/sites/default/files/resources/downloads/wppl_10.pdf.

Ware, A., Wilson, C., Tapp, J., & Moore, E. (2016). Mentalisation-based therapy (MBT) in a high-secure hospital setting: Expert by experience feedback on participation. *The Journal of Forensic Psychiatry & Psychology*, *27*(5), 722–744. doi:10.1080/14789949.2016.1174725.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence (version 2)*. Burnaby, Canada: Mental Health. Law, and Policy Institute, Simon Fraser University. doi:10.4135/9781412959537.n137.

Wilson, H. A. (2014). Can antisocial personality disorder be treated? A meta-analysis examining the effectiveness of treatment in reducing recidivism for individuals diagnosed with ASPD. *International Journal of Forensic Mental Health*, *13*(1), 36–46. doi:10.1080/14999013.2014.890682.

Wong, S. C., Gordon, A., Gu, D., Lewis, K., & Olver, M. E. (2012). The effectiveness of violence reduction treatment for psychopathic offenders: Empirical evidence and a treatment model. *International Journal of Forensic Mental Health*, *11*(4), 336–349. doi:10.1080/14999013.2012.746760.

Young, J. E., Arntz, A., & Giesen-Bloo, J. (2006). Therapy Adherence and Competence Scale. http://www.epp.unimaas.nl.

Young, J. E., & Brown, G. (1998). *Young schema questionnaire: Short form*. New York: Cognitive Therapy Center. doi:10.1037/t67023-000.

Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema therapy: A practitioner's guide*. New York: Guilford Press. doi:10.1017/S1352465804211869.

https://doi.org/10.1017/S0033291721001161 Published online by Cambridge University Press