Behaviour Research and Therapy 47 (2009) 961–973

Contents lists available at ScienceDirect

Behaviour Research and Therapy

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

Implementation of outpatient schema therapy for borderline personality disorder with versus without crisis support by the therapist outside office hours: A randomized trial

Marjon Nadort [a,*], Arnoud Arntz [b], Johannes H. Smit [a], Josephine Giesen-Bloo [b], Merijn Eikelenboom [a], Philip Spinhoven [d], Thea van Asselt [e], Michel Wensing [c], Richard van Dyck [a]

[a] GGZ inGeest, Department of Psychiatry and EMGO Institute, VU University Medical Center Amsterdam, A.J. Ernststraat 887, 1081 HL Amsterdam, The Netherlands
[b] Maastricht University, Department of Clinical Psychological Science, The Netherlands
[c] Radboud University Nijmegen Medical Centre, Scientific Institute for Quality of Healthcare, The Netherlands
[d] Leiden University, Department of Psychology, The Netherlands
[e] Department KEMTA Academic Hospital Maastricht, The Netherlands

Keywords:
Borderline personality disorder
Outpatient therapy
Schema therapy
Implementation
Crisis support outside office hours

A B S T R A C T

Objective: This study aimed to evaluate the success of implementing outpatient schema focused therapy (ST) for borderline patients in regular mental healthcare and to determine the added value of therapist telephone availability outside office hours in case of crisis (TTA).

Methods: To enhance the implementation, the following adaptations regarding the original ST protocol were applied: a reduction in the frequency and duration of the therapy; training therapists of eight regular healthcare centers in ST with a structured and piloted program supported by a set of films (DVDs) with examples of ST techniques; training and supervision given by Dutch experts. Telephone availability outside office hours was randomly allocated to 50% of the therapists of each treatment center. Patient's outcome measures were assessed with a semi-structured interview and self-report measures on BPD, quality of life, general psychopathology and an ST questionnaire, before, during and after treatment.

Results: Data on 62 DSM-IV defined BPD patients were available. Intention-to-treat analyses showed that after 1.5 years of ST 42% of the patients had recovered from BPD.
No added value of therapist telephone availability (TTA) was found on the BPDSI score nor on any other measure after 1.5 years of ST.

Conclusions: ST for BPD can be successfully implemented in regular mental healthcare. Treatment results and dropout were comparable to a previous clinical trail. No additional effect of extra crisis support with TTA outside office hours ST was found.

Introduction

Borderline Personality Disorder (BPD) has long been viewed as severe and difficult to treat. However, during recent years several promising treatment possibilities have been developed. Among them, Schema Therapy (ST) was found to be effective regarding all aspects of BPD. How well ST can be delivered in regular mental healthcare practice is unknown, but it was expected that its implementation poses challenges. BPD is marked by chronic instability in multiple areas (emotional dysregulation, self-harm, impulsivity and identity disturbance). The lifetime prevalence of BPD in the general population is 2%. In psychiatric outpatient settings 10% of the patients suffer from BPD, in psychiatric inpatients settings 20% (APA, 2005). The medical and societal costs for BPD are substantial (Ten Have, Lorsheyd, van Bijl, & Osterthun, 1995; van Asselt, Dirksen, Arntz, Giesen-Bloo, & van Dyck, 2008; van Asselt, Dirksen, Arntz, & Severens, 2007). About 10% of the BPD patients die because of suicide (Paris, 1993, 2008).

However, recent years showed progress in the development of treatment options (Arntz & van Genderen, 2009; Bateman & Fonagy, 2004; Linehan, 1993a, 1993b; van Genderen & Arntz, 2005; Yeomans, Clarkin, & Kernberg, 2002; Young, Klosko, & Weishaar, 2003) that are supported by randomized controlled trials (Bateman & Fonagy, 1999; Giesen-Bloo et al., 2006; Linehan, Armstrong,

* Corresponding author.
E-mail address: m.nadort@ggzingeest.nl (M. Nadort).

0005-7967/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.brat.2009.07.013

Suarez, Allmon, & Heard, 1991; Linehan et al., 2006; Paris, 2008; Verheul et al., 2003). These treatments demonstrated effectiveness on symptom level, as manifested by reduced suicide attempts, fewer acts of self-harm or hospitalizations. In an RCT which compared Schema therapy (ST; also called Schema Focused Therapy (SFT)) and Transference Focused Psychotherapy (TFP) (Giesen-Bloo et al., 2006) both therapies showed a significant change in personality that was maintained at 1-year follow up (Giesen-Bloo et al., 2009). This study showed that three years of ST and TFP proved to bring about a significant change in patient's personality, shown by reductions in all BPD symptoms and general psychopathologic dysfunction, increases in quality of life, and changes in associated personality features. While both treatment conditions showed positive results in the treatment of many aspects of BPD, ST was superior to TFP with respect to reduction in BPD manifestations, general psychopathologic dysfunction, and change in ST/TFP personality concepts. ST had a recovery rate of 45.5% and a reliable change rate of 65.9%. The dropout rate for ST was significantly lower than for TFP.

Based on these positive results, a study of the implementation of ST in regular mental healthcare practice was conducted. One of the premises in the therapeutic approach of ST (Arntz & van Genderen, 2009; van Genderen & Arntz, 2005; Young et al., 2003) and Dialectical Behavior Therapy (Linehan, 1993a, 1993b; Linehan et al., 1991; Linehan et al., 2006; Verheul et al., 2003) is that borderline patients need extra support from the therapist between sessions when they are in crisis or in emotional need. For this reason patients are offered a special phone number where they can reach their therapist outside of office hours. This personal connection between sessions is suggested to help to refute the patient's beliefs that there is nobody who really cares and can help to prevent or overcome crisis. In a pilot study of ST crisis support in the form of therapist phone accessibility outside office hours was one of the most controversial topics (Giesen-Bloo, Arntz, Dyck, Spinhoven, & van Tilburg, 2001) and led some therapists to withdraw from the project. In general mental healthcare there is much discussion about this topic because of the financial consequences, the burden to and responsibility of the therapist, and the possible risk of violation of boundaries. Therefore, telephone accessibility outside office hours was perceived as an important barrier for the successful implementation of ST in regular practice. The RCT by Giesen-Bloo et al. (2006) demonstrated that ST is a successful treatment, but it remains unknown whether the crisis support by the therapist was crucial to outcomes. Since the issue of crisis support outside office hours by the therapist makes it difficult to implement ST in regular practice and its effect has never been examined, we decided to investigate the role of the crisis support outside office hours in the implementation study by randomly allocating the crisis support outside office hours to 50% of the therapists.

In sum, this study tested the implementation of ST for BPD in regular mental healthcare and compared two modalities: one with extra crisis support by the therapist outside office hours and one without such telephone support. The study had three aims. First, to assess whether patient outcomes after 1.5 years of ST would be the same when implemented in regular practice, compared to what was found in a clinical trial of this therapy. Since rigorous evaluations such as RCTs always imply controlled conditions, it is unclear to what extent their positive effects can be generalized to regular clinical practice. Treatment effects may be more modest outside RCTs because of different circumstances (Rothwell, 1995; Weersing, 2005; Wilson, 1995). The second aim was to assess the added value of therapist telephone availability outside office hours in case of crisis (TTA) during those 1.5 years of ST. The third aim was to assess the problems that arose during the implementation process.

Method

Study design

The study was a multicenter randomized two-group design for studying the added value of therapist phone support outside office hours. It was also a clinical evaluation of implementing ST for BPD and a comparison of the regular mental healthcare treatment results with a clinical trial, using the so-called benchmark strategy. Benchmarking contains four elements: defining the problem, population and treatment model; selecting or creating a gold-standard outcome benchmark from the research literature; measuring outcome in the applied setting, using comparable methods as in the benchmark and finally comparing outcomes and exploring reasons for any differences (Weersing, 2005). The interventions and assessments were done between December 2005 and February 2009. The medical ethical committees of the participating centers approved the study.

Treatment setting

Twenty mental healthcare centers were approached and invited to take part in the implementation study. Selection criteria were a) at least two therapists on each location so that peer supervision groups could be formed, b) therapists agree in executing the telephone availability outside office hours and managers had to give their permission to do so, c) both therapists and managers had to agree in making the necessary time reservations for monthly supervision and weekly peer supervision. Eight general mental healthcare centers, covering 31 clinicians were willing to participate in the study. These were regional institutes, covering urbanized areas located in different parts of the Netherlands: Amersfoort, Amsterdam (3), Leeuwarden, Utrecht, Rotterdam en Zaandam.

Treatment setting, phone support and therapists

To prevent regional influences TTA had to be equally spread over the eight different sites. Therefore we used a stratified randomization procedure. The stratification procedure was performed by a study-independent person and concealed for participating therapists, patients and researchers. All the therapists in the study had agreed to provide the extra phone support outside office hours if so randomized. Stratified per center, in total 16 therapists were randomly allocated to the condition with extra phone support and 15 therapists to the condition without extra phone support. Each therapist treated 2 patients either with or without phone support dependent upon the randomization.

Patients were then randomly assigned to one of the therapists in the different institutions or regions.

Patients and procedures

Inclusion criteria were a DSM-IV based main diagnosis of BPD, age between 18 and 60 years, a BPDSI-IV score above 20, and Dutch literacy. General exclusion criteria were psychotic disorders (except short, reactive psychotic episodes), bipolar disorder, dissociative identity disorder, antisocial personality disorder, attention deficit hyperactivity disorder, addiction of such severity that clinical detoxification was indicated (after which entering treatment was possible), psychiatric disorders secondary to medical conditions and mental retardation. Co morbid axis-I and axis-II disorders were allowed, as was medication use.

Most of the patients were referred by therapists in secondary and tertiary community mental health institutes in each center's area. Some patients were referred by primary care physicians or

psychotherapists with private practices. All patients were referred based on a clinical diagnosis of BPD. Patients were then assessed at each site. The M.I.N.I. (Pinninti, Madison, Musser, & Rissmiller, 2003; Sheehan et al., 1998) was used for assessing the Axis-I diagnosis. The BPD section of the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (SCID II, First, Gibbon, Spitzer, Williams, & Benjamin, 1997; Weertman, Arntz, & Kerkhofs, 2000) was used for assessing the diagnosis BPD. Only the BPD section of the SCID II was used. If APD was suspected patients were not included. Patients were further screened using a semi-structured clinical interview, the Borderline Personality Disorder Severity Index, fourth version (BPDSI-IV; range 0–90; Arntz et al., 2003; Giesen-Bloo, Wachters, Schouten, & Arntz, 2006b). A BPDSI-IV cut-off score of $\geq$20 discriminates patients with BPD from patients with other personality disorders (Giesen-Bloo, Wachters, et al., 2006b). Further, if illiteracy was suspected, the Dutch Adult Reading Test (Schmand, Lindeboom, & van Harskamp, 1992) was administered. A positive screening procedure took 2 months, and this interval served as a patient's motivational check for undergoing intensive psychotherapy. See Fig. 1 of patient flow.

Signed informed consent was obtained after full explanation of the procedures and of both conditions of ST before the first assessment and randomization. Participants did not receive compensation for screening or assessments. Participating in assessments was obligatory to receiving the treatments studied.

Therapists

Thirty-one therapists treated two patients each. One therapist held a doctoral degree (ST with extra phone support), 29 therapists held master's degrees (15 ST with and 14 without extra phone support), one therapist held a bachelor's degree with postgraduate training (ST without extra phone support), with no between-group differences ($P = 0.29$). Twenty-five therapists had previous therapy experience with patients with BPD, six therapists had no previous therapy experience with patients with BPD (3 in each condition). Length of experience with BPD as in the ST with extra phone support condition $M = 7.88$ [SD 6.58] years; in the ST without extra phone support condition $M = 6.60$ [SD 5.37] years, with no between-group differences, $t(29) = 0.59$, $P = 0.56$). Most therapists had little experience with ST (mean length of experience [SD]: ST with extra phone support 1.19[1.94] years; ST without extra phone support 0.47[1.13] years), with no between-group differences, $t(29) = 1.25$, $P = 0.22$. There were 9 male and 22 female therapists. 3 men and 12 women in the condition with phone support and 6 men and 10 women in the condition without phone support with no between-group difference, $\chi^2(1) = 1.15$, $P = 0.28$.

Implementation interventions

On the basis of explorations of possible facilitators and barriers, the following implementation interventions were applied to enhance successful implementation (Nadort et al., 2009). Firstly, therapists, managers and assistants of different mental healthcare centers were informed of the study. Secondly agreements were made with the therapists and managers about the time investment for the treatment protocol (sessions twice a week, peer supervision weekly and supervision once a month during the first year and less frequently during the second year) and financial aspects. Thirdly therapists and research assistants were trained and support on organizational level was offered. The process evaluation aimed to assess the impact of these implementation interventions on the delivery of ST for BPD patients and to analyze the problems that might occur during the implementation process.

Training and supervision

As the primary aim of the study was to assess whether ST could be successfully implemented in regular mental healthcare practice, we made the following adjustments compared to the Giesen-Bloo et al. trial. In the study of Giesen-Bloo et al. (2006) the ST therapists were trained and supervised by the originator of ST, Jeffrey Young. In the implementation study the therapists were trained and supervised by Dutch experts (Nadort, 2005; Nadort & Giesen-Bloo, 2005). The training was based on a structured and piloted program supported by a set of DVDs with examples of ST techniques (website for schematherapy: www.schematherapie.nl (2005)) (see Nadort et al., 2009).

Therapists were trained in a 50 h training program (eight days during a period of two months). Essential to the treatment is expert supervision and peer supervision. During the first year monthly supervision was provided at each site, in the second year supervision was provided every two months. The therapists had weekly peer supervision on each site. There was a 1-day central supervision for all therapists once a year.

Frequency of sessions and treatment period

In the RCT (Giesen-Bloo et al., 2006) the treatment period was three years with sessions twice a week. In the implementation study there were also sessions twice a week in the first year, but sessions once a week in the second year. In the implementation study we decided to do the first evaluation after a treatment period of eighteen months. This was done for several reasons: different treatments have shown positive results after 1–1.5 years of treatment (Bateman & Fonagy, 1999; Linehan et al., 1991, 2006; Giesen-Bloo et al., 2006), effectiveness already became apparent after one year (Giesen-Bloo et al., 2006) and most drop outs occurred during the first 1.5 years of therapy (Giesen-Bloo et al., 2006; Linehan et al., 1991).

Treatment protocol

Treatment was offered in 45-min sessions twice a week. Treatment protocols addressed the theoretical model, treatment frame, different phases and the use of strategies and techniques (Arntz & van Genderen, 2009; van Genderen & Arntz, 2005; Nadort, 2005; Nadort & Giesen-Bloo, 2005; Young & Klosko, 1999; Young, Klosko, & Weishaar, 2005; Young et al., 2003). Central to ST is the assumption of 5 schema modes specific for BPD. Schema modes are sets of schemas expressed in pervasive patterns of thinking, feeling and behaving (Lobbestael, Arntz, & Sieswerda, 2005; Lobbestael, van Vreeswijk, & Arntz, 2008). Change is achieved through a range of behavioral, cognitive and experiential techniques that focus on (1) the therapeutic relationship, (2) daily life outside therapy and (3) past (traumatic) experiences. Recovery in ST is achieved when dysfunctional schemas no longer control or rule the patient's life.

Measurements

Procedures

Each patient's first assessment occurred after inclusion and before randomization. Then assessments were made every 6 months for 1.5 years by independent research assistants.

Eleven experienced and trained research assistants with higher vocational training in psychology assessed patients for treatment outcome measures. Study researchers, screeners, research assistants and therapists were masked to treatment allocation during the screening period and the first assessment.

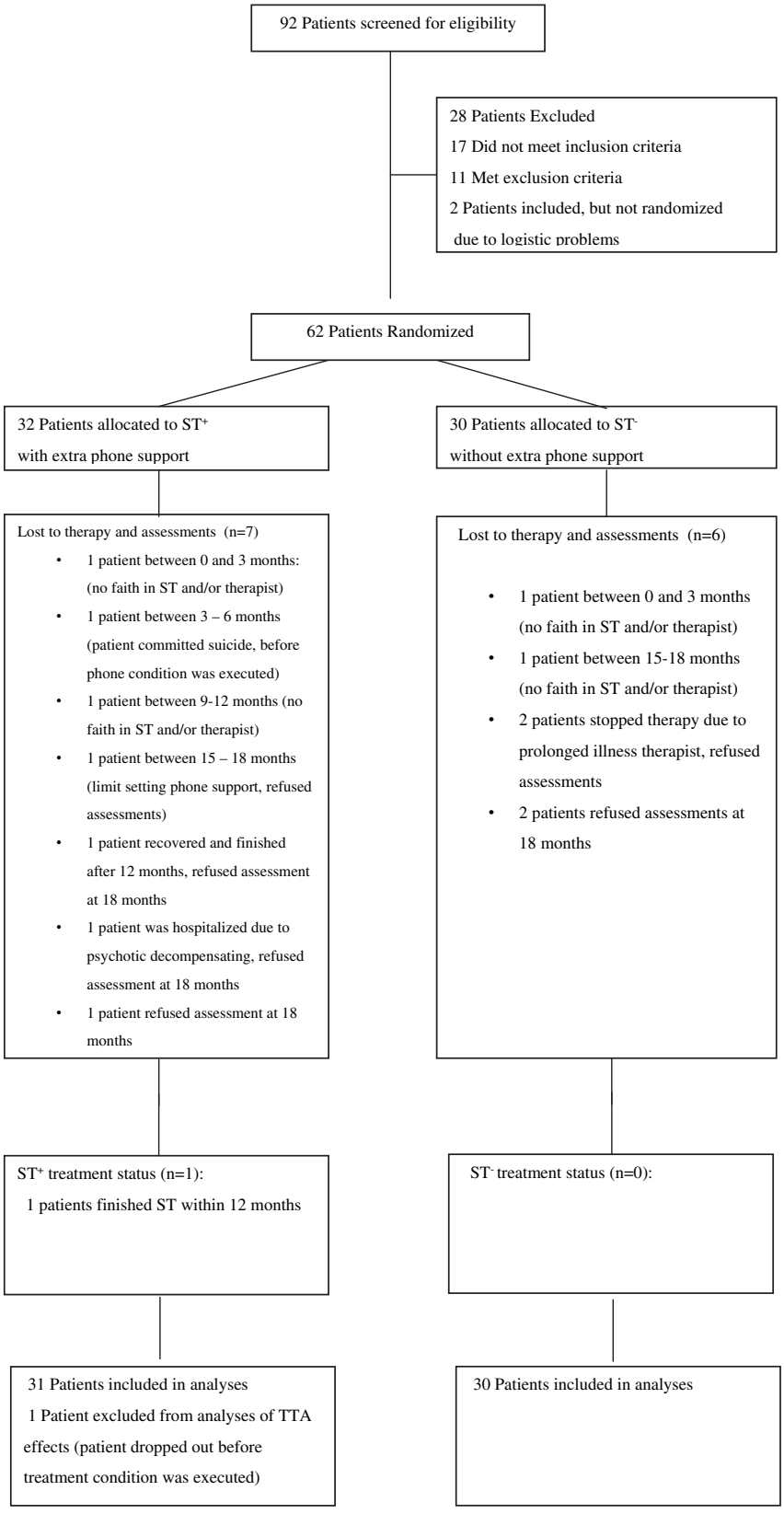

Fig. 1. Patient flow of implementation trial.

Treatment outcomes

Primary outcome measure: BPDSI-IV

The primary outcome measure was the score on the *BPDSI-IV*, a DSM-IV BPD criteria- based semi-structured interview: this 70-item index represents the current severity and frequency of the DSM-IV BPD manifestations. This instrument showed excellent psychometric features (Cronbach's alpha = 0.85, interrater reliability, 0.99; validity and sensitivity to change; Arntz et al., 2003; Giesen-Bloo, Wachters, et al., 2006b). Previous research (Arntz et al., 2003; Giesen-Bloo, Wachters, et al., 2006b) found a cut-off score (Jacobson & Truax, 1991) of 15 between patients with BPD and controls, with a specificity of 0.97 and a sensitivity of 1.00.

Recovery criterion

The recovery criterion was defined as achieving a BPDSI-IV score of less than 15 and maintaining this score until the last assessment.

Reliable change

A second criterion was reliable change (Jacobson & Truax, 1991), which reflects individual clinically significant improvement. For the BPDSI-IV, reliable change was achieved when improvement was at least 11.70 points at the last assessment (Giesen-Bloo et al., 2009).

Secondary outcome measures

EuroQol, WHOQol. A secondary outcome measure was quality of life, assessed by means of two widely used and psychometrically sound self-report questionnaires: the *EuroQol-thermometer and EQ 5D* and the *World Health Organization Quality of Life Questionnaire* (Brooks, 1996; Dolan, 1997; EuroQol Group, 1990; WHOQOL Group, 1998). The vertical EuroQol-thermometer rating indicates one's experienced level between best (100) and worst (0) imaginable health status. The EQ 5D contains 5 dimensions: mobility, self-care, daily activities, pain/discomfort and depression/anxiety. Each dimension is rated at three levels: no problems, some problems and major problems. EQ 5D health states can be converted into utility scores. The WHOQOL is a 100-item self-report questionnaire, and through the domains of physical health, psychological health, environment, personal convictions, social relationships and extent of independency, the WHO concept of quality of life is assessed.

BPD-47, SCL-90, Young Schema Questionnaire. Other secondary outcome measures consisted of general psychopathologic measures and measures of ST personality concepts, all in self-report format and with robust psychometric properties. These measures included the *BPD Checklist* on the burden of BPD-specific symptoms (Giesen-Bloo, Arntz, et al., 2006a) and the *Symptom Checklist-90* for subjective experience of general psychopathology (Arrindell & Ettema, 1986; Derogatis, Lipman, & Covi, 1973). A theory specific instrument was the *Young Schema Questionnaire* on schemas underlying Young's theory (Rijkeboer, 2005, Rijkeboer, van den Bergh, & van den Bout, 2005; Schmidt, Joiner, Young, & Telch, 1995; Sterk & Rijkeboer, 1997).

Treatment adherence

Treatment adherence was monitored by means of supervision. All sessions were audio taped. The audio tapes were saved for evaluation. Of 62 patients 62 audio tapes between 5 and 12 months of treatment were randomly selected with 30 tapes from the condition with, and 28 tapes from the condition without TTA. There were 4 missing tapes, 3 because there were no tapes from these particular patients and one missing tape because of bad sound quality. Twenty separate tapes were rated by independent raters to assess the intra class correlation coefficient (ICC). All the raters were independent of the study and masked to treatment condition and outcome. The raters were psychologists trained in ST. We used the ST Therapy Adherence and Competence Scale for BPD (Young, Arntz, & Giesen-Bloo, 2006). This instrument consists of visual analog scale and Likert scale items and has a competence cutoff score of at least 60.

Registration of the phone contact

All therapists in the condition with phone support outside office hours had to monitor the telephone contacts on standardized forms with the following specifications: duration of the contact (minutes), time (weekday/nights or weekend), point of time (day, evening, night), reason of the phone contact (crisis, therapeutic, administrative). All contacts were registered and were used for calculating the number of therapeutic and crisis contacts outside office hours. The data will also be used for another yet to publish cost-outcome article.

Registration of therapy sessions

All sessions were audio taped. The number of sessions was monitored and the content of the sessions and used ST-techniques registered on standardized forms.

Problems during the implementation process

These were monitored by the researcher, recorded in a log book, and discussed with the project group during monthly meetings and with the therapists during the monthly supervision. Topics that were discussed were the experiences of therapists and research assistants with the project, no show or dropouts of patients, therapists and research assistants quitting the project, support of therapists by management, peers, and crisis facilities, and organizational changes influencing the implementation process like reorganization.

Analysis

The BPDSI-IV power calculation was based on the aim of showing a difference at the patient level between the conditions with extra phone support of the therapist outside office hours versus the condition without such support. Because we did not know what the effect of the extra phone support might be, it was decided to use a medium effect size of 0.5, according to Cohen (1977), for the power calculation. With a minimum of 2 conditions $\times$ 30 patients per condition, the power to demonstrate such a difference between the two conditions with two-tailed alpha of 0.05 is 84.

It will be concluded that extra crisis support is both helpful and clinically relevant if at the patient level a medium effect difference is found between the conditions with and without extra phone support. A possible small difference in effect would, although indicating that the extra support is helpful, probably not convince clinicians to implement this extra availability in their regular practice. Using Cohen's formula, effect sizes were calculated as $X_1 - X_2/\mathrm{SD}_{\mathrm{pooled}}$, where X_1 represents the pre-treatment scores, X_2 the post-treatment scores, and $\mathrm{SD}_{\mathrm{pooled}}$ represents the pooled standard deviations of the pre- and post-treatment scores.

The statistical analyses were based on the intention-to-treat as randomized principle. Treatment effects were tested with survival analysis for dichotomous variables, and one-sample *t*-tests and ANCOVAs for continuous variables with baseline as covariate and condition as between-group factor. As no severe deviations from

distributional assumptions were detected, parametric ANCOVAs were used.

All the tests were two-tailed with a significance level of 5%. Analyses were performed using the Statistical Package for Social Sciences, version 15.0 for Windows (survival analyses, within-group analyses, Chi-square tests)

Selection of the benchmark

A systematic review of the literature showed that only one RCT for individual ST could be found that was completed (Giesen-Bloo et al., 2006). One trial of ST for BPD used a group format (Farrell, Shaw, & Webber, 2009) and other RCTs of ST (Bamelis et al., 2006–2010; Bernstein, Arntz, & de Vos, 2007) are not on BPD and results are not available yet.

The characteristics of the RCT and the Implementation Study were highly comparable: Both studies had the same inclusion and exclusion criteria, the same outcome measures and both studies used experienced therapists. The sample size in the RCT was 44, in the implementation study 62, the drop out in the RCT was 18% and in the implementation study 21%.

Comparison of the treatment effects

In the RCT (Giesen-Bloo et al., 2006) a pre- to post-treatment effect size difference of $d = 1.24$ (Cohen's d) was found on the main patient outcome measure BPDSI on 18 months. In the present study, the same treatment was less intensive and executed in non-academic practice, so that a lower effectiveness was expected. We therefore tentatively estimated the pre–post-difference to $d = 1.0$.

Results

Treatment results

Sample characteristics

The patient flow is presented in Fig. 1.

Of 92 patients referred to the study centers, 30 patients (32.6% of referrals) were not eligible for participation: 17 patients did not meet the inclusion criteria (either because they had no BPD diagnosis or their BPDSI-IV scores were below 20) and 11 patients met exclusion criteria (1 bipolar disorder, 9 psychotic disorder and 1 ADHD). Another 2 patients were included in the study but could not be randomized because of logistic problems at one of the institutes (there were problems with the waiting list). Therefore 62 patients (67.4% of referrals) were included in the study.

Treatment and dropout

After 1.5 years of treatment 49 patients (79%) were still in treatment. Fig. 1 presents the moments in time that patients dropped out of treatment. One patient who was allocated to the treatment condition with extra crisis support committed suicide before she received the therapist's phone number that could be used outside office hours. As she passed away before the experimental manipulation was executed, her assessments were excluded from the condition-specific analyses. In the general analyses her assessments were included. Two patients dropped out of treatment due to prolonged illness of their therapist and due to uncertainty whether their therapist would return. They were unsatisfied with the support given by another therapist and decided to look for another treatment. Both refused to do the assessments. Two other patients refused to do the assessments, but were still in treatment. They were considered study dropouts, but not treatment dropouts. One patient successfully terminated treatment after one year. Nevertheless she refused to do the assessment at 18 months, because she found it too burdensome. Therefore she was considered a study, but not a treatment dropout.

Treatment groups at baseline

Table 1 shows baseline characteristics of the patients.

There were no significant differences between treatment conditions regarding age, sex, educational level, employment status and psychotropic medication. Most of the patients were women. There were only two men who participated in the study. The mean age of the participants was around thirty years. Most of them had average educational levels. About 58% of the patients used psychotropic medication. At baseline, the treatment groups only differed significantly on the BPDSI-IV. Numbers of co morbid axis-I and axis-II disorders were equally distributed at baseline. History of recent suicide planning, steps and/or attempts was similar between conditions.

Treatment outcomes

Results of the primary and secondary outcome measures

Results of the primary and secondary outcome measures are given in Table 2 and Fig. 2.

Although we used the last observation carried forward method (LOCF), Table 2 shows some missing data for the EuroQol thermometer, EQ 5D, SCL-90, WHOQol and YSQ. This was caused by patients who did the assessments but failed to complete some parts of the written questionnaires.

Main effects

Significant effects at 1.5 years of ST for the whole group emerged for patients' reduction of BPDSI scores ($M = 13.65$, SE $= 1.40$, $t(61) = 9.72$, $p < 0.001$, $d = 1.55$), reduction on the BPD-47 scores $M = 23.98$, SE $= 4.34$, $t(61) = 5.52$, $p = 0.00$, $d = 0.80$, SCL-90 scores $M = 37.21$, SE $= 9.38$, $t(59) = 3.97$, $p = 0.00$, $d = 0.57$ and Young Schema Questionnaire $M = 113.96$, SE $= 19.57$, $t(61) = 5.82$, $p = 0.00$, $d = 0.69$). Patients of both conditions improved significantly on all DSM-IV BPD criteria ($p < 0.001$ for abandonment, unstable relationships, identity disturbance, affect instability, emptiness and anger, for parasuicidality $p = 0.027$, impulsivity $p = 0.011$ and paranoid and dissociative ideation $p = 0.002$). There was also an improvement in quality of life on EQ 5D utility scores $M = 0.12$, SE $= 0.05$, $t(61) = 2.51$, $p = 0.02$, $d = 0.35$) and WHOQOL total score $M = -1.07$, SE $= 0.19$, $t(61) = 5.56$, $p < 0.001$, $d = 0.58$), there was no significant effect on improvement in quality of life measured by the EuroQol thermometer, $M = 4.53$, SE $= 2.47$, $t(59) = 1.84$, $p = 0.07$, $d = 0.23$. Although there were significant differences on all measures except the EuroQol thermometer at 1.5 years, a comparison of the assessments between 12 months and 18 months shows that patients did not make progress in this period or deteriorated. This could be temporary or could be due to the fact that sessions changed from twice a week to once a week. When we compare these results with the Giesen-Bloo study, we see the same results in their study in the period between 12 and 18 months.

Number of therapy sessions

The mean number of treatment sessions was 69 (SD 32.6) with a range from 1 to 142 sessions. The mean number of sessions in the condition with phone support outside office hours was 71 (SD 34.57) (range 2–142) and in the condition without phone support 67 (SD 30.85) (range 1–130) with no between-group difference ($t(60), = 0.46$, $p = 0.65$).

Table 1
Socio-demographic and clinical characteristics of 62 study participants.[a]

	Schema Focused Group, With extra phone support SFT$^+$ No. (%) ($N = 32$)	Schema Focused Group, Without extra phone support SFT$^-$ No. (%) ($N = 30$)	*P* value
Age, mean (SD), y	31.81 (9.24)	32.13 (9.01)	0.890[b]
Women	31 (96.9)	29 (96.7)	0.963[c]
Men	1 (3.1)	1 (3.3)	
Education			
Graduate/ professional	1 (3.1)	1(3.3)	
College graduate	3 (9.4)	5 (16.7)	0.909[c]
Some college	16 (50.0)	12 (40.0)	
High school graduate	9 (28.1)	9 (30.0)	
Grades 7–11	3 (9.4)	3 (10.0)	
Employment status			
House wife	4 (12.5)	3 (10.0)	
Student	3 (9.4)	0 (0.0)	
Employed	8 (25.0)	8 (26.7)	0.466[c]
Disability	14 (43.8)	14 (46.7)	
Welfare	3 (9.4)	5 (16.7)	
Psychotropic medication at baseline	19 (59.4)	17 (56.7)	0.829[c]
Recent suicide planning, steps and/or attempts[e]	10 (31.3)	8 (26.7)	0.691[c]
Recent non-suicidal self-injury[f]	18 (56.3)	15 (50.0)	0.622[c]
Meeting DSM-IV BPD criterion 5	19 (59.4)	17 (56.7)	0.829[c]
	Mean (SE [95%CI])	**Mean (SE [95%CI])**	
Number of Axis-I diagnoses	2.16 (0.27 [1.61–2.70])	2.40 (0.24 [1.90–2.90])	0.503[d]
Number of SCID II BPD criteria	6.63 (0.18 [6.25–7.00])	7.07 (0.24 [6.58–7.56]	0.146[d]
Number of treatment modalities before baseline[g]	2.19 (0.27 [1.63–2.74])	2.53 (0.30 [1.92–3.15]	0.396[d]

Abbreviation: BPD, borderline personality disorder; BPDSI-IV, Borderline Personality Disorder Severity Index, fourth version; CI, Confidence Interval; SCID II, Structured Clinical Interview for DSM-IV Axis-II Personality Disorders; M.I.N.I., Mini International Neuropsychiatric Interview, SE Standard Error.

[a] Data are given as number (percentage) except where otherwise indicated.
[b] Based on the *t*-test.
[c] Based on the Pearson χ^2 test.
[d] Based on analysis of variance.
[e] According to BPDSI-IV 5.11–5.13 over the previous three months.
[f] According to BPDSI-IV 5.1–5.8 over the previous three months.
[g] Range 0–6; individual treatment, group treatment, family/couples therapy, daily medication, clinical treatment and otherwise.

Treatment adherence

Fifty-eight tapes of the first year therapy period were rated (interrater reliability was assessed with twenty tapes that were rated twice).

Therapist's adherence to ST, i.e., overall appropriateness of used methods and techniques in ST, was excellent (mean = 94.90; SD = 3.26). The agreement between the raters was very high; all scores were between 88.30 and 99.55. Due to lack of variance across raters' scores ICC couldn't be calculated. If we, however, computed the interrater reliability according to the definition of Finn (1970, p. 73), cited in Tinsley and Weiss (1975), that takes the total possible range from 0 to 100 into account, we found an interrater reliability of 0.99. Therapist's mean competence/quality level for applying ST methods was good (mean 78.17; ICC = 0.79), as was the mean global competence/quality therapist rating (mean 77.60; ICC = 0.66). Analysis of both conditions showed no significant between-group difference regarding adherence and appropriateness of used methods and techniques [$t\,(56) = 1.41$, $p = 0.16$]. With respect to competence/quality level for applying ST methods a significant difference [$t\,(56) = 2.12$, $p = 0.038$] was found between the two groups of therapists. Mean competence/quality level for applying ST methods was 81.36 (SD 10.41) for the therapist group of the condition without phone support, and 75.19 (SD 11.65) for the therapist group of the condition with phone support. Regarding the mean global competence/quality level of the therapists there was also a significant difference (mean 80.89; SD 12.13) for the therapists of the condition without phone support, and (mean 74.53; SD 10.95) for the therapists of the condition with phone support [$t\,(56) = 2.10$, $p = 0.04$].

Condition with and without phone support

Number of phone contacts outside office hours in case of crisis and for therapeutic contacts

32 patients were allocated to the treatment condition with extra crisis support outside office hours. One patient who was allocated to this condition passed away before the experimental manipulation was executed. Therefore phone contacts of 31 patients were available. All the 'crisis' phone contacts outside office hours and the 'therapeutic' phone contacts outside office hours were added. The mean number of phone contacts outside office hours was 18.23 (4.15 [9.76–26.69]), range (0–79). Although the mean number of phone contacts outside office hours is 18, there are large differences between a group of patients ($n = 8$) who never phoned outside office hours and two patients who phoned 74 and 79 times.

Using Cohen's formula, large effect sizes for both conditions were found on the BPDSI-IV. ES for the condition with phone support was 1.42, ES for the condition without phone support was 1.82. Analysis of covariance (ANCOVA) with condition and baseline BPDSI-IV as covariates showed no significant difference between the two conditions $F\,(1, 58) = 0.44$, $p = 0.51$.

ANCOVA tests at the last observation of all BPDSI-IV subscales showed no difference between the two conditions with respect to all subscales.

ANCOVA on EuroQol thermometer of last observation means at 1.5 years with condition and baseline EuroQol thermometer as covariates showed no significant difference between the two conditions for 59 patients $F\,(1, 56) = 0.78$, $p = 0.38$, ANCOVA on EQ utility scores of last observation means at 1.5 years with condition and baseline EQ utility score as covariates showed no significant difference between the two conditions for 61 patients $F\,(1, 58) = 0.68$, $p = 0.41$, ANCOVA on WHOQol scores of last observation means at 1.5 years with condition and baseline WHOQol score as covariates showed no significant difference between the two conditions for 61 patients $F\,(1, 58) = 0.35$, $p = 0.56$, ANCOVA on BPD-47 scores of last observation means at 1.5 years with condition and baseline BPD-47 score as covariates showed no significant difference between the two conditions for 61 patients $F\,(1, 58) = 0.07$, $p = 0.79$, ANCOVA on SCL90 scores of last observation means at 1.5 years with condition and baseline SCL-90 score as covariates showed no significant difference between the two conditions for 59 patients $F\,(1, 56) = 0.03$, $p = 0.87$, ANCOVA on Young Schema Questionnaire scores of last observation means at 1.5 years with condition and baseline YSQ scores as covariates showed no significant difference between the two conditions for 61 patients $F\,(1, 58) = 2.07$, $p = 0.16$.

Because the results of the Treatment Adherence and Competence scale demonstrated significant differences between the two groups of therapists regarding main competence/quality level for applying ST methods and mean global competence/quality level, it

Table 2
Primary and secondary outcome measures in 61 study participants.[a]

	SFT with extra phone support No. (%) ($N = 31$)	SFT without extra phone support No. (%) ($N = 30$)	*P* value[b]
Recovery criterion 15, yes	13 (41.94)	13 (43.33)	0.91[b]
Reliable Change, yes	16 (51.61)	19 (63.33)	0.36[b]
	Mean (SE [95%CI])	**Mean (SE [95%CI])**	***P* value[c]**
BPDSI-IV total score (score range, 0–90)*			
Baseline	28.92 (1.05 [26.77–31.07])	32.73 (1.40 [29.88–35.59])	0.03
6-mo treatment	20.21 (1.48 [17.18–23.24])	24.70 (1.69 [21.23–28.16])	
12-mo treatment	16.37 (1.78 [12.74–20.00])	17.93 (1.51 [14.84–21.03])	
18-mo treatment	17.07 (1.95 [13.09–21.06])	16.77 (1.81 [13.06–20.48])	0.51
EuroQol-thermometer scores (range 0–100)**			
Baseline	53.35 (3.75 [45.70–61.01])	59.53 (3.14 [53.11–65.96])	0.21
6-mo treatment	64.94 (3.40 [57.99–71.88])	62.50 (3.02 [56.32–68.68])	
12-mo treatment	65.36 (4.17 [56.82–73.90]) $N = 29$	65.83 (2.99 [59.71–71.95]) $N = 29$	
18-mo treatment	61.03 (4.03 [52.80–69.27])	60.39 (3.65 [52.91–67.88]) $N = 28$	0.38
Utility score EQ (5) D**			
Baseline	0.44 (0.06 [0.31–0.56])	0.48 (0.05 [0.38–0.57])	0.63
6-mo treatment	0.60 (0.05 [0.49–0.70]) $N = 30$	0.64 (0.04 [0.55–0.73])	
12-mo treatment	0.67 (0.05 [0.56–0.78]) $N = 30$	0.63 (0.05 [0.52–0.74]) $N = 29$	
18-mo treatment	0.54 (0.07 [0.39–0.68])	0.62 (0.06 [0.50–0.74])	0.41
WHOQOL total scores (range 4–20)**			
Baseline	11.38 (0.31 [10.74–12.03])	10.71 (0.27 [10.17–11.26])	0.11
6-mo treatment	12.20 (0.40 [11.39–13.02])	11.39 (0.29 [10.80–11.99])	
12-mo treatment	12.63 (0.43 [11.75–13.51]) $N = 30$	12.12 (0.31 [11.49–12.75]) $N = 30$	
18-mo treatment	12.32 (0.40 [11.51–13.13])	11.97 (0.35 [11.26–12.68])	0.56
BPD-47*			
Baseline	111.39 (5.09 [100.99–121.79])	121.29 (4.96 [111.14–131.44])	0.17
6-mo treatment	94.52 (5.34 [83.60–105.43])	98.07 (5.27 [87.30–108.85])	
12-mo treatment	91.60 (6.04 [79.24–103.96]) $N = 30$	87.52 (5.28 [76.72–98.32]) $N = 30$	
18-mo treatment	90.94 (5.19 [80.33–101.54])	92.87 (6.51 [79.55–106.18])	0.79
SCL-90***			
Baseline	243.01 (10.97 [220.61–265.42])	254.49 (9.36 [235.32–273.67]) $N = 29$	0.43
6-mo treatment	210.02 (11.54 [186.44–233.59])	224.30 (9.50 [204.88–243.72])	
12-mo treatment	198.12 (11.66 [174.28–221.97]) $N = 30$	199.83 (10.96 [177.42–222.24])	
18-mo treatment	207.62 (12.34 [182.42–232.81])	214.93 (13.71 [186.85–243.00]) $N = 29$	0.87
YSQ L2***			
Baseline	686.12 (26.71 [631.56–740.68])	728.46 (26.61 [674.04–782.89])	0.27
6-mo treatment	615.93 (30.57 [553.50–678.37])	645.89 (31.83 [580.79–710.98])	
12-mo treatment	589.70 (34.74 [518.66–660.75]) $N = 30$	596.31 (30.64 [533.55–659.07]) $N = 29$	
18-mo treatment	603.05 (31.69 [538.33–667.76])	578.79 (34.41 [508.41–649.17])	0.16

Abbreviation: BPDSI-IV, Borderline Personality Disorder Severity Index, fourth version; CI, Confidence Interval; SE Standard Error; EuroQol, European Quality of Life, WHOQOl, World Health Organization Quality of Life Assessment; SCL-90, Symptoms Checklist-90; YSQ L2, Young Schema Questionnaire Long Version 2.
* Higher scores indicate more severe BPD pathology.
**Higher scores indicate higher levels of quality of life.
***Higher scores indicate more psycho- and personality pathology.
[a] Data are given as number (percentage) except where otherwise indicated.
[b] Based on the Pearson χ^2 test.
[c] Based on ANCOVA.

was decided to use these scores as covariates to find out if this might have influenced outcome.

Analysis of covariance (ANCOVA) with condition, baseline BPDSI-IV and main competence/quality level for applying ST methods as covariates showed no significant difference between the two conditions, $F(1, 54) = 0.12$, $p = 0.73$. Analysis of covariance (ANCOVA) with condition, baseline BPDSI-IV and mean global competence/quality level as covariates showed no significant difference between the two conditions, F (1, 54) = 0.22, p = 0.64. ANCOVAs for the BPD-47, EuroQol thermometer, EQ 5D utility scores, WHOQol scores, SCL-90 and Young Schema Questionnaire scores with the same covariates neither demonstrated significant differences.

Recovery criterion and reliable change

Survival analysis on the BPDSI-IV recovery criterion with condition and baseline BPDSI-IV as predictors (covariates) showed no significant differences between the two conditions. Also, when

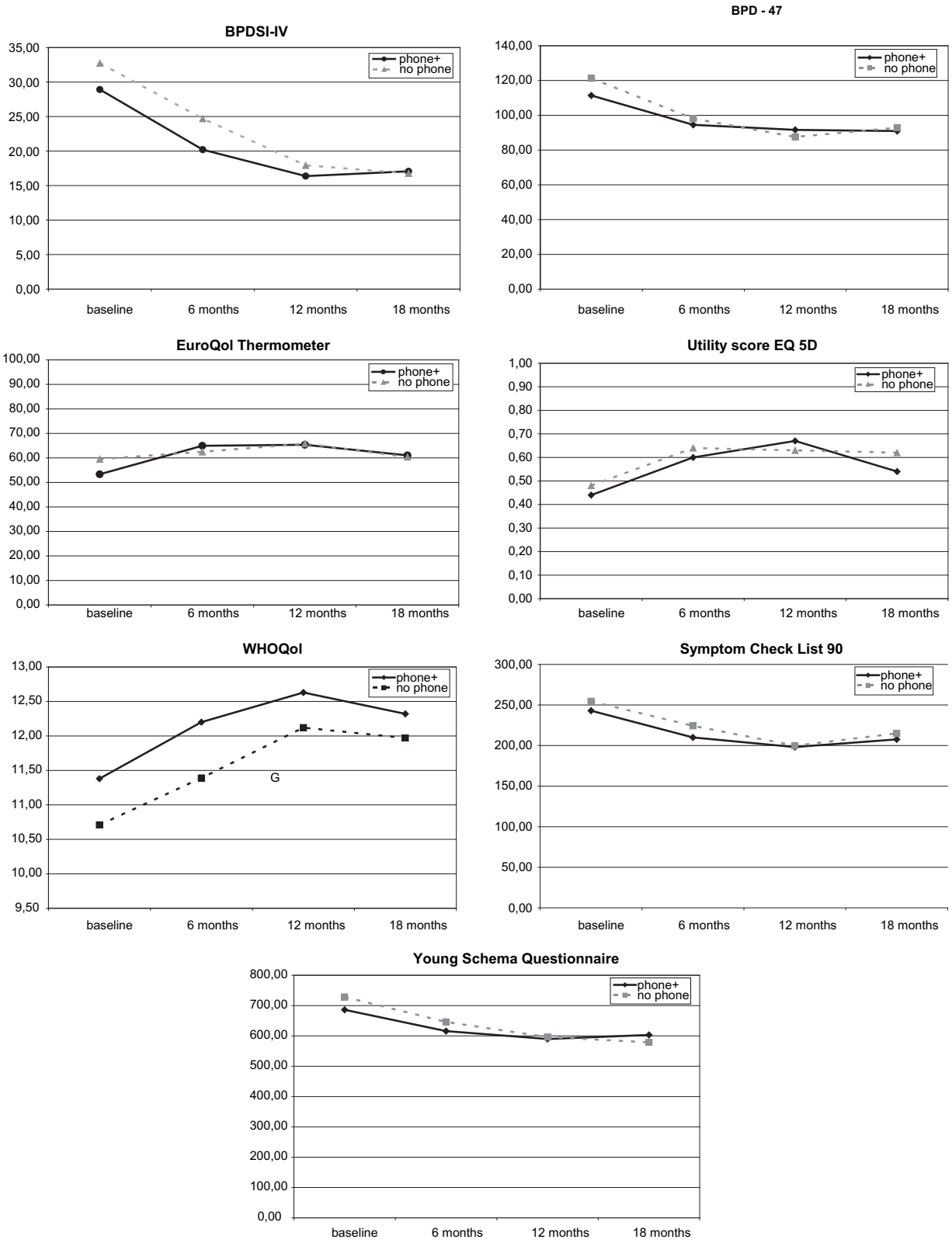

Fig. 2. Mean primary and secondary outcome measure scores. BPDSI-IV, BPD-47, EuroQoL, EQ 5D, WHOQOL, SCL-90, and YSQ.

Table 3
Survival analysis on recovery criterion and reliable change.

Recovery for 62 patients: therapist telephone availability effects					
$N = 61$	Wald	*P* value	Exp (B)	95% CI for Lower	Exp (B) Upper
Predictors					
Phone	0.032	0.857	0.932	0.432	2.012
Phone, BPDSI	0.003	0.957	0.978	0.446	2.148
Phone, medication	0.050	0.823	0.916	0.424	1.979
Phone, BPDSI, medication	0.019	0.890	0.946	0.431	2.077
Recovery for 61 patients: therapist telephone availability effects					
$N = 61$	Wald	*P* value	Exp (B)	95% CI for Lower	Exp (B) Upper
Phone	0.080	0.777	0.895	0.414	1.932
Phone, BPDSI	0.029	0.865	0.934	0.423	2.061
Phone, medication	0.096	0.757	0.885	0.410	1.913
Phone, BPDSI, medication	0.055	0.814	0.910	0.413	2.004
Reliable change for 62 patients: therapist telephone availability effects					
$N = 62$	Wald	*P* value	Exp (B)	95% CI for Lower	Exp (B) Upper
Phone	0.500	0.480	1.272	0.653	2.477
Phone, BPDSI	0.079	0.779	1.104	0.552	2.208
Phone, medication	0.572	0.449	1.295	0.663	2.532
Phone, BPDSI, medication	0.081	0.776	1.107	0.550	2.226
Reliable change for 61 patients: therapist telephone availability effects					
$N = 61$	Wald	*P* value	Exp (B)	95% CI for Lower	Exp (B) Upper
Phone	0.345	0.557	1.221	0.627	2.378
Phone, BPDSI	0.002	0.962	1.017	0.504	2.054
Phone, medication	0.406	0.524	1.243	0.637	2.427
Phone, BPDSI, medication	0.004	0.951	1.022	0.504	2.073

psychotropic medication was entered as a covariate, no significant differences were found between the two conditions (see Table 3).

At 1.5 years, 41.9% of the condition with and 43.3% of the condition without telephone support outside office hours were recovered. The overall recovery rate of 42% was higher but not significantly different from the recovery rate of 27% at 1.5 years in the Giesen-Bloo et al. trial, $\chi^2\,(1) = 2.41$, $p = 0.12$.

Survival analysis on reliable change status with baseline BPDSI-IV as covariate showed no significant difference between the conditions. Also, when psychotropic medication was entered as a covariate, no significant differences were found between the two conditions, see Table 3. At 1.5 years, 51.6% of the condition with and 63.3% of the condition without telephone support outside office hours were reliably changed. The overall reliable change rate of 56.5% was not significantly different from the reliable change rate of 45.5% at 1.5 years in the Giesen-Bloo et al. trial, $\chi^2\,(1) = 0.04$, $p = 0.26$.

Comparison between the Implementation Study and the RCT

A comparison of the baseline characteristics of the patients of the Implementation Study and the RCT shows that both groups were reasonably comparable, but the Implementation sample had on average lower BPDSI scores, used less medication, and had higher quality of life scores, suggesting that it was slightly less severe.

In Table 4 and Fig. 3 a comparison is given between the primary and secondary outcome measures of the Implementation Study and the RCT.

Table 4 and Fig. 3 show that the expected shrinkage in effect was not found.

Table 4
Primary and secondary outcome measures implementation and RCT.[a]

	Implementation No. (%) ($N = 62$)[1]	RCT No. (%) ($N = 44$)	*P* value[b]
Recovery criterion 15, yes	26 (42%)	12 (27%)	0.12
Reliable Change, yes	35 (56.5%)	20 (45.5%)	0.26
Drop out rate	13 (21%)	8 (18%)	0.17
Effect size on BPDSI	$d = 1.55$	$d = 1.24$	
	Mean (SD)	**Mean (SD)**	
BPDSI-IV total score (score range, 0–90)*			
Baseline	30.92 (7.01)	33.14 (7.13)	
6-mo treatment	22.67 (9.14)	24.34 (11.33)	
12-mo treatment	17.48 (9.42)	22.41 (10.40)	
18-mo treatment	17.27 (10.60)	21.90 (11.02)	
EuroQol-thermometer scores (range 0–100)**	Mean	Median	
Baseline	56.13 (19.21)	50	
6-mo treatment	63.35 (17.80)	50	
12-mo treatment	65.16 (19.48)	56	
18-mo treatment	60.38 (20.84)	55	
Utility score EQ 5 D**			
Baseline	0.44 (0.31)	0.46 (0.31)	
6-mo treatment	0.60 (0.28)	0.50 (0.31)	
12-mo treatment	0/63 (0.31)	0.54 (0.32)	
18-mo treatment	0.56 (0.37)	0.55 (0.32)	
WHOQOL total scores (range 4–20)**			
Baseline	11.04 (1.63)	10.33	
6-mo treatment	11.77 (1.96)	10.34	
12-mo treatment	12.33 (2.05)	11.17	
18-mo treatment	12.11 (2.06)	10.92	
BPD-47*			
Baseline	116.48 (27.82)	Composite score	
6-mo treatment	96.81 (29.20)		
12-mo treatment	90.22 (31.06)		
18-mo treatment	92.50 (32.25)		
SCL-90***			
Baseline	250.49 (57.54)	Composite score	
6-mo treatment	219.44 (61.05)		
12-mo treatment	201.71 (64.58		
18-mo treatment	213.69 (72.86)		
YSQ L2***			
Baseline	709.40 (147.66)	Composite score	
6-mo treatment	634.35 (172.51)		
12-mo treatment	597.39 (178.61)		
18-mo treatment	595.44 (183.04)		

Abbreviation: BPD, borderline personality disorder; BPDSI-IV, Borderline Personality Disorder Severity Index, fourth version; CI, Confidence Interval; SCID II, Structured Clinical Interview for DSM-IV Axis-II Personality Disorders; M.I.N.I., Mini International Neuropsychiatric Interview, SE Standard Error.
[1]According to BPDSI-IV 5.11–5.13 over the previous three months.
[2]According to BPDSI-IV 5.1–5.8 over the previous three months.
[3]Range 0–6; individual treatment, group treatment, family/couples therapy, daily medication, clinical treatment and.
[1]Due to missing values not always 62 respondents.
[a] Data are given as number (percentage) except where otherwise indicated.
[b] Based on the Pearson χ^2 test.

Fig. 4 shows that the results of the recovery rates and the reliable change rates of both studies are highly comparable.

Qualitative study results

During the monthly supervision meetings the experiences of the therapists were discussed. One of the topics was the phone support outside office hours. For some therapists the phone support was very burdensome, especially when patients were suicidal, angry or phoned when they had been drinking. Another topic was no show. A group of young patients (between 18 and 21 years) came to treatment on a very irregular basis. Although therapists explained

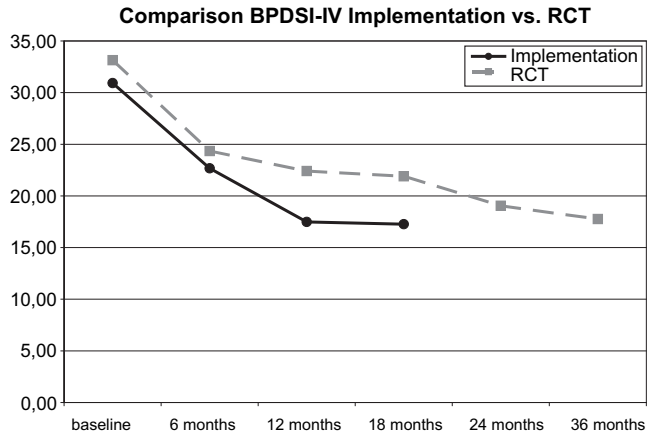

Fig. 3. BPDSI – IV comparison RCT and implementation.

them the rationale of the model (sessions twice a week during the first year) and motivated them in a very personal way (limited reparenting) some of these patients only came once a month. We had the impression that for some young borderline patients a therapy program with sessions twice a week was not feasible. Not only the therapists, but also the research assistants suffered from the no show of some patients. Some patients cancelled the half-yearly assessments more than ten times for different reasons. Our impression was that for some patients the assessments were too confronting, especially for those patients who did not make any progress.

Discussion

The present study had several aims. First, to assess the effectiveness of a less intensive form of ST for BPD when implemented in regular health care and compare the results with the originally study by van Giesen-Bloo et al. as a benchmark RCT. Second, to assess whether there was a clinically relevant effect of therapist availability for crisis support outside office hours in the implemented treatment. Third, to describe the impact of the implementation and the problems that emerged during the implementation process. As to the first aim, no evidence for the expected shrinkage of effectiveness was found. In this study the effects were comparable to the RCT and the implemented treatment was at least as effective as 1.5-year ST in the original trial. Similarly, treatment retention was very similar to that of the original trial (81% in the implementation study versus 82% in the RCT in 1.5 years). There are several possible factors that may have contributed to these positive results. During recent years negative views on BPD and its treatability have changed, ST has become well known in the Netherlands, and ST has become more embedded in general mental healthcare than during the period that the RCT was conducted. Another possible contributing factor is the structured training program supported by the DVDs demonstrating the therapy techniques and the monthly supervision on location. Also the Dutch translation of the Schema Therapy Practitioners Guide by Young et al. (2003) in 2005 and the publication of Schema Therapy for Borderline Personality Disorder protocol by Van Genderen and Arntz in 2005 made it easier for the therapists to learn about ST. Another contributing factor might be the use of psychotropic medication at baseline. In the implementation study 58% of the patients used psychotropic medication at baseline, compared with 77% of the patients in the RCT. Since the findings of the RCT by van Giesen-Bloo showed that psychotropic medication use had a significant negative association with recovery this could be a possible explanation why the recovery rate in the implementation study is higher than in the RCT. Note however that in the implementation study no difference in effectiveness was found between the group that used psychotropic medication compared to the group that did not use medication. Lastly, patients in the present study were on average less severe than those of the Giesen-Bloo et al. trial. This might relate to quicker recovery.

As to the second aim, no significant differences between the two conditions appeared, indicating that within 1.5 years of outpatient ST this extra crisis support did not have a medium or larger effect on treatment outcome. Treatment retention was also highly

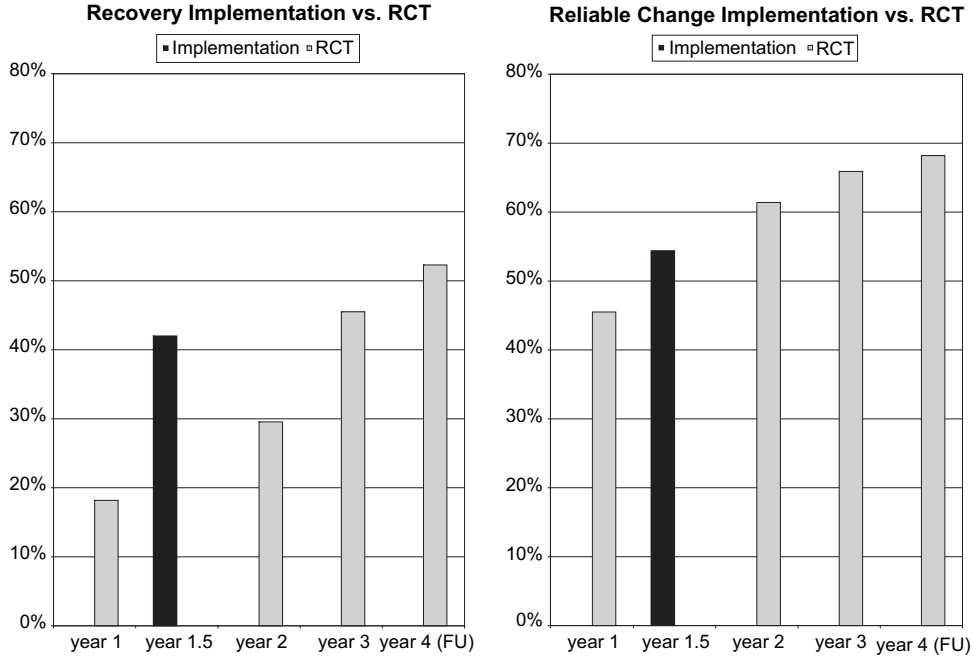

Fig. 4. Comparison: recovery and reliable change in the implementation study and the RCT.

comparable between conditions, thus there was no evidence that the extra crisis support prevented early dropout from treatment.

The implementation study was powered to demonstrate a medium or higher effect of TTA. The failure to detect any difference between conditions, does not mean that they are equivalent, only that differences, if any, would be small. A small difference however does not imply lack of clinical significance. A power analysis shows that a sample of more than 3100 patients is necessary to detect a small effect with a significance level of 0.05 and a power of 80. According to Rothwell (1995) overall results may not always be generalisable to individual patients. If we would be able to predict which patients might need TTA to recover, more focused tests could of course be set-up.

A remarkable finding was the relatively small number of phone contacts outside office hours with the exception of two patients. One of the explanations can be the good mental healthcare system in the Netherlands. When patients are in crisis, they can contact a general practitioner or the emergency service of a hospital, so they don't always need a therapist. The number of emergency contacts and contacts with GP's will be reported in another paper on cost-effectiveness. In some other healthcare systems where it is more complicated to organize good care the phone support outside office hours might be of greater importance. Given the results of this implementation study that no medium effect of the phone support could be demonstrated and the heavy burden for the therapist we don't think that nationwide implementation of ST with TTA is recommended in the Netherlands. We would suggest offering phone support outside office hours only in special cases. Therapists doing so might need extra training and support to set limits to the small number of patients that overuse TTA.

Regarding the results of the Treatment Adherence and Competence scale we would like to make some comments. The treatment integrity check took place at the end of the study. When patients were randomized nothing was known yet about a possible difference regarding the competence of the therapists. All therapists received the same training program and supervision.

Because the results of the Treatment Adherence and Competence scale demonstrated significant differences between the two groups of therapists regarding main competence/quality level for applying ST methods and mean global competence/quality level, it was decided to use these scores as covariates to find out if this might have influenced outcome. The results showed that after controlling for these variables no significant differences were found on outcome scores, meaning that the lack of difference between the conditions with phone support versus the condition without phone support could not be explained by the difference in competence scores of the therapists.

One of the limitations of the present study is that it assessed a highly structured and controlled implementation program, with a structured and intensive training, supervision by an expert, peer supervision, high adherence to the protocol, and institutes and managers supporting the implementation of ST. This is without doubt how implementation should be done and shows that with a good implementation strategy, results are very promising. Our impression is that it may not be fully representative how ST is actually applied in regular practice. Future studies should investigate what effects therapists obtain with BPD patients if they apply ST (or their own version of ST) in a less structured and supportive context. If effects are smaller, such studies might learn what the essential factors are that are responsible for the shrinkage of effects.

Another limitation is the question whether the treatment results will also remain sufficient in the long term. Given the rather high turnover of therapists and managers the implementation might lose some of its strength and effectiveness. Therefore it is important to analyze the results of the 1.5-year follow-up period after the initial 1.5-year treatment period. This follow-up period will be completed in August 2010.

Another limitation of this study may be the generalizability to countries with different healthcare systems. In the Netherlands, when patients are in crisis, they can contact a general practitioner or the emergency service of a hospital, so they don't always need a therapist. In some other healthcare systems where it is more complicated to organize good care the phone support outside office hours may be of greater importance.

To conclude, the results of this study suggest that implementing outpatient schema therapy for borderline patients in general mental healthcare can be successful and that treatment results were comparable with the RCT results. The positive outcome should facilitate the decision for healthcare providers to adopt ST for BPD in their institution so that more BPD patients can recover. To our knowledge, this is the first well-designed 1.5-year controlled implementation study for BPD that investigates the extra phone support outside office hours. An additional 1.5-year follow-up period after the initial 1.5-year treatment period will be completed in August 2010.

Funding/support

This research was financially supported by the Health Care Efficiency Research Program: subprogram Implementation (ZonMw) (Grant 945-16-313).

Role of the sponsor

The sponsor played no role in the data collection and analysis, manuscript preparation, or authorization for publication.

Dutch trial registry

NTR: TC = 1781.

Acknowledgements

We wish to acknowledge the contributions of the participating BPD patients, trainers, therapists and research assistants. Furthermore we acknowledge the statistical advice of Adriaan Hoogendoorn, PhD.

References

American Psychiatric Association. (2005). *Diagnostic and statistical manual of mental disorders*. text revision (4th ed.). Washington, DC: American Psychiatric Association.

Arntz, A., van den Hoorn, M., Cornelis, J., Verheul, R., van den Bosch, W., & de Bie, A. (2003). Reliability and validity of the borderline personality disorder severity index. *Journal of Personality Disorder, 17*(1), 45–59.

Arntz, A., & van Genderen, H. (2009). *Schema therapy for borderline personality disorder*. Chichester: Wiley.

Arrindell, W. A., & Ettema, J. H. M. (1986). *Klachtenlijst (SCL-90)*. Lisse: Swets & Zeitlinger.

Bamelis, L., Arntz, A., Bernstein, D., Widdershoven, G., Spinhoven, P., Evers, S., et al. (2006–2010). *Psychological treatments of personality disorders; a multicentered randomized controlled trial on the (cost-) effectiveness of schema-focused therapy*. Promotietraject Universiteit Maastricht.

Bateman, A., & Fonagy, P. (1999). Effectiveness of partial hospitalisation in the treatment of borderline personality disorder: a randomised controlled trial. *American Journal of Psychiatry, 156*, 1563–1569.

Bateman, A., & Fonagy, P. (2004). *Psychotherapy for borderline personality disorder. Mentalization-based treatment*. New York: Oxford University Press.

Bernstein, D., Arntz, A., & de Vos, M. E. (2007). Schemagerichte therapie in forensische settings: Theoretisch model en richtlijnen voor best clinical practice. *Tijdschrijft voor Psychotherapie, 33*, 120–139.

Brooks, R. (1996). EuroQol: the current state of play. *Health Policy, 37*, 53–72.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences (revised edition)*. New York: Academic Press.

Derogatis, L. R., Lipman, R. S., & Covi, L. (1973). SCL-90: an outpatient psychiatric rating scale – preliminary report. *Psychopharmacology Bulletin, 9*, 13–28.

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care, 35*, 1095–1108.
EuroQol Group. (1990). EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy, 16*, 199–208.
Farrell, J. M., Shaw, I. A., & Webber, M. A. (2009). A schema-focused approach to group psychotherapy for outpatients with borderline personality disorder: a randomized controlled trial. *Journal of Behavior Therapy and Experimental Psychiatry, 40*(2), 317–328.
First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *User's guide for the structured clinical interview for DSM-IV axis II personality disorders (SCID-II)*. Washington, DC: American Psychiatric Press.
van Genderen, H., & Arntz, A. (2005). *Schemagerichte cognitieve therapie bij borderline persoonlijkheidsstoornis*. Amsterdam: Nieuwezijds.
Giesen-Bloo, J. H., Arntz, A., van Dyck, R., Spinhoven, P., & van Tilburg, W. (2001). Outpatient treatment of borderline personality disorder: analytical psychotherapy versus cognitive behavior therapy. Paper presented at the World Congress of Behavioral and Cognitive Therapies, July 17–21, Vancouver, Canada.
Giesen-Bloo, J., Arntz, A., & Schouten, E. (2006a). The Borderline Personality Disorder Checklist: psychometric evaluation and factorial structure in clinical and nonclinical samples. In J. Giesen-Bloo (Eds.), *Crossing borders; theory, assessment and treatment in borderline personality disorder* (pp. 85–101). [Proefschrift]. Maastricht: Universitaire Pers.
Giesen-Bloo, J., Wachters, L., Schouten, E., Arntz, A. (2006b). Assessment of borderline personality disorder with the Borderline Personality Disorder Severity Index-IV: psychometric evaluation and dimensional structure. In J. Giesen-Bloo (Eds.), *Crossing borders; theory, assessment and treatment in borderline personality disorder* (pp. 69–83). [Proefschrift]. Maastricht: Universitaire Pers.
Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., Asselt, Th., et al. (2006). Outpatient psychotherapy for borderline personality disorder: a randomized trial of schema-focused therapy vs. transference-focused psychotherapy. *Archives of General Psychiatry, 63*, 649–658.
Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., et al. (2009). One-year follow-up of schema focused therapy and transference focused psychotherapy for BPD, and the influence of drop-out status, treatment status and medication.
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.
Linehan, M. M. (1993a). *Cognitive–behavioral treatment of borderline personality disorder*. New York: The Guilford Press.
Linehan, M. M. (1993b). *Skills training manual for treating borderline personality disorder*. New York: The Guilford Press.
Linehan, M. M., Arnstrong, H. E., Suarez, A., Allmon, D., & Heard, H. L. (1991). Cognitive–behavioural treatment of chronically parasuicidal borderline patients. *Archives of General Psychiatry, 48*, 1060–1064.
Linehan, M. M., Comtois, K. A., Murray, A. M., Brown, M. Z., Gallop, R. J., Heard, H. L., et al. (2006). Two year randomized trial and follow-up of dialectical behaviour therapy vs. treatment – by experts for suicidal behaviours and borderline personality disorder. *Archives of General Psychiatry, 63*(7).
Lobbestael, J., Arntz, A., & Sieswerda, S. (2005). Schema modes and childhood abuse in borderline and antisocial patients. *Journal of Behaviour Therapy and Experimental Psychology, 36*, 240–253.
Lobbestael, J., van Vreeswijk, M. F., & Arntz, A. (2008). An empirical test of mode conceptualisations in personality disorders. *Behaviour Research and Therapy, 46*, 854–860.
Nadort, M. (2005). Schematherapie voor de Borderline Persoonlijkheidsstoornis. Therapietechnieken. DVD-box.
Nadort, M., van Dyck, R., Smit, J. H., Giesen-Bloo, J., Eikelenboom, M., Wensing, M., et al. (2009). Three preparatory studies for promoting implementation of outpatient schema therapy for borderline personality disorder in general mental health care. *Behaviour Research and Therapy, 47*(11), 938–945.
Nadort, M., & Giesen-Bloo, J. (2005). *Pilot implementation study of ST for borderline patients*. Eindrapportage college voor Zorgverzekeraars.
Paris, J. (1993). The treatment of borderline personality disorder in light of research on its long term outcome. *Canadian Journal of Psychiatry, 38*, S28–S34.
Paris, J. (2008). Clinical trials of treatment for personality disorders. *The Psychiatry Clinics of North America, 31*(3), 517–526.
Pinninti, N. R., Madison, H., Musser, E., & Rissmiller, D. (2003). MINI International Neuropsychiatric Schedule: clinical utility and patient acceptance. *European Psychiatry, 18*(7), 361–364.
Rijkeboer, M. M., van den Bergh, H., & van den Bout, J. (2005). Stability and discriminative power of the Young Schema Questionnaire in a Dutch clinical versus non-clinical population. *Journal of Behavior Therapy and Experimental Psychiatry, 36*(2), 129–144.
Rijkeboer, M. M. (2005). *Assessment of early maladaptive schemas. On the validity of the Dutch schema-questionnaire*. Academisch proefschrift. Universiteit Utrecht.
Rothwell, P. M. (1995). Can overall results of clinical trials be applied to all patients? *The Lancet, 345*, 1616–1619.
Schmand, B., Lindeboom, J., & van Harskamp, F. (1992). *NLV, Nederlandse Leestest voor Volwassenen, Handleiding*. [*DART, Dutch adult reading test, manual*]. Lisse: Swets & Zeitlinger.
Schmidt, N. B., Joiner, T. E., Young, J. E., & Telch, M. J. (1995). The Schema Questionnaire: investigation of psychometric properties and the hierarchical structure of a measure of maladaptive schemas. *Cognitive Therapy and Research, 19*, 295–321.
Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59*(Suppl. 20), 22–33.
Sterk, F., & Rijkeboer, M. M. (1997). *Schema-Vragenlijst*. [*Schema questionnaire*]. Utrecht: Ambulatorium Utrecht University.
Ten Have, M. L., Lorsheyd, J. J. G., van Bijl, R., & Osterthun, P. (1995). *Jaarboek Geestelijke Gezondheidszorg 1995/1996*. [*Annual report mental health care 1995/1996*]. Utrecht: De Tijdstroom.
Tinsley, H. E. A., & Weiss, D. J. (1975). *Journal of Counseling Psychology, 22*(4), 358–376.
Van Asselt, A. D. I., Dirksen, C. D., Arntz, A., Giesen-Bloo, J. H., van Dyck, R., Spinhoven, P., et al. (2008). Outpatient psychotherapy for borderline personality disorder: cost effectiveness of schema-focused therapy versus transference focused psychotherapy. *British Journal of Psychiatry, 192*, 450–457.
Van Asselt, A. D. I., Dirksen, C. D., Arntz, A., & Severens, J. L. (2007). The cost of borderline personality disorder: societal cost of illness in BPD-patients. *European Psychiatry, 22*, 354–361.
Verheul, R., van den Bosch, M. C., Koeter, M. W. J., de Ridder, M. A. J., Stijnen, T., & van den Brink, W. (2003). Efficacy of dialectical behavior therapy: a Dutch randomised controlled trial. *British Journal of Psychiatry, 182*, 135–140.
Website for schematherapy: website voor schematherapie: www.schematherapie.nl (2005).
Weersing, V. R. (2005). Benchmarking the effectiveness of psychotherapy: program evaluation as a component of evidence-based practice. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*, 1058–1062.
Weertman, A., Arntz, A., & Kerkhofs, M. L. M. (2000). *Gestructureerd klinisch interview voor DSM-IV persoonlijkheidsstoornissen (SCID-II)*. [*Structured clinical interview for DSM-IV axis II personality disorders*]. Lisse: Swets & Zeitlinger.
WHOQOL Group TW. (1998). The World Health Organization Quality of Life assessment (WHOQOL): development and general psychometric properties. *Social Science & Medicine, 46*(12), 1569–1585.
Wilson, G. T. (1995). Empirically-validated treatments as a basis for clinical practice: problems and prospects. In S. C. Heyes, V. M. Follete, R. M. Dawes, & K. E. Grady (Eds.), *Scientific standards of psychological practice: Issues and recommendations* (pp. 163–196). Reno, NV: Context Press.
Yeomans, F. E., Clarkin, J. F., & Kernberg, O. F. (2002). *A primer for transference focused psychotherapy for the borderline patient*. Northvale, NJ: Jason Aronson.
Young, J., Arntz, A., & Giesen-Bloo, J. (2006). *Therapy adherence and competence scale*. http://www.epp.unimaas.nl. Accessed 1.5.2006.
Young, J. E., & Klosko, J. S. (1999). *Leven in je leven. Leer de valkuilen in je leven kennen*. Lisse: Swets & Zeitlinger.
Young, J. E., Klosko, J., & Weishaar, M. E. (2003). *Schema therapy: A practioner's guide*. New York: Guilford.
Young, J. E., Klosko, J. S., & Weishaar, M. E. (2005). *Schemagerichte therapie; handboek voor therapeuten*. Houten: Bohn Stafleu van Loghum.