



Prediction of treatment discontinuation and recovery from Borderline Personality Disorder: Results from an RCT comparing Schema Therapy and Transference Focused Psychotherapy



Arnoud Arntz ^{a, b, *}, Snežana Stupar-Rutenfrans ^{b, c}, Josephine Bloo ^{b, d}, Richard van Dyck ^e, Philip Spinhoven ^f

^a Department of Clinical Psychology, University of Amsterdam, The Netherlands

^b Department of Clinical Psychological Science, Maastricht University, The Netherlands

^c International Media and Entertainment Management Academy for Digital Entertainment NHTV University of Applied Sciences, Breda, The Netherlands

^d Community Mental Health Center Virenze RIAGG, Vaals, The Netherlands

^e VU Medical Centre Amsterdam, The Netherlands

^f Institute of Psychology, Leiden University, Leiden, The Netherlands

ARTICLE INFO

Article history:

Received 6 May 2015

Received in revised form

12 September 2015

Accepted 14 September 2015

Available online 21 September 2015

Keywords:

Borderline Personality Disorder

Outcome prediction

Treatment discontinuation

Dissociation

Schema Therapy

Transference Focused Psychotherapy

ABSTRACT

Knowing what predicts discontinuation or success of psychotherapies for Borderline Personality Disorder (BPD) is important to improve treatments. Many variables have been reported in the literature, but replication is needed and investigating what therapy process underlies the findings is necessary to understand why variables predict outcome. Using data of an RCT comparing Schema Therapy and Transference Focused Psychotherapy as treatments for BPD, variables derived from the literature were tested as predictors of discontinuation and treatment success. Participants were 86 adult outpatients (80 women, mean age 30.5 years) with a primary diagnosis of BPD who had on average received 3 previous treatment modalities. First, single predictors were tested with logistic regression, controlling for treatment type (and medication use in case of treatment success). Next, with multivariate backward logistic regression essential predictors were detected. Baseline hostility and childhood physical abuse predicted treatment discontinuation. Baseline subjective burden of dissociation predicted a smaller chance of recovery. A second study demonstrated that in-session dissociation, assessed from session audiotapes, mediated the observed effects of baseline dissociation on recovery, indicating that dissociation during sessions interferes with treatment effectiveness. The results suggest that specifically addressing high hostility, childhood abuse, and in-session dissociation might reduce dropout and lack of effectiveness of treatment.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Although the last decades showed a reduction in pessimism about the possibilities to treat Borderline Personality Disorder (BPD), BPD remains one of the more challenging psychiatric disorders, often needing specialized long-term treatment. Effective psychological treatments have been developed, but premature discontinuation of treatment and lack of effectiveness of treatments in a substantial proportion of patients remain areas where improvements can be achieved. Many studies have tried to detect

predictors of discontinuation and treatment success. Although causality usually remains an issue to be further investigated, knowledge of predictors of treatment discontinuation and success is a first step in the process to develop more acceptable and effective treatments. Unfortunately, the field of psychotherapy prediction studies is characterized by a lack of attempts to replicate findings of previous studies, so that it is unclear whether findings were accidental or not.

In the area of (B)PD treatment, a recent review (McMurrin, Huband, & Overton, 2010) reported the following patient-related pretreatment characteristics in single or multiple studies being prognostic of treatment discontinuation, which we grouped in the following categories.

Biographical variables: younger age; lower education; lower

* Corresponding author. Department of Clinical Psychology, University of Amsterdam, PO Box 19268, 1000 GG Amsterdam, The Netherlands.

E-mail address: A.R.Arntz@UvA.nl (A. Arntz).

occupational level; unemployment; juvenile conviction; parents divorced before patient was 10; spending less time alone; being in a relationship for less than 6 months.

Treatment history: less previous experience with mental health care; prior hospitalization.

BPD features: BPD-severity; impulsivity; fewer suicide attempts; current/previous substance abuse.

Defense mechanisms and personality organization: poor ego structure; defense level.

Axis-II comorbidity: more PD diagnoses/traits; dependent, obsessive–compulsive, histrionic, or antisocial PD; Cluster-A or B PDs; no specific PD; narcissism.

Negative emotional problems: less depression (level/diagnosis); high trait anxiety.

Childhood abuse and neglect: childhood emotional neglect.

Others: pre-contemplation stage of change; lower general level of functioning; low problem solving capacities; better social competence; lower persistence; problems in one area vs. several areas; and higher avoidance (one study experiential avoidance, another harm avoidance). Contradicting the last, avoidant PD was associated with treatment completion. Conflicting results between studies were found as to interpersonal distress.

In addition, we detected in the literature the following predictors of treatment discontinuation: male gender (Arntz, 1999; Links, Mitton, & Steiner, 1990; Löffler-Stastka et al., 2003; Thormählen et al., 2003); being single/divorced (Links et al., 1990); and anger-hostility (Smith, Koenigsberg, Yeomans, Clarkin, & Selzer, 1995; Rüscher et al., 2008 (trend)).

As to the prediction of treatment response, Gunderson et al. (2006) and Barnicott et al. (2012) offer overviews of empirical findings. Gunderson et al. found similar predictors as for treatment discontinuation. Additional predictors for worse outcome included: low IQ; affective instability; distractibility; family mental illness; early psychiatric contact; various forms of childhood maltreatment and loss; disability insurance benefits. In Barnicott et al.'s review it is concluded that higher severity predicts larger improvements, and that demographic variables generally have no influence (though some studies report effects of age, gender or employment), as does social adjustment. Negative effects were found for suicide attempts (two studies), and mixed effects for axis-1 symptom severity, self-harm (four studies no effect, two negative and one positive effect), medication use, and dissociation. As to dissociation, one study found that baseline level of dissociative symptoms predicted worse outcome (Kleindienst et al., 2011); however another (smaller) study did not find evidence for dissociation predicting less treatment response (Braakmann et al., 2007). Barnicott et al. comment that mixed findings might be related to whether change or remission is taken as outcome. Gunderson et al.'s (2006) own empirical findings indicated that BPD-severity, level of functioning, childhood trauma, and quality of current relationships (incl. BPD traits 1 and 2) predict outcome. In addition, we found two additional studies with contradictory findings as to self-injury (Chiesa & Fonagy, 2007; Plakun, 1991).

As said, one of the major problems with this overwhelming list of discontinuation and treatment success predictors is the lack of replication. Moreover, usually only univariate analyses were done, obscuring which predictors are essential. Systematic replication of previous findings is needed followed by meta-analytic studies to disentangle which predictors are robust, and which might have been accidental findings. Moreover, there is a need for studies that investigate *how* predictors affect the treatment. Without understanding “the how” little theoretical and clinical progress can be made.

The present paper consists of two related studies. Study 1 aimed to test previously reported predictors. Using data of a multicenter RCT on the effectiveness of Schema Therapy (ST) and Transference

Focused Psychotherapy (TFP) for BPD (Giesen-Bloo et al., 2006), we tested previously reported predictors (if included in our baseline assessment) of treatment discontinuation and outcome (recovery). In short, results showed that treatment dropout from TFP was significantly higher than from ST (50.0% vs. 27.3%), and that ST was superior to TFP in terms of recovery from BPD (45.5% vs. 23.8%), changes in dimensional indices of BPD severity (including 6 of the 9 DSM-IV BPD-criteria), changes in secondary variables, and cost-effectiveness (Giesen-Bloo et al., 2006; van Asselt et al., 2008). Moreover, medication use at baseline, as well as during treatment, predicted a smaller chance of recovery. Given the results of the RCT, we controlled for (significant) effects of condition and medication (the last for predicting recovery). Predictors that were (trend) significant when tested singly were subsequently tested in a multivariate approach, so that their unique contribution could be assessed. The second study aimed at testing whether the effect of a predictor of failure to recover was mediated by a process during treatment sessions, based on ratings of recordings of treatment sessions.

2. Study 1 method

2.1. Design and participants

This study used data from an RCT on the effectiveness of ST and TFP. In short, 86 patients with a primary diagnosis of BPD were randomized to these two treatments, and followed up during three years of treatment. Axis-1 and -2 diagnoses were based on SCID interviews (First, Spitzer, Gibbon, & Williams, 1996, 1997; Groenestijn, Akkerhuis, Kupka, Schneider, & Nolen, 1999; Weertman, Arntz, & Kerkhofs, 2000). An extensive structured interview, the Borderline Personality Disorder Severity Index version IV (BPDSI-IV; Giesen-Bloo, Wachters, Schouten, & Arntz, 2010) was used to assess severity of BPD-manifestations every 3 months. The BPDSI-IV has been recommended as a good outcome measure for BPD treatment research by an international group of experts (Zanarini et al., 2010), and has been validated in various studies (Arntz et al., 2003; Giesen-Bloo et al., 2010; Kröger et al., 2013; Leppänen, Lindeman, Arntz, & Hakko, 2013; Leppänen, Hakko, Sintonen, & Lindeman, 2015) and used in various treatment studies as outcome instrument (e.g., Bales et al., 2012; Bellino, Paradiso, & Bogetto, 2006, 2015; Dickhaut & Arntz, 2014; Leppänen et al., 2015; Nadort et al., 2009; Schuppert et al., 2012; Verheul et al., 2003). A statistically derived criterion based on the BPDSI total score (when < 15) defined recovery from BPD with specificity .97 and sensitivity 1.00 (Giesen-Bloo et al., 2006; 2010). Treatment discontinuation was also monitored during the study (Giesen-Bloo et al., 2006). Inclusion criteria were BPD as primary diagnosis, age 18–60 years, BPDSI-IV score >20, and Dutch literacy. Exclusion criteria were psychotic disorders (except psychotic episodes covered by BPD-criterion 9), bipolar disorder, dissociative identity disorder, antisocial personality disorder, attention-deficit/hyperactivity disorder, addiction needing clinical detoxification (after detox participation was allowed; addictions not needing clinical detox were not excluded), psychiatric disorders secondary to medical conditions, and mental retardation. When mental retardation was suspected, a formal IQ test was planned to be taken; however none of the candidates was suspected of mental retardation.

At baseline, the 86 participants (6 men) were on average 30.6 years (SD 7.8), had a BPDSI-IV total score of 33.97 (SD 7.97), 2.7 (SD 1.6) Axis-1 and 2.1 (SD 1.2) PD-diagnoses, and had 2.9 (SD 1.3) previous treatment modalities; 65 (75.6%) used psychotropic medication (58 (67.4%) antidepressants; 28 (32.6%) anxiolytics; 17 (19.8%) antipsychotics). Fifty-four (62.8%) didn't have a partner; 69

(80.2) didn't have a paid job. More details about the study can be found in Giesen-Bloo et al. (2006).

2.2. Predictors

All predictors were assessed before randomization by self-report, independent research assistants, or clinicians that did not take and SCIDs.

Biographical variables, treatment history, medication use, employment status, disability compensation, educational level were derived from structured interviews taken by research assistants at baseline assessment. In the Netherlands, the educational system has different levels of diplomas, which we ordered in categories ranging from primary school to university. This is also a proxy for IQ, as research indicates correlations between IQ and educational level in the .50–.85 range (Calvin, Fernandes, Smith, Visscher, & Deary, 2010; Deary, Strand, Smith, & Fernandes, 2007; Gerritsen, Berg, & Deelman, 2001; Luteijn & Barelds, 2004; Plassman et al., 1995; Tambs, Sundet, Magnus, & Berg, 1989; Wechsler, 2005).

BPD-severity was assessed with the total BPDSI-IV-score, representing objectifiable frequency and extent of BPD-manifestations during the last 3 months, and with the BPD-checklist, a self-report of the subjective burden of BPD-symptoms during the last month. The internal consistency of the BPDSI-IV total score is .96 in a heterogeneous sample, and .85 in a BPD-sample; of the BPD-checklist .93 (see Giesen-Bloo et al., 2010 for both instruments).

Level of recent suicidality was assessed by the sum of BPDSI-IV items 5.11–5.13 (plans, steps, attempts; Cronbach alpha = .63 (Giesen-Bloo et al., 2010)), and with BPD-checklist items 9 and 26 (Cronbach alpha = .84).

Level of recent self-injurious behavior was assessed with BPDSI-IV items 5.1–5.8 (Cronbach alpha = .73), and with BPD-checklist item 6.

Level of recent substance abuse was assessed with the sum of items BPDSI-IV 4.4–4.7, and with the sum of BPD-checklist items 12 (alcohol) and 17 (drugs). The internal consistency of the first scale was .35, of the second .37. The low reliabilities point at the relatively low association between drugs and alcohol abuse.

Level of recent interpersonal instability was assessed with summed BPDSI-IV criterion 1 and 2 total scores (Cronbach alpha = .85), and a similar sum for the BPD-checklist (Cronbach alpha = .91).

Level of recent impulsivity was assessed with BPDSI-IV criterion 4 total score (Cronbach alpha = .67), and similar for the BPD-checklist (Cronbach alpha = .70).

Level of recent anger regulation problems was assessed with BPDSI-IV criterion 8 total score (Cronbach alpha = .78), and similar for the BPD-checklist (Cronbach alpha = .78).

Level of recent dissociation problems was assessed with BPDSI-IV criterion 9 dissociation items sum score (with the paranoia items excluded; Cronbach alpha = .77), and similar for the BPD-checklist (Cronbach alpha = .78).

Level of defense mechanisms was assessed with the 3 subscales (Mature, Neurotic and Immature Defenses) of the Defensive Style Questionnaire (Andrews, Pollock, & Stewart, 1989). In the Giesen-Bloo et al. (2010) data internal consistencies were .52 (Mature Defenses), .71 (Neurotic Defenses), and .88 (Immature Defenses).

Level of personality organization (ego pathology) was assessed with the Inventory of Personality Organization (Lenzenweger, Clarkin, Kernberg, & Foelsch, 2001). Five subscales were assessed: Lower Level Defenses, Identity Diffusion, Pathological Object Relations, Alterations in Reality Testing, Superego Pathology. Internal consistencies range from .78 to .93 (Berghuis, Kamphuis, Boedijn, & Verheul, 2009).

Childhood abuse and neglect was assessed with the Interview for Traumatic Events in Childhood (ITEC; Lobbestael, Arntz, Harkema-

Schouten, & Bernstein, 2009). The interview yields dimensional severity scores of emotional (Cronbach $\alpha = .83$), physical ($\alpha = .88$) and sexual abuse ($\alpha = .89$), as well as emotional ($\alpha = .75$) and physical neglect ($\alpha = .58$).

Personality Disorder comorbidity was assessed with the SCID-II total of PD-diagnoses and sum scores per personality disorder (range of interrater reliability (ICC) per PD .69–.95, mean ICC = .84; Lobbestael, Leurgans, & Arntz, 2011; range of internal consistency: .55–.87, mean = .73; data from Arntz et al., 2009). For narcissism, an additional variable was explored, the entitlement score of the Young Schema Questionnaire (YSQ; Rijkeboer, van den Bergh, & van den Bout, 2005; Cronbach alpha = .76; Rijkeboer & van den Bergh, 2006).

Level of depression, anxiety and hostility were assessed with subscales of the SCL-90 (Derogatis, Lipman, & Covi, 1973). Medians of internal consistencies over various Dutch samples are .90 (depression), .88 (anxiety) and .77 (hostility) (Arrindell & Ettema, 2003).

2.3. Statistical analysis

Predictors of discontinuation were first separately tested with logistic regression, controlling for treatment condition, given that condition predicted treatment dropout, with ST having significantly less dropout than TFP (Giesen-Bloo et al., 2006). Next, predictors with a significance level $<.10$ were together entered and tested with backward logistic regression, with a two-tailed significance level of $p = .05$ as criterion. Lastly, it was tested whether the resulting predictors were moderated by condition, by adding the predictor by condition interactions to the regression equation. These moderation tests were a second reason to include treatment condition as covariate, as moderation tests should be done with main effects as covariates forced into the equation. The same approach was used for predicting recovery, but now with both condition and baseline medication use as covariates given the findings of the original RCT, where condition and medication use were significant predictors of recovery (with ST superior to TFP in recovery, and those using medication achieving less often recovery; Giesen-Bloo et al., 2006). Analyses were based on intent-to-treat principle, that is all available data was included. For the recovery analysis, the last available assessment was therefore used (in the 3 year period). Analyses were checked for robustness by redoing them with survival analysis, which yielded similar findings.

We did not correct for number of tests that were done, for instance by a Bonferroni correction, for the following reasons. First, we wanted to test whether each predictor reported previously in the literature would survive replication. Using a corrected p-level (e.g., $p < .0009$ instead of $p < .05$, when a Bonferroni correction is used for 55 predictors) would lead to an increased chance of rejecting predictors as not significant whilst they actually are valid predictors. Thus, to give the suggested predictors a reasonable chance to emerge, we used a significance level of .05 for each individual predictor. Second, only unrealistically powerful predictors would survive a corrected p-level with the current sample size, and it seemed unlikely that such powerful predictors exist, as only a few could exist and such obvious predictors would have already be detected in previous research. Thus, instead of aiming to demonstrate that (a) specific variable(s) predict treatment discontinuation or success, we aimed to contribute to building an empirical knowledge base that requires replication and accumulation of findings over studies to detect predictors.

We used a backward procedure to assess which predictors, that were (trend) significant when separately tested, would survive when controlled for each other. Stepwise procedures have been criticized for leading to overfitting of the model, including inflated

beta's, R^2 values etc. This implies that replication with the same set of predictors is necessary to come to unbiased estimates of parameters, R^2 , etc. Note however that the aim of the use of the stepwise procedure was not to create a numerical prediction model to predict chances of dropout or recovery for new patients, but to assess whether predictors that were (trend) significant when tested initially alone would survive when controlled for each other, that is to assess to what degree their contribution was unique. Given the relatively small sample size, a multivariate model with all predictors entered simultaneously was impossible, hence we first tested the predictors separately. The relatively small sample size also makes any multivariate model with relatively many predictors problematic, which calls for a selection procedure. This implies that the explanatory power of the resulting model after backwards elimination might be overestimated, and that replication in an independent dataset is essential to get unbiased estimation. Thus, the current procedures should be seen as aiming to further select candidate predictors that need further testing in the future. Note on the other hand that, in contrast to most of the previous research, we choose variables that were reported in previous research to be significant predictors, thus our study did not use a simple shotgun approach.

3. Results

3.1. Prediction of treatment discontinuation

Table 1 gives an overview of the tests of the predictors separately, controlled for condition, as well as of the final model after entering the predictors with $p < .10$ and stepwise deleting those with $p > .05$. The final model contained three predictors: condition (TFP vs ST), hostility and childhood physical abuse. Patients with higher levels of hostility (SCL-90) and with more severe childhood physical abuse (ITEC) had a higher chance to discontinue treatment prematurely. It was also explored whether conditions differed in predictive strength of hostility and childhood physical abuse, but predictor by condition interactions failed to reach significance, hostility (SCL-90) by condition: OR = 1.033, $p = .75$, 95%CI (.845; 1.263); childhood physical abuse: OR = .944, $p = .77$, 95%CI (.644; 1.383).

3.2. Prediction of recovery

Table 2 presents an overview of the tests of the predictors separately, controlled for condition and medication, as well as of the final model after entering the predictors with $p < .10$ and stepwise deleting those with $p > .05$. The tests of separate predictors showed, controlled for condition and medication, burden of total BPD-symptoms, burden of dissociative BPD-symptoms, and burden of suicidality symptoms, all assessed with the BPD-checklist, to be significant (Table 2). The backward regression procedure showed burden of dissociative symptoms to remain a significant predictor, whilst other predictors with initial $p < .10$ being removed. In sum, the final model showed that TFP (vs. ST), baseline medication use, and higher levels of subjective burden of dissociative symptoms to be predictive of less chance of recovery. The moderation tests showed that treatment condition did not significantly influence the association of subjective burden of dissociation with recovery: OR = .851, $p = .37$, 95%CI (.598; 1.211), controlled for baseline medication. Fig. 1 illustrates the relationship between level of subjective burden of dissociation (grouped in quintiles (each quintile represents 20% of sample) and chance of recovery by condition, controlled for medication use (based on results of the logistic regression). The influence of dissociation appears to be large. For example, at lowest level 73% recovers with

ST, which reduces to 32% at the highest level of dissociation; for TFP this is 47% respectively 14%.

4. Discussion

Evidence was found that treatment discontinuation was related to a history of childhood physical abuse and baseline level of hostility, but also that discontinuation was not related to numerous other variables that have been mentioned in the literature. Specifically, variables mentioned in the McMurrin et al. (2010) review as most prominent over studies, i.e. younger age, lower education, lower occupational levels, and greater avoidance were *not* found to predict discontinuation. The association of childhood physical abuse with discontinuation is at odds with a previous study that failed to detect such a relationship, but found emotional neglect to be a predictor (Perry, Bond, & Roy, 2007). However, the latter study's sample was mixed as to diagnoses and the study was conducted in a natural (uncontrolled) context. Moreover, forms of childhood abuse tend to correlate (Lobbetael et al., 2009), making the detection of essential predictors difficult. Hostility as predictor of discontinuation has previously been reported by Smith et al. (1995) for TFP and Rüscher et al. (2008) for inpatient DBT (though at trend level). The range of psychotherapy types of which hostility is predictive of discontinuation is interesting and makes it a potential candidate for a relatively universal discontinuation predictor.

As to treatment success, a quite different predictor was found: subjective burden of dissociation. This replicates a study where dissociation was found to predict worse outcome of DBT (Kleindienst et al., 2011). Note however that another (smaller) study did not find evidence for dissociation predicting less treatment response of inpatient DBT (Braakmann et al., 2007). Still another study reported that pre-treatment dissociation severity did not predict remission from self-harm (Harned, Jackson, Comtois, & Linehan, 2010). The last two studies were criticized as having low quality (Barnicott et al., 2012). The finding that baseline dissociation predicts a poorer response raised the hypothesis that perhaps patients reporting to highly suffer from dissociation, also do so during psychotherapy sessions, which might interfere with profiting from the session and thus explain the lack of success. Study 2 aimed to investigate this hypothesis.

5. Study 2

The prediction of recovery by self-reported dissociation at baseline raised the question whether perhaps this effect was caused by those reporting high levels of dissociation at baseline also suffering from dissociation during treatment sessions. If so, profiting from treatment could be compromised by in-session dissociation, dissociation limiting the capacity to process the experiences and insights from the session and/or prohibiting the therapist to use the full range of treatment techniques. This hypothesis was tested by having independent judges rate the level of dissociation manifest during an audio-recorded session and testing whether the in-session level of dissociation statistically mediated the relationship between baseline dissociation and recovery.

6. Method

6.1. Tape ratings

To investigate the possible mediation of negative effects of pretest dissociation on treatment outcome by dissociation during treatment sessions, a 15-item scale was developed to rate manifestations of dissociation on the basis of session recordings, as a

Table 1
Prediction of treatment discontinuation.^a

Predictor	β	s.e.	Wald	p-value	OR	95% Confidence interval	
						Lower	Upper
Condition (TFP vs ST)	.981	.458	4.585	.032	2.667	1.087	6.544
Biographical characteristics							
Gender (male vs female)	−1.091	1.138	.918	.338	.336	.036	3.128
Age (in years)	.014	.030	.233	.630	1.014	.957	1.076
Educational level (5 levels, low to high)	.061	.182	.115	.735	1.063	.745	1.518
Highest employment level (7 levels, low to high)	.096	.121	.638	.424	1.101	.869	1.395
Having a partner	.469	.475	.977	.323	1.599	.631	4.056
Having a paid job	.498	.565	.778	.378	1.645	.544	4.976
Receiving disability compensation	.434	.626	.481	.488	1.544	.452	5.272
Treatment history							
Medication use at baseline	.129	.534	.058	.810	1.137	.399	3.238
Having had psychological therapy	−.382	.646	.351	.554	.682	.192	2.418
Number of previous treatments	−.016	.178	.008	.929	.984	.694	1.395
History of inpatient treatment	.105	.489	.046	.830	1.111	.426	2.894
BPD features							
BPD severity (BPDSI-IV total)	−.005	.029	.035	.851	.995	.940	1.052
BPD severity (burden of BPD symptoms)	.008	.009	.844	.358	1.008	.991	1.027
Level of suicidality (BPDSI-IV)	.813	.486	2.795	.095	2.255	.869	5.848
Burden of suicidality (BPD-checklist)	−.048	.118	.169	.681	.953	.757	1.200
Level of self-injury (BPDSI-IV)	−.168	.192	.770	.380	.845	.580	1.231
Burden of self-injury (BPD-checklist)	−.215	.119	1.159	.282	.807	.546	1.192
Level of substance abuse (BPDSI-IV)	−.036	.190	.036	.850	.965	.665	1.400
Burden of substance abuse (BPD-checklist)	.077	.133	.331	.565	1.080	.831	1.403
Level of interpersonal instability (BPDSI-IV)	−.193	.181	1.137	.286	.824	.577	1.176
Burden of interpersonal instability (BPD-checklist)	.004	.027	.026	.871	1.004	.952	1.059
Level of impulsivity (BPDSI-IV)	.029	.210	.020	.889	1.030	.682	1.556
Burden of impulsivity (BPD-checklist)	.045	.049	.852	.356	1.046	.950	1.152
Level of anger problems (BPDSI-IV)	−.007	.130	.003	.957	.993	.770	1.280
Burden of anger problems (BPD-checklist)	.140	.072	3.777	.052	1.151	.999	1.326
Level of dissociation (BPDSI-IV)	.018	.021	.749	.387	1.018	.977	1.061
Burden of dissociation (BPD-checklist)	.007	.068	.011	.918	1.007	.881	1.151
Quality of current relations (WHOQOL)	−.141	.099	2.033	.154	.868	.715	1.054
Defense mechanisms & personality organization							
Mature defense mechanisms (DSQ)	.008	.021	.135	.713	1.008	.967	1.050
Neurotic defense mechanisms (DSQ)	−.004	.013	.107	.744	.966	.970	1.022
Immature defense mechanisms (DSQ)	.012	.010	1.680	.195	1.012	.994	1.032
Lower level defenses (IPO)	−.009	.022	.173	.678	.991	.950	1.034
Identity diffusion (IPO)	.006	.019	.113	.737	1.006	.970	1.045
Pathological object relations (IPO)	.000	.013	.001	.981	1.000	.976	1.026
Alterations in reality testing (IPO)	−.005	.255	.000	.986	.995	.604	1.640
Superego pathology (IPO)	.170	.307	.307	.579	1.186	.649	2.165
Axis-II comorbidity							
Number of personality disorders (SCID-II)	−.180	.200	.813	.367	.835	.564	1.236
Avoidant traits (SCID-II)	−.014	.058	.063	.803	.986	.880	1.104
Dependent traits (SCID-II)	−.031	.050	.386	.534	.969	.879	1.069
Obsessive-compulsive traits (SCID-II)	−.033	.072	.211	.646	.968	.840	1.114
Paranoid traits (SCID-II)	−.046	.066	.482	.488	.955	.840	1.087
Schizotypal traits (SCID-II)	−.010	.096	.011	.917	.990	.820	1.195
Schizoid traits (SCID-II)	−.018	.152	.014	.907	.982	.729	1.324
Histrionic traits (SCID-II)	−.017	.091	.033	.855	.984	.823	1.175
Narcissistic traits (SCID-II)	−.095	.121	.608	.435	.910	.717	1.154
Antisocial traits (SCID-II)	.119	.089	1.779	.182	1.126	.946	1.341
Entitlement (YSQ)	.005	.027	.029	.865	1.005	.952	1.060
Negative emotional problems							
Anxiety symptoms (SCL-90)	.049	.031	2.511	.113	1.050	.989	1.115
Hostility (SCL-90)	.121	.051	5.545	.019	1.128	1.020	1.247
Depressive symptoms (SCL-90)	.018	.018	1.005	.316	1.018	.983	1.054
Childhood abuse and neglect							
Physical abuse (ITEC)	.221	.097	5.161	.023	1.247	1.031	1.509
Sexual abuse (ITEC)	.054	.107	.259	.611	1.056	.857	1.301
Emotional abuse (ITEC)	.365	.216	2.851	.091	1.441	.943	2.203
Physical neglect (ITEC)	−.165	1.446	.013	.909	.848	.050	14.430
Emotional neglect (ITEC)	.192	.218	.772	.380	1.211	.790	1.858
Final multivariate model							
Condition (TFP vs ST)	1.051	.504	4.353	.037	2.860	1.066	7.673
Hostility (SCL-90)	.146	.057	6.491	.011	1.158	1.034	1.296
Childhood physical abuse (ITEC)	.261	.103	6.491	.011	1.298	1.061	1.588

Bold values denote $p < .05$.

^a All predictors following condition controlled for condition.

Table 2
Prediction of recovery.^a

Predictor	β	s.e.	Wald	p-value	OR	95% Confidence interval	
						Lower	Upper
Condition (TFP vs ST)	−1.226	.516	5.640	.018	.294	.107	.807
Medication (use at baseline)	−1.499	.565	7.042	.008	.223	.074	.676
Biographical characteristics							
Gender (male vs female)	−.429	.964	.198	.656	.651	.098	4.304
Age (in years)	.003	.031	.012	.914	1.003	.944	1.067
Educational level (5 levels, low to high)	−.047	.193	.058	.809	.955	.654	1.392
Highest employment level (7 levels, low to high)	−.001	.130	.000	.995	.999	.775	1.289
Having a partner	.005	.508	.000	.992	1.005	.372	2.719
Having a paid job	−.595	.639	.867	.352	.552	.158	1.931
Receiving disability compensation	.790	.664	1.414	.234	2.203	.599	8.101
Treatment history							
Having had psychological therapy	−.001	.721	.000	.999	.999	.243	4.105
Number of previous treatments	−.326	.213	2.331	.127	.722	.475	1.097
History of inpatient treatment	−.091	.516	.031	.860	.913	.332	2.510
BPD features							
BPD severity (BPDSI-IV total)	−.014	.032	.200	.655	.986	.926	1.049
BPD severity (burden of BPD symptoms)	−.026	.011	5.758	.016	.974	.953	.995
Level of suicidality (BPDSI-IV)	−.140	.497	.080	.778	.869	.328	2.301
Burden of suicidality (BPD-checklist)	−.339	.167	4.130	.042	.712	.514	.988
Level of self-injury (BPDSI-IV)	−.068	.212	.103	.749	.934	.616	1.416
Burden of self-injury (BPD-checklist)	−.373	.239	2.440	.118	.689	.431	1.100
Level of substance abuse (BPDSI-IV)	.167	.202	.690	.406	1.182	.796	1.755
Burden of substance abuse (BPD-checklist)	.004	.140	.001	.977	1.004	.763	1.320
Level of interpersonal instability (BPDSI-IV)	.036	.187	.037	.848	1.037	.718	1.496
Burden of interpersonal instability (BPD-checklist)	−.030	.029	1.053	.305	.970	.916	1.028
Level of impulsivity (BPDSI-IV)	.025	.226	.013	.910	1.026	.659	1.597
Burden of impulsivity (BPD-checklist)	−.098	.059	2.778	.096	.906	.807	1.071
Level of anger problems (BPDSI-IV)	.010	.138	.005	.942	1.010	.771	1.323
Burden of anger problems (BPD-checklist)	−.162	.083	3.786	.052	.850	.722	1.001
Level of dissociation (BPDSI-IV)	−.030	.024	1.520	.218	.971	.926	1.018
Burden of dissociation (BPD-checklist)	−.230	.088	6.885	.009	.794	.669	.943
Quality of current relations (WHOQOL)	−.025	.101	.062	.803	.975	.800	1.188
Defense mechanisms & personality organization							
Mature defense mechanisms (DSQ)	.018	.023	.624	.430	1.018	.974	1.064
Neurotic defense mechanisms (DSQ)	−.013	.015	.833	.361	.987	.959	1.015
Immature defense mechanisms (DSQ)	−.005	.010	.237	.626	.995	.975	1.015
Lower level defenses (IPO)	−.015	.023	.455	.500	.985	.942	1.030
Identity diffusion (IPO)	−.010	.020	.235	.625	.990	.952	1.030
Pathological object relations (IPO)	−.013	.014	.863	.353	.988	.962	1.014
Alterations in reality testing (IPO)	−.094	.283	.110	.740	.910	.523	1.585
Superego pathology (IPO)	−.195	.329	.349	.555	.823	.432	1.570
Axis-II comorbidity							
Number of personality disorders (SCID-II)	−.262	.223	1.377	.241	.770	.497	1.192
Avoidant traits (SCID-II)	−.058	.063	.832	.362	.944	.834	1.068
Dependent traits (SCID-II)	−.130	.072	3.256	.071	.878	.763	1.011
Obsessive-compulsive traits (SCID-II)	−.071	.076	.867	.352	.931	.802	1.082
Paranoid traits (SCID-II)	−.007	.069	.011	.917	.993	.867	1.137
Schizotypal traits (SCID-II)	.002	.102	.000	.988	1.002	.819	1.224
Schizoid traits (SCID-II)	−.312	.224	1.949	.163	.732	.472	1.134
Histrionic traits (SCID-II)	.073	.094	.610	.435	1.076	.895	1.293
Narcissistic traits (SCID-II)	−.086	.121	.511	.475	.917	.724	1.162
Antisocial traits (SCID-II)	.020	.092	.046	.831	1.020	.852	1.220
Entitlement (YSQ)	−.020	.031	.408	.523	.981	.924	1.041
Negative emotional problems							
Anxiety symptoms (SCL-90)	−.050	.033	2.359	.125	.951	.892	1.014
Hostility (SCL-90)	−.113	.059	3.742	.053	.893	.796	1.002
Depressive symptoms (SCL-90)	−.003	.019	.022	.882	.997	.960	1.035
Childhood abuse and neglect							
Physical abuse (ITEC)	−.023	.094	.061	.805	.977	.813	1.174
Sexual abuse (ITEC)	.096	.115	.696	.404	1.101	.878	1.380
Emotional abuse (ITEC)	.166	.218	.578	.447	1.180	.770	1.811
Physical neglect (ITEC)	1.894	1.538	1.516	.218	6.648	.326	135.532
Emotional neglect (ITEC)	.109	.230	.226	.634	1.115	.711	1.750
Final multivariate model							
Condition (TFP vs ST)	−1.241	.537	5.335	.021	.289	.101	.829
Medication use	−1.259	.589	4.564	.033	.284	.089	.901
Burden of dissociation (BPD-checklist)	−.230	.088	6.885	.009	.794	.669	.943

Bold values denote $p < .05$.^a All predictors following condition and medication controlled for these two variables.

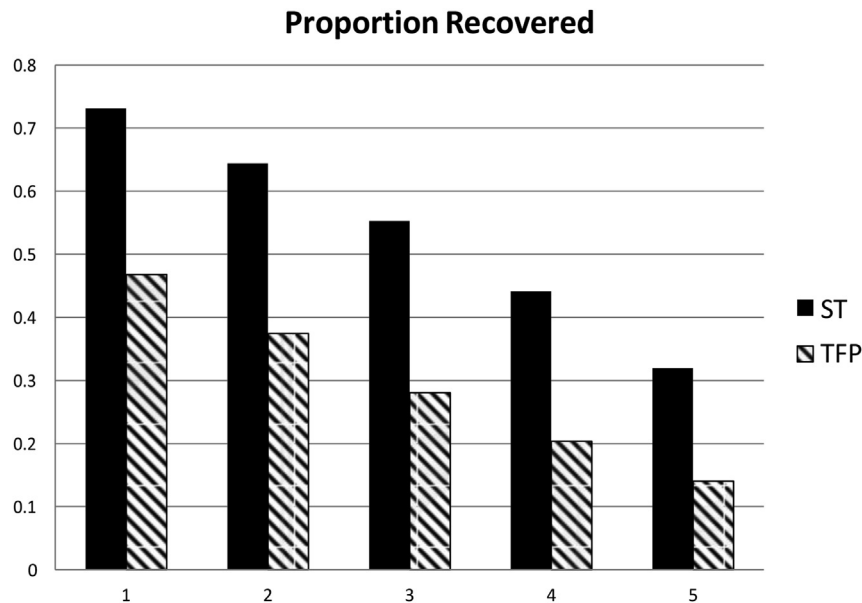


Fig. 1. Relationship between therapy success (recovery) and level of burden of dissociative symptoms at baseline (grouped by quintiles) by treatment condition, estimates from the logistic regression, controlled for baseline medication use.

literature search did not yield a suitable observation rating scale of dissociation. Items were derived from various sources, including textbooks (e.g., Spiegel, 1994) and reviews (e.g., Korzekwa, Dell, Links, Thabane, & Fougere, 2009), self-report scales (Stiglmayr, Shapiro, Stieglitz, Limberger, & Bohus, 2001; Vanderlinden, Van Dyck, Vandereycken, Vertommen, & Verkes, 1993), and interviews (Giesen-Bloo et al., 2010; Steinberg, 1993). Dissociative areas captured by items included depersonalization, derealization, amnesia, and disproportional anger, fear and distrust of the therapist (see Appendix for the scale). The disproportional responses towards the therapist items were added as expressions of more severe forms of dissociation where these phenomena result from a shift to a dissociated emotional state (usually viewed as related to traumatic experiences) that drives them and makes them disproportional. Three independent raters rated a random selection of tapes of months 2–12 of therapy (one tape per participant), per rater 33–36 tapes were double rated. Four items (3,6,8,11) showed very little variance, probably because raters were unable to infer these rather internal experiences from the audiotapes, and were therefore disregarded for further analysis. The interrater agreement per item was high, mean ICC = .82, median ICC = .85. Item ratings were next averaged over raters. A principal component analysis supported a one factor solution, with mean factor loading .60, range .45–.73. For further computations the factor scores were used.

Recordings were available for 80 of the 86 participants.

6.2. Mediation test

Mediation was tested using the Preacher and Hayes approach (2008a). Fig. 2 illustrates the approach. First, the association between baseline dissociation and in-session dissociation was tested with linear regression. Second, the association between in-session dissociation and recovery (controlled for condition and medication) was assessed with logistic regression. Third, the change of the association between baseline dissociation and recovery, controlling for condition and medication, with adding in-session dissociation was assessed with logistic regression. Lastly, the significance of the indirect path between baseline dissociation and recovery via in-session dissociation (the mediator), controlling for condition and medication, was assessed with an SPSS macro by Preacher and

Hayes (2008a, b), which uses a bootstrap percentile approach to estimate the 95% CI of the indirect path and is suitable for logistic regression with multiple covariates. Significance is concluded if the 95% CI does not contain zero. To achieve high precision of the estimation, the number of bootstrap samples was set at 50,000 (Koehele, Brown, & Haneuse, 2009). Mediation is concluded when first, second and last tests are significant, and there is a reduction of the predictive power of the original predictor in the third test. As tapes were unavailable for 6 participants, analyses were repeated with estimated in-session ratings for these missings: (1) by assigning the mean of the in-session dissociation factor scores to the missings; and (2) by assigning a regression-based score to the missings, with baseline dissociation, condition and medication as predictors.

7. Results

Table 3 presents the results of the mediation analyses. In all analyses the relationship between baseline dissociation and in-session dissociation was significant, as was the relationship between in-session dissociation and recovery. The association-strength between baseline dissociation and recovery reduced after adding in-session dissociation as predictor, to non-significant levels in the two analyses of the full sample. The lowest row of the table shows that the mediation path (the contribution of baseline dissociation to recovery via in-session dissociation) was significant in the full sample analyses, as the 95% CIs did not contain zero. For the reduced sample ($N = 80$) the upper limit of the 95% CI was zero.

8. Discussion

The findings, despite rating only a single session, at least partially supported the hypothesis that in-session dissociation might account for the effect of baseline dissociation on treatment success. For the reduced sample ($N = 80$) the results of the mediation analysis were a bit ambiguous, as the 95%CI of the mediation path has zero as upper limit, instead of being smaller than zero. The effect of baseline dissociation reduced but not to a nonsignificant level after adding the mediator to the model, suggesting what is called “partial mediation”. The ambiguous findings seem related to

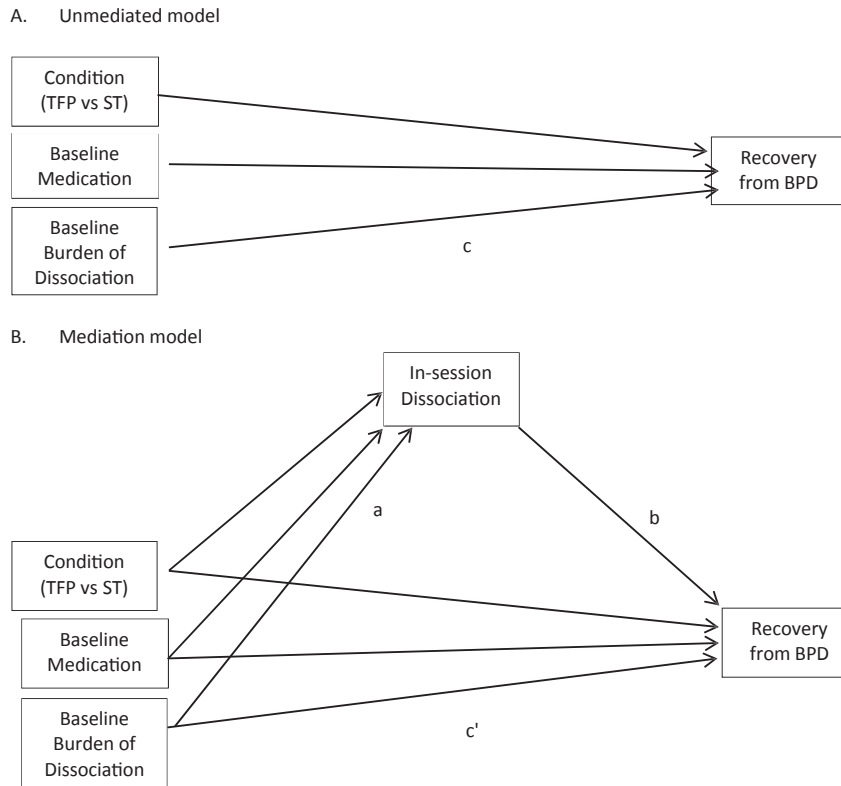


Fig. 2. Illustration of the mediation analysis of the effect of baseline burden of dissociation on recovery by in-session dissociation. Mediation is concluded when the direct effect *c* (in upper model A) reduces when controlled for the mediator (*c'* (lower model B) should reduce compared to *c*), and effects *a* and *b* are significant, and the indirect effect (the mediation path) *ab* is significant. When *c'* becomes nonsignificant, “full mediation” is concluded. The mediation was tested controlling for condition and baseline medication as covariates.

Table 3
Results of mediation tests of the dissociation effect.

	N = 80			N = 86, missing replacement by mean			N = 86, missing replacement by regression estimate		
	β	s.e.	p	β	s.e.	p	β	s.e.	p
Baseline dissociation → in-session dissociation	.086	.031	.008	.085	.030	.005	.083	.030	.007
In-session dissociation → recovery ^a	-.946	.452	.036	-.993	.473	.036	-1.061	.486	.029
Baseline dissociation → recovery ^a	-.241	.089	.007	-.230	.088	.009	-.230	.088	.009
Baseline dissociation → recovery, controlled for in-session dissociation ^a	-.189	.093	.043	-.175	.092	.057	-.170	.092	.064
	95% CI			95% CI			95% CI		
Mediation path ^a	Lower	Upper		Lower	Upper		Lower	Upper	
	-.2091	.0000		-.2230	-.0002		-.2377	-.0023	

^a (also) controlled for condition and medication.

the fact that we did not have recordings of six participants, as with both models using estimated mediation scores for these six participants, unequivocal evidence for full mediation was obtained.

9. General discussion

The first study reported in this manuscript aimed to test predictors of dropout and treatment success mentioned in previous studies on (B)PD in data from an RCT on ST and TFP as treatments for BPD. We found evidence for hostility and childhood physical abuse as predictors of treatment discontinuation, and for subjective burden of dissociation as predictor of failure to recover from BPD. The second study explored whether in-session dissociation explains the predictive power of burden of baseline dissociation. Mediation tests confirmed at least partially this hypothesis, as in the full sample (where in-session mediation had to be estimated because of absence of audio-recordings in 6 participants) mediation was significant and the direct effect of baseline dissociation on

recovery became nonsignificant when controlled for mediation.

However, in the subsample with recordings available, the mediation 95% CI had exactly zero as upper limit and the direct path remained significant, though reduced in strength, thus yielding mixed evidence.

Hostility and childhood physical abuse were predictive of treatment discontinuation, irrespective of treatment condition. Previous studies found evidence for hostility as predicting discontinuation (Rüsch et al., 2008; Smith et al., 1995) and we now have evidence that it predicts discontinuation of DBT, ST and TFP. Patients struggling with high levels of hostility might find it difficult to go through the frustrating process of treatment, demand a quick fix, and might easily feel hostile towards their therapist. Therapists on the other hand often find it difficult to deal with hostile patients, and this might weaken the collaboration in treatment. Studies investigating such processes are needed to better understand why hostility increases chance of discontinuation. A cautious conclusion that can be drawn from our finding is that therapists should

develop ways to better tolerate and deal with hostility, to prevent premature discontinuation of treatment. Childhood abuse has been reported previously as relevant predictor, but this was not always specifically physical abuse (Gunderson et al., 2006; Perry et al., 2007).

As different forms of childhood abuse tend to correlate (e.g., physical abuse usually takes place in a context of emotional abuse and neglect, Lobbestael et al., 2009), we should be cautious to conclude it is specifically physical abuse that predicts discontinuation. Indeed, emotional abuse showed a trend significant relationship with discontinuation. However, physical abuse might be an index of very severe abuse that damages trust in others, including therapists. Clearly, future studies should try to replicate whether childhood physical abuse is a specific predictor of treatment discontinuation.

As hostility might also be related to a trauma history, the finding that both hostility and childhood physical abuse predict treatment discontinuation suggests that it is important to effectively process trauma with empirically supported methods integrated in the offered treatment. This is actually part of the ST protocol, and recent studies have demonstrated that trauma processing can also be effectively integrated in DBT (Bohus et al., 2013; Harned, Korslund, & Linehan, 2014). The inclusion of (childhood) trauma processing in ST might be one of the explanations why dropout was less and recovery higher in ST than in TFP.

The finding that baseline burden of dissociation was predictive of a diminished chance to recover was not only a replication of a similar finding by Kleindienst et al. (2011), but was also further explained by in-session dissociation accounting for this. One explanation is that dissociation during therapy sessions reduces the impact of corrective experiences and information, because dissociation interferes with information processing. It should be noted that there are indications that the interfering effect of dissociation on information processing specifically concerns highly negative and traumatic material (Olsen & Beck, 2012), which are of course topics of psychotherapy. Another explanation is that therapists when confronted with dissociation in their patients find it more difficult to use the full range of techniques. The effects of dissociation were quite powerful with only a few of the highest dissociative patients recovering (see Fig. 1, highest 20%). This calls for revising treatments to better deal with in-session dissociation. Interestingly, one of the explanations Braakman et al. (2007) offered why in their study dissociation did not predict worse outcome was their effectively addressing dissociation during treatment. Interestingly, the last decade ST-therapists reported having become less afraid of dissociation during sessions, using methods to get patients out of dissociative states, and teaching patients to prevent a massive shift into a dissociative state. Future studies will show us whether indeed the field has improved in better handling dissociation during treatment so that treatment effects are less affected.

The findings from the original RCT that ST had less dropout and more recovery than TFP, and medication had a negative influence on recovery, were maintained. Thus these effects turned out to be robust for the detected predictors. Moreover, we failed to find interactions indicating that predictors had a different effect in the two treatments. It is not clear why medication use predicts poorer effects. However, it should be noted that medication use during treatment was reduced, notably in antidepressants (from $N = 58$ (67.4%) to $N = 36$ (41.9%); Giesen-Bloo et al., 2006). Nevertheless, the negative association between medication and recovery persisted when the use of psychotropic medication was analyzed as a time-dependent covariate in a survival analysis of recovery (13 assessments; Wald statistic = 6.21; $P = .01$; RR = .38; 95% CI, .18–.81; Giesen-Bloo et al., 2006). Another study of TFP found similar effects (Doering et al., 2010), though a study on ST failed to replicate the

medication effect (Nadort et al., 2009). Moreover, medication effects were not found in other treatment types (Bateman & Fonagy, 1999; Black et al., 2009; Ryle & Golyunkina, 2000), though samples sizes were considerably smaller in first and last studies, and Black et al. investigated a short-term program to be added to regular treatment. However, if the medication effect exists, how to explain it? Medication use might be a marker of the more difficult patient, although baseline severity did not differ between medication users and non-users. Another possibility is that medication interferes with the psychological change processes used in ST and TFP. For example, medications like antidepressants and antipsychotics might flatten emotional arousal (Moncrieff & Cohen, 2009; Moncrieff, Cohen, & Mason, 2009; Price, Cole, & Goodwin, 2009), possibly interfering with the levels of emotional arousal necessary for psychological treatments like ST and TFP (Price & Goodwin, 2009), e.g. by dampening brain areas involved in (corrective) emotional learning like the amygdala (Harmer, 2008), and/or dampening cognitive biases and other processes that should be corrected through experiences in treatment (Merens, Van der Does, & Spinhoven, 2007; Pringle, Browning, Cowen, & Harmer, 2011). Or, patients might attribute symptom reduction to medication, instead of to (stable) psychological changes and their own effort in therapy, a kind of attribution that is known to predict relapse in the long-term in the treatment of anxiety disorders and depression (Basoglu, Marks, Kilic, Brewin, & Swinson, 1994; Moradveisi, Huibers, & Arntz, 2015; Powers, Smits, Whitley, Bystritsky, & Telch, 2008). Clearly, an RCT comparing specialized psychotherapy with vs. without medication is needed to test causality implied in the hypothesis that medication interferes with psychotherapy.

At least as important are alleged predictors that turned out to have no significant predictive power. Sociodemographic variables like age, gender, education level, unemployment, and disability compensation were not significant. Moreover, indices of severity of BPD including level of self-injury and suicidality, did not predict dropout and recovery. These findings are in line with reviews pointing out that most assumed predictors have in fact no predictive power (Barnicot et al. 2012; McMurrin et al., 2010). This indicates that we should be reluctant in using such variables in treatment allocation; more specifically we should not withhold specialist treatment to patients with a long treatment history, specific age, lower educational level, or poor social functioning, etc. Moreover, indices of “low-level” borderline organization (impulsivity problems, substance abuse, self-injury, anger control problems) did not predict discontinuation or recovery, suggesting that “low-level borderline” should not be used as exclusion criterion.

Some of the predictors were based on the baseline BPDSI-IV, including the BPDSI-IV total score as one of the indices of baseline BPD-severity. The same instrument was used to define recovery. Usually one sees a positive correlation between baseline and posttest on the same instrument, which might be partially caused by the fact that the very same instrument is used. Interestingly, no evidence for a correlation between predictors based on the baseline BPDSI-IV and recovery was found, which indicates that treatment and/or time caused fundamental changes so that the correlation between baseline BPDSI-IV and recovery was nil (see also Giesen-Bloo et al., 2006). A clinical implication is that recovery by specialized psychotherapies like ST and TFP is not predicted by baseline BPD-severity, indicating that this should not be used to select patients.

Several limitations should be mentioned. First, although we were able to investigate many of the alleged predictors mentioned in the literature, there was a substantial set we couldn't assess due to lack of instruments assessing the pertinent construct. These included a wide range of variables, e.g. juvenile conviction, parent divorce before the age of 10, trait anxiety, distractibility, and pre-contemplation stage of change. Second, the large number of tests

and the use of uncorrected p-levels increase the chance of accidental findings. However, it is important that we tested whether previously documented predictors survived replication. Instead of trying to prove indisputable predictors in a single study, it seems better to work on establishing an empirical data base over studies that can be used in meta-analyses to document the power of various predictors of treatment discontinuation and success. Third, although we explored whether the strength of the predictors found to be significant in the whole sample differed between treatments, our sample size was insufficient to detect more subtle differences (medium effect size and smaller). Such differences are of obvious importance as they inform us about the best treatment for a specific patient. Fourth, although with estimations of the six missing recordings evidence for full mediation was obtained, this does not prove that in-session dissociation causes the effect: in-session dissociation might also be a proxy of another process (for example, continuous dissociation, poor sleep quality, etc.) that accounts for the relationship. In other words, a third variable explaining the association is still possible. Nevertheless, the mediation finding indicates that it is worthwhile to test whether a revised treatment that reduces in-session dissociation would lead to better recovery in patients who suffer from high levels of dissociation. Fifth, due to financial limitations, only one recording per patient was rated, limiting the representativeness of the ratings per individual. On the other hand, with only one recording we found evidence that in-session dissociation mediated the predictive relationship of baseline dissociation and recovery, which pleads for the power of the effect. However, an additional limitation of having only one session ratings of dissociation is that the relationship between individual level of dissociation (and its change during treatment), and treatment response, could not be studied on an individual level. For that repeated assessments of level of dissociation and treatment response are needed, and advanced multilevel tests distinguishing intra- and inter-individual sources of (co-)variance. Sixth, in the absence of validated dissociation observation rating scales that could be used to rate audio recordings, we had to develop one, of which the validity is (except for the present findings) unknown. Seventh, it is unclear why subjective burden of dissociation (BPD-checklist) and not frequency of dissociation (BPDSI-IV) predicted recovery. Similarly, it is not clear why hostility assessed with the SCL-90 predicted discontinuation, and not seemingly related variables like anger control problems; and why childhood physical abuse predicted discontinuation, and not other types of abuse and neglect. Future studies are needed to shed light on these issues. Eighth, as the RCT was not planned as a prediction study some of the predictors had to be constructed from the available assessment instruments, while for some perhaps more reliable or valid instruments are available. Some of the null findings might be related to this, though reliabilities of many predictors were reasonable. Ninth, stepwise regression methods can lead to overfitting of the model, with inflated beta's, R^2 values etc. Replication with the same set of predictors is necessary to come to unbiased estimation. Note however that the aim of the stepwise procedure was not to create a numerical prediction model to predict chances of dropout or recovery for new patients, but to assess whether predictors that were (trend) significant when tested initially alone would survive multivariate tests. Interestingly the resulting models were not or only in a very limited way different from the models of the single predictors: For recovery, the model with burden of dissociative

symptoms, condition and medication resulted, which was the very same as the initial test of burden of dissociative symptoms – hence no inflation took place. For dropout, the two single significant predictors were selected (hostility and childhood physical abuse), explaining each by and large the same as they did when tested singly – although a bit of inflation took probably place, it was marginal compared to the initial tests and did not lead to different conclusions. Taken together, replication in an independent dataset remains essential, and the current procedures should be seen as aiming to select candidate predictors that need further testing in the future.

In conclusion, this study found that most variables that were previously reported to predict treatment discontinuation or effectiveness did not survive replication. However, we replicated hostility and childhood abuse (though only physical) as predictors of discontinuation, and dissociation as predictor of reduced treatment success. Moreover we found evidence that in-session dissociation explains the effect of baseline dissociation on recovery, suggesting that we need to better deal with in-session dissociation to improve effects of treatment. Similarly, therapists might need to learn to better deal with high levels of hostility in their patients, and with the aftermaths of severe childhood (physical) abuse to prevent premature discontinuation.

Declarations

Financial support

The RCT from which the data are derived that form the basis of this manuscript was funded by grant OG-97.002 from the Dutch Health Care Insurance Board. The Dutch National Fund of Mental Health supported central training of the therapists. The grant organizations played no role in the data collection and analysis, manuscript preparation, or authorization for publication.

Conflict of interest

None.

Ethics

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The medical ethics committees of the participating centers approved the study (see [Giesen-Bloo et al., 2006](#)).

Acknowledgments

Thanks are due to Eva van den Hurk, Lauren van Litsenburg and Rinske Zopfi for rating the sessions tapes. We acknowledge the contributions of the participating patients, therapists and consultants, the screening psychologists, and the research assistants.

Appendix. Dissociation rating scale

Indicate to which degree the phenomenon is shown by the patient during the session (not caused by medical condition or substance abuse).

1	2	3	4	5	6	7
not, or very little		somewhat, sometimes		considerably, regularly		extensively, often

1. **Depersonalisation:** Not feeling oneself anymore, like an outside observer of oneself, or experiencing oneself as in a movie or dream.
2. **Depersonalisation:** Feeling of alienation or unreality that is usually frightening and is related to one's own thoughts, emotions, or body, during which reality testing is intact.
3. **Depersonalisation:** The body feels different than normal, sometimes one feels nothing, or it seems as if body parts are not one's own.
4. **Derealisation:** Observing or experiencing the world around oneself totally differently so that it seems very odd or unreal (e.g. others look unfamiliar or like 'robots')
5. **Derealisation:** Feeling of alienation or unreality that is usually frightening and relates to the environment, during which reality testing is intact.
6. **Derealisation:** Hearing or seeing things not good or they sound or look different than normal.
7. **Dissociative amnesia:** Not being able to remember important things.
8. **Dissociative amnesia:** Inability to recall important personal information that is usually of a traumatic or stressful nature.
9. **Dissociative amnesia:** Not knowing anymore what one has done or where one is.
10. **Disproportional or even inexplicable anger toward the therapist:** Angry behavior toward the therapist, consisting of raised voice, changes of timbre (e.g., higher harmonic, even with the same undertone), and message content. **Note:** Exercises excluded.
11. **Disproportional or even inexplicable anger toward the therapist:** Aggressive behavior such as yelling, scolding, threatening, etc. or physically aggressive behavior. **Note:** Exercises excluded.
12. **Disproportional mistrust of the therapist:** Constantly questioning the good intentions of the therapist. Asking controlling questions.
13. **Disproportional fear of the therapist:** Constantly asking for confirmation to get constancy/certainty.
14. **Disproportional mistrust of the therapist:** Lying or concealing things to the therapist out of disproportionate mistrust.
15. **Disproportional fear of the therapist:** Telling little or nothing, or telling something very superficial. Remaining very correct. Constantly keeping an eye on what the therapist does.

References

- Andrews, G., Pollock, C., & Stewart, G. (1989). The determination of defense style by questionnaire. *Archives of General Psychiatry*, 46, 455–460.
- Arntz, A. (1999). Do personality disorders exist? On the validity of the concept and its cognitive-behavioral formulation and treatment. *Behaviour Research and Therapy*, 37, 97–134.
- Arntz, A., Bernstein, D., Gielen, D., Nieuwenhuyzen van, M., Penders, K., Haslam, N., et al. (2009). Taxometric evidence for the dimensional structure of cluster-c, paranoid, and borderline personality disorders. *Journal of Personality Disorders*, 23(6), 606–628.
- Arntz, A., van den Hoorn, M., Cornelis, J., Verheul, R., van den Bosch, W. M., & de Bie, A. J. (2003). Reliability and validity of the borderline personality disorder severity index. *Journal of Personality Disorders*, 17(1), 45–59.
- Arrindell, W. A., & Ettema, J. H. (2003). *SCL-90 Symptom checklist: Handleiding bij een multidimensionele psychopathologie-indicator*. Lisse: Swets Test Publishers [SCL-90 Symptom checklist: manual of a multidimensional psychopathology indicator].
- van Asselt, A. D. I., Dirksen, C. D., Arntz, A., Giesen-Bloo, J. H., Dyck van, R., Spinhoven, P., Tilburg van, W., Kremers, I. P., Nadort, M., & Severens, J. L. (2008). Outpatient psychotherapy for borderline Personality disorder. Cost-effectiveness of schema-focused therapy versus transference-focused psychotherapy. *British Journal of Psychiatry*, 192(6), 450–457.
- Bales, D., van Beek, N., Smits, M., Willemsen, S., Busschbach, J. J., Verheul, R., et al. (2012). Treatment outcome of 18-month, day hospital mentalization-based treatment (MBT) in patients with severe borderline personality disorder in the Netherlands. *Journal of Personality Disorders*, 26(4), 568–582.
- Barnicot, K., Katsakou, C., Bhatti, N., Savill, M., Fearn, N., & Priebe, S. (2012). Factors predicting the outcome of psychotherapy for borderline personality disorder: a systematic review. *Clinical Psychology Review*, 32, 400–412.
- Basoglu, M., Marks, I. M., Kilic, C., Brewin, C. R., & Swinson, R. P. (1994). Alprazolam and exposure for panic disorder with agoraphobia: attribution of improvement to medication predicts subsequent relapse. *British Journal of Psychiatry*, 164, 652–659.
- Bateman, A., & Fonagy, P. (1999). The effectiveness of partial hospitalisation in the treatment of borderline personality disorder: a randomised controlled trial. *American Journal of Psychiatry*, 156, 1563–1569.
- Bellino, S., Bozzatello, P., & Bogetto, F. (2015). Combined treatment of borderline personality disorder with interpersonal psychotherapy and pharmacotherapy: predictors of response. *Psychiatry Research*, 226(1), 284–288.
- Bellino, S., Paradiso, E., & Bogetto, F. (2006). Efficacy and tolerability of quetiapine in the treatment of borderline personality disorder: a pilot study. *Journal of Clinical Psychiatry*, 67(7), 1042–1046.
- Berghuis, H., Kamphuis, J. H., Boedijn, G., & Verheul, R. (2009). Psychometric properties and validity of the Dutch inventory of personality organization (IPO-NL). *Bulletin of the Menninger Clinic*, 73(1), 44–60.
- Black, D. W., Allen, J., St John, D., Pfohl, B., McCormick, B., & Blum, N. (2009). Predictors of response to systems training for emotional predictability and problem solving (STEPPS) for borderline personality disorder: an exploratory model. *Acta Psychiatrica Scandinavica*, 120, 53–61.
- Bohus, M., Dyer, A. S., Priebe, K., Krüger, A., Kleindienst, N., Schmahl, C., et al. (2013). Dialectical behaviour therapy for post-traumatic stress disorder after childhood sexual abuse in patients with and without borderline personality disorder: a randomised controlled trial. *Psychotherapy and Psychosomatics*, 82(4), 221–233.
- Braakmann, D., Ludewig, S., Milde, J., Stiglmeier, C. E., Hüppe, M., Sipos, V., et al. (2007). Dissoziative symptome im Verlauf der Behandlung der Borderline-Persönlichkeitsstörung [Dissociative symptoms during treatment of borderline personality disorder]. *Psychotherapie Psychosomatische Medizinische Psychologie*, 57, 154–160.
- Calvin, C. M., Fernandes, C., Smith, P., Visscher, P. M., & Deary, I. J. (2010). Sex, intelligence and educational achievement in a national cohort of over 175,000 11-year-old schoolchildren in England. *Intelligence*, 38, 424–432.
- Chiesa, M., & Fonagy, P. (2007). Prediction of medium-term outcome in cluster B personality disorder following residential and outpatient psychosocial treatment. *Psychotherapy and Psychosomatics*, 76, 347–353.
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13–21.
- Derogatis, L. R., Lipman, R. S., & Covi, L. (1973). SCL-90: an outpatient psychiatric rating scale: preliminary report. *Psychopharmacological Bulletin*, 9, 13–28.
- Dickhaut, V., & Arntz, A. (2014). Combined group and individual schema therapy for borderline personality disorder: a pilot study. *Journal of Behavior Therapy and Experimental Psychiatry*, 45, 242–251.
- Doering, S., Horz, S., Rentrop, M., Fischer-Kern, M., Schuster, P., Benecke, C., et al. (2010). Transference-focused psychotherapy versus treatment by community psychotherapists for borderline personality disorder: randomised controlled trial. *British Journal of Psychiatry*, 196, 389–395.
- 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.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured clinical interview for DSM-IV axis I disorders—patient edition (SCID-I/P), version 2.0*. New York: Biometrics Research Department, New York State Psychiatric Institute.
- Gerritsen, M., Berg, I., & Deelman, B. (2001). De Sniijders-Oomen niet-verbale intelligentietest: bruikbaar bij ouderen? [The Sniijders-Oomen nonverbal intelligence test: feasible for elderly?] *Tijdschrift voor Gerontologie en Geriatrie*, 32, 24–28.
- 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.
- Giesen-Bloo, J. H., Wachters, L. M., Schouten, E., & Arntz, A. (2010). The borderline personality disorder severity index-IV: psychometric evaluation and dimensional structure. *Personality and Individual Differences*, 49, 136–141.
- Groenestijn, M. A. C., Akkerhuis, G. W., Kupka, R. W., Schneider, N., & Nolen, W. A. (1999). *Structured clinical interview for DSM-IV Axis I disorders*. Lisse, the Netherlands: Swets & Zeitlinger BV.
- Gunderson, J. G., Daversa, M. T., Grilo, C. M., McGlashan, T. H., Zanarini, M. C., Shea, M. T., ... Stout, R. L. (2006). Predictors of 2-year outcome for patients with borderline personality disorder. *American Journal of Psychiatry*, 163, 822–826.
- Harmer, C. J. (2008). Serotonin and emotional processing: does it help explain antidepressant drug action? *Neuropharmacology*, 55(6), 1023–1028.
- Harned, M. S., Jackson, S. C., Comtois, K. A., & Linehan, M. M. (2010). Dialectical behavior therapy as a precursor to PTSD treatment for suicidal and/or self-injuring women with borderline personality disorder. *Journal of Traumatic Stress*, 23(4), 421–429.
- Harned, M. S., Korslund, K. E., & Linehan, M. M. (2014). A pilot randomized controlled trial of dialectical behavior therapy with and without the dialectical behavior therapy prolonged exposure protocol for suicidal and self-injuring women with borderline personality disorder and PTSD. *Behaviour Research*

- and Therapy, 55, 7–17.
- Kleindienst, N., Limberger, M. F., Ebner-Priemer, U. W., Keibel-Mauchnik, J., Dyer, A., Berger, M., ... Bohus, M. (2011). Dissociation predicts poor response to dialectical behavioral therapy in female patients with borderline personality disorder. *Journal of Personality Disorders, 25*(4), 432–447.
- Koehle, E., Brown, E., & Haneuse, S. J. P. A. (2009). On the assessment of Monte Carlo error in simulation-based statistical analyses. *The American Statistician, 63*, 155–162.
- Korzekwa, M. I., Dell, P. F., Links, P. S., Thabane, L., & Fougere, P. (2009). Dissociation in borderline personality disorder: a detailed look. *Journal of Trauma & Dissociation, 10*(3), 346–367.
- Kröger, C., Vonau, M., Kliem, S., Röpke, S., Kosfelder, J., & Arntz, A. (2013). Psychometric properties of the German version of the borderline personality disorder severity INDEX – version IV. *Psychopathology, 46*, 396–403.
- Lenzenweger, M. F., Clarkin, J. F., Kernberg, O. F., & Foelsch, P. A. (2001). The inventory of personality organization: psychometric properties, factorial composition, and criterion relations with affect, aggressive dyscontrol, psychosis proneness, and selfdomains in a non-clinical sample. *Psychological Assessment, 13*, 577–591.
- Leppänen, V., Hakko, H., Sintonen, H., & Lindeman, S. (2015). Comparing effectiveness of treatments for borderline personality disorder in communal mental health care: the Oulu BPD study. *Community Mental Health Journal, 1*–12.
- Leppänen, V., Lindeman, S., Arntz, A., & Hakko, H. (2013). Preliminary evaluation of psychometric properties of the Finnish borderline personality disorder severity index: Oulu-BPD-study. *Nordic Journal of Psychiatry, 67*, 312–319.
- Links, P. S., Mitton, J. E., & Steiner, M. (1990). Predicting outcome for borderline personality disorder. *Comprehensive Psychiatry, 31*(6), 490–498.
- Lobbetael, J., Arntz, A., Harkema-Schouten, P., & Bernstein, D. (2009). Development and psychometric evaluation of a new assessment method for childhood maltreatment experiences: the interview for traumatic events in childhood. *Child Abuse and Neglect, 33*, 505–517.
- Lobbetael, J., Leurgans, M., & Arntz, A. (2011). Inter-rater reliability of the structural and clinical interview for DSM-IV axis (SCID I) and axis 2 disorders (SCID II). *Clinical Psychology and Psychotherapy, 18*, 75–79.
- Löffler-Stastka, H., Voracek, M., Leithner, K., Fischer-Kern, M., Kunz, C., & Meissel, T. (2003). Predicting psychotherapy utilization for patients with borderline personality disorder. *Psychotherapy Research, 13*(2), 255–264.
- Luteijn, F., & Barelids, D. P. H. (2004). *Groninger Intelligentie test 2: Handleiding [Groninger intelligence test 2: manual]*. Amsterdam: Harcourt Assessment B.V.
- McMurrin, M., Huband, N., & Overton, E. (2010). Non-completion of personality disorder treatments: a systematic review of correlates, consequences, and interventions. *Clinical Psychology Review, 30*, 277–287.
- Merens, W., Van der Does, A. W., & Spinhoven, P. (2007). The effects of serotonin manipulations on emotional information processing and mood. *Journal of Affective Disorders, 103*(1), 43–62.
- Moncrieff, J., & Cohen, D. (2009). How do psychiatric drugs work? *BMJ, 338*.
- Moncrieff, J., Cohen, D., & Mason, J. P. (2009). The subjective experience of taking antipsychotic medication: a content analysis of internet data. *Acta Psychiatrica Scandinavica, 120*(2), 102–111.
- Moradveisi, L., Huibers, M. J., & Arntz, A. (2015). The influence of patients' attributions of the immediate effects of treatment of depression on long-term effectiveness of behavioural activation and antidepressant medication. *Behaviour Research and Therapy, 69*, 83–92.
- Nadort, M., Arntz, A., Smit, J. H., Giesen-Bloo, J., Eikelenboom, M., Spinhoven, P., et al. Dyck van, 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*, 961–973.
- Olsen, S. A., & Beck, J. G. (2012). The effects of dissociation on information processing for analogue trauma and neutral stimuli: a laboratory study. *Journal of Anxiety Disorders, 26*, 225–232.
- Perry, J. C., Bond, M., & Roy, C. (2007). Predictors of treatment duration and retention in a study of long-term dynamic psychotherapy: childhood adversity, adult personality, and diagnosis. *Journal of Psychiatric Practice, 13*, 221–232.
- Plakun, E. M. (1991). Prediction of outcome in borderline personality disorder. *Journal of Personality Disorders, 5*, 93–101.
- Plassman, B. L., Welsh, K. A., Helms, M., Brandt, J., Page, W. F., & Breitner, J. C. S. (1995). Intelligence and education as predictors of cognitive state in late life: a 50-year follow-up. *Neurology, 45*, 1446–1450.
- Powers, M. B., Smits, J. A., Whitley, D., Bystritsky, A., & Telch, M. J. (2008). The effect of attributional processes concerning medication taking on return of fear. *Journal of Consulting and Clinical Psychology, 76*(3), 478.
- Preacher, K. J., & Hayes, A. F. (2008a). *SPSS macro for multiple mediation, written by A.F. Hayes*. The Ohio State University Accessed 27.03.15 <http://www.afhayes.com>.
- Preacher, K. J., & Hayes, A. F. (2008b). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–891.
- Price, J., Cole, V., & Goodwin, G. M. (2009). Emotional side-effects of selective serotonin reuptake inhibitors: qualitative study. *The British Journal of Psychiatry, 195*(3), 211–217.
- Price, J., & Goodwin, G. M. (2009). Emotional blunting or reduced reactivity following remission of major depression. *Medicographia, 31*, 152–156.
- Pringle, A., Browning, M., Cowen, P. J., & Harmer, C. J. (2011). A cognitive neuro-psychological model of antidepressant drug action. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 35*(7), 1586–1592.
- Rijkeboer, M. M., & van den Bergh, H. (2006). Multiple group confirmatory factor analysis of the young schema-questionnaire in a Dutch clinical versus non-clinical population. *Cognitive Therapy and Research, 30*(3), 263–278.
- 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 a non-clinical population. *Journal of Behavior Therapy and Experimental Psychiatry, 36*, 129–144.
- Rüsch, N., Schiel, S., Corrigan, P. W., Leihener, F., Jacob, G. A., Olschewski, M., et al. (2008). Predictors of dropout from inpatient dialectical behaviour therapy among women with borderline personality disorder. *Journal of Behavior Therapy and Experimental Psychiatry, 39*, 497–503.
- Ryle, A., & Golyukina, K. (2000). Effectiveness of time-limited cognitive analytic therapy of borderline personality disorder: factors associated with outcome. *British Journal of Medical Psychology, 73*, 197–210.
- Schuppert, H. M., Timmerman, M. E., Bloo, J., van Gemert, T. G., Wiersema, H. M., Minderaa, R. B., ... Nauta, M. H. (2012). Emotion regulation training for adolescents with borderline personality disorder traits: a randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(12), 1314–1323.
- Smith, T. E., Koenigsberg, H. W., Yeomans, F. E., Clarkin, J. F., & Selzer, M. A. (1995). Predictors of dropout in psychodynamic psychotherapy of borderline personality disorder. *Journal of Psychotherapy Practice and Research, 4*, 205–213.
- Spiegel, D. (1994). *Dissociation: Culture, mind, and body*. Washington, DC: American Psychiatric Press.
- Steinberg, M. (1993). *Structured clinical interview for DSM-IV dissociative disorders (SCID-D)*. Arlington, VA, US: American Psychiatric Association.
- Stiglmayr, C. E., Shapiro, D. A., Stieglitz, R. D., Limberger, M. F., & Bohus, M. (2001). Experience of aversive tension and dissociation in female patients with borderline personality disorder—a controlled study. *Journal of Psychiatric Research, 35*(2), 111–118.
- Tambs, K., Sundet, J. M., Magnus, P., & Berg, K. (1989). Genetic and environmental contributions to the covariance between occupational status, educational attainment, and IQ: a study of twins. *Behavior Genetics, 19*, 209–222.
- Thormählen, B., Weinryb, R. M., Norén, K., Vinnars, B., Bågedahl-Strindlund, M., & Barber, J. P. (2003). Patient factors predicting dropout from supportive-expressive psychotherapy for patients with personality disorders. *Psychotherapy Research, 13*, 493–509.
- Vanderlinden, J., Van Dyck, R., Vandereycken, W., Vertommen, H., & Verkes, J. R. (1993). The dissociation questionnaire (DIS-Q): development and characteristics of a new self-report questionnaire. *Clinical Psychology & Psychotherapy, 1*(1), 21–27.
- Verheul, R., van den Bosch, L. M., Koeter, M. W., De Ridder, M. A., Stijnen, T., & Van Den Brink, W. (2003). Dialectical behaviour therapy for women with borderline personality disorder 12-month, randomised clinical trial in the Netherlands. *British Journal of Psychiatry, 182*(2), 135–140.
- Wechsler, D. (2005). *WAIS-III-NL Nederlandstalige bewerking. Technische handleiding (herziene uitgave 2005) [WAIS-III-NL Dutch adaptation. Technical Manual (revised edition 2005)]*. Amsterdam: Harcourt Test Publishers.
- Weertman, A., Arntz, A., & Kerkhofs, M. L. M. (2000). *Structured clinical interview for DSM-IV Axis II personality disorders*. Lisse, the Netherlands: Swets & Zeitlinger BV.
- Zanarini, M. C., Stanley, B., Black, D. W., Markowitz, J. C., Goodman, M., Pilonis, P., et al. (2010). Methodological considerations for treatment trials for persons with borderline personality disorder. *Annals of Clinical Psychiatry: Official Journal of the American Academy of Clinical Psychiatrists, 22*(2), 75.