



UvA-DARE (Digital Academic Repository)

Differential effectiveness of dialectical behavioural therapy and schema therapy in patients with borderline personality disorder

a secondary analysis of a randomised clinical trial

Assmann, Nele; Kaiser, Tim; Herzog, Philipp; Arntz, Arnoud; Klein, Jan Philipp; Fassbinder, Eva; Schaich, Anja

DOI

[10.1016/j.brat.2025.104899](https://doi.org/10.1016/j.brat.2025.104899)

Publication date

2025

Document Version

Final published version

Published in

Behaviour Research and Therapy

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Assmann, N., Kaiser, T., Herzog, P., Arntz, A., Klein, J. P., Fassbinder, E., & Schaich, A. (2025). Differential effectiveness of dialectical behavioural therapy and schema therapy in patients with borderline personality disorder: a secondary analysis of a randomised clinical trial. *Behaviour Research and Therapy*, 195, Article 104899. <https://doi.org/10.1016/j.brat.2025.104899>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)



Differential effectiveness of dialectical behavioural therapy and schema therapy in patients with borderline personality disorder: a secondary analysis of a randomised clinical trial

Nele Assmann^{a,*}, Tim Kaiser^b, Philipp Herzog^{c,d}, Arnoud Arntz^e, Jan Philipp Klein^a, Eva Fassbinder^{f,1}, Anja Schaich^{a,1}

^a Department of Psychiatry, Psychosomatics and Psychotherapy, University of Lübeck, Lübeck, Germany

^b Division of Methods and Evaluation/Quality Assurance, Freie Universität Berlin, Berlin, Germany

^c Department of Psychology, University of Kaiserslautern–Landau (RPTU), Landau, Germany

^d Department of Psychology, Harvard University, Cambridge, MA, 2138, USA

^e Department of Clinical Psychology, University of Amsterdam, Amsterdam, the Netherlands

^f Department of Psychiatry and Psychotherapy, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

ARTICLE INFO

Keywords:

Borderline personality disorder
Differential effectiveness
Treatment heterogeneity
Dialectical behaviour therapy
Schema therapy
Personalized advantage index

ABSTRACT

Background: Borderline personality disorder (BPD) can be treated successfully with specific psychological treatments, but there is no clear evidence of superiority of one specific treatment at the group level. Due to high heterogeneity in BPD, individual patients might benefit differently from specific treatments.

Methods: Based on a randomised trial comparing 18 months of dialectical behaviour therapy (DBT) and schema therapy (ST) for BPD, differential effectiveness was examined using causal forest analyses. Baseline variables on BPD criteria, general psychopathology, traumatic childhood experiences, rejection sensitivity, level of functioning, coping skills, schemas and medication were included to predict the Borderline Personality Disorder Severity Index (BPDSI-IV) during treatment and follow-up (24 and 30 months after start of treatment).

Results: A subgroup was identified that benefited significantly more from DBT compared to ST. This group showed a significantly greater reduction in symptoms post treatment (post-treatment difference of 5.79 BPDSI points, SMD = 0.65, $p = .028$), but no longer at follow-up ($p = .771$). The group that showed better results with DBT displayed a pattern of specific baseline characteristics: higher levels of functioning, less frequent emotional neglect and sexual abuse, more severe anxiety symptoms and more pronounced schema 'failure to achieve'. No pattern of variables was identified associated with a superiority of ST.

Conclusion: Moderators of the short-term effect of DBT versus ST were found. However, this moderator effect was no longer significant at follow-up. Identifying patient characteristics associated with differential treatment effect might be a promising way to improve BPD treatment outcomes faster.

Retrospectively registered (German Clinical Trials Register: DRKS00011534) without protocol changes.

1. Introduction

Borderline personality disorder (BPD) is a difficult-to-treat, prevalent mental disorder associated with severe psychosocial impairments and a high number of comorbidities (Bohus et al., 2021; Tomko et al., 2014). Fortunately, ample evidence exists that BPD can be treated successfully with BPD-specific psychological treatments (Stoffers-Winterling et al., 2022; Storebø et al., 2020), including, among others, dialectical

behaviour therapy (DBT), mentalisation-based therapy (MBT), and schema therapy (ST). So far, however, there is no clear evidence that a particular disorder-specific treatment is superior to others. Furthermore, no treatment has shown universal effectiveness, meaning there is always a group of patients who do not respond satisfactorily (Woodbridge et al., 2021). Given the high burden of disease related to BPD and the insufficient capacity of treatment places (Iliakis et al., 2019), research on treatment variability is crucial. Even though emotional instability is a

* Corresponding author. Department of Psychiatry and Psychotherapy University of Lübeck, Ratzeburger Allee 160, 23538, Lübeck, Germany.

E-mail address: nele.assmann@uksh.de (N. Assmann).

¹ these authors share last authorship.

<https://doi.org/10.1016/j.brat.2025.104899>

Received 9 July 2025; Received in revised form 1 October 2025; Accepted 16 October 2025

Available online 23 October 2025

0005-7967/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

central feature of BPD (D'Aurizio et al., 2023), there is a high heterogeneity among patients with BPD indicating that individuals may respond to different psychological treatments in different ways. Enhancing knowledge of which patient characteristics are related to better outcomes in specific psychological treatments could enable more precise treatment recommendations (Cohen et al., 2021).

The variability in treatment effects is made up of random variability and nonrandom variability. The latter can be associated with patient characteristics and has been referred to as the heterogeneity of treatment effects (HTE) (Varadhan & Seeger, 2013). To improve the outcome of psychotherapy, detecting and quantifying HTE is of special interest. A recent analysis showed that HTE can be assumed to exist for established psychological BPD treatments (Kaiser & Herzog, 2023).

The most common approach to analyzing HTE are subgroup analyses, which can identify subgroups in a population that respond better (or worse) to a specific treatment (Varadhan & Seeger, 2013). Whereas in subgroups usually one variable at a time is evaluated as a potential predictor, the personalised advantage index (PAI) (Chekroud et al., 2021; DeRubeis et al., 2014) combines multiple potential moderators and thereby enables the prediction of treatment response for a specific patient receiving a given treatment.

Most randomised controlled trials on BPD used non-specialised treatment control groups and did not compare two specialised treatment arms. Given the limited resources available, an empirically based treatment selection tool would help to increase the effectiveness of mental health care. Thus, this promising approach has been used several times in patients with depression (e.g. Friedl et al., 2020; van Bronswijk, Bruijnijk, et al., 2021; van Bronswijk, DeRubeis, et al., 2021), but only one study has used the PAI for treatment selection in BPD patients (Keefe et al., 2021). The results of this study indicated that patients reporting more emotional abuse, showing more depressive symptoms, dependent personality traits, and social maladjustment showed greater improvement when they were treated with DBT, whereas patients with more general symptoms and more impulsive BPD symptoms benefited more from psychodynamic psychiatric management.

We recently compared outpatient DBT and ST for the first time in a head-to-head randomised clinical trial in patients with severe BPD (Assmann et al., 2024). In this study, patients in both treatment conditions improved significantly over time, but there was no difference between DBT and ST in the overall reduction of BPD severity. Thus, the average treatment effect (ATE), defined as the average difference of outcomes between DBT and ST, was not significantly different from zero. In addition to determining the average treatment effect, however, one important objective of the study was to investigate personalisation by identifying patients who especially benefit from DBT or ST, respectively. Even though the ATE is near zero, individual treatment effects might vary, which statistically manifests as HTE.

We explored which clinical baseline patient characteristics moderated the effect of treatment (DBT vs. ST) on the reduction of BPD symptom severity. The second aim was to develop a PAI that indicates which patients benefit relatively more from which of the two treatments. Since this is the first such analysis with BPD patients treated with DBT and ST, we exploratively analysed which characteristics are associated with better outcome in which treatment.

2. Methods

2.1. Sample

In this study, we analysed data from $N = 164$ patients with BPD participating in the PRO*BPD study (Assmann et al., 2024). The study was designed to compare the effectiveness of DBT and ST using a randomised clinical trial. Patients were randomly assigned to one of the two treatment options.

The study was reviewed and approved by the ethics committee of Lübeck University, reference number 13–005. All patients provided

written informed consent to participate.

2.2. Intervention

Patients received one group therapy sessions (100 min) and one individual therapy session (60 min) per week for up to 18 months of either DBT or ST. Although both ST and DBT have their roots in CBT, there are major differences between the two methods (for a detailed description see Fassbinder et al., 2016). In DBT, the main focus is on teaching emotion regulation skills (Linehan, 2015a, 2015b), whereas ST aims to change core dysfunctional schemas developed from aversive childhood experiences (Arntz & van Genderen, 2009; Young et al., 2003). Therapists were advanced DBT and ST therapists as well as CBT-experienced therapists who were new to ST or DBT, all of whom received training before administering the treatment. All therapists received weekly supervision sessions. For a detailed description of both interventions, please refer to the study protocol (Fassbinder et al., 2018).

2.3. Measures

Assessments were completed at baseline, six months after the start of treatment, 12 months after the start of treatment, post-treatment (18 months after start of treatment), and at two follow-up assessments (24 months and 30 months after start of treatment).

2.4. Outcome

2.4.1. Borderline Personality Disorder Severity Index for DSM-IV (BPDSI-IV)

The BPDSI (Arntz et al., 2003) is a semi-structured clinical interview assessing manifestations of BPD criteria over the past three months. It was developed according to the DSM-IV criteria and consists of 70 items related to the nine BPD-criteria: (1) abandonment, (2) interpersonal relationships, (3) identity, (4) impulsivity, (5) parasuicidal behaviour, (6) affective instability, (7) feelings of emptiness, (8) bursts of anger, (9) dissociation and paranoid ideation. Since criteria of BPD did not change from DSM-IV to DSM-5 the BPDSI is also valid for DSM-5. Individual cluster criteria scores from the baseline assessment were included as predictors. The BPDSI yields scores per BPD-criterion, as well as a total score, indexing the severity of BPD expressions during the last three months. It has excellent psychometric features (Cronbach's alpha = 0.85; interrater reliability 0.99, high validity and sensitivity to change) (Giesen-Bloo et al., 2010; Kröger et al., 2013).

We used the total BPDSI score as the outcome variable. Like Keefe et al. (2021), we intended to predict the reduction of BPD severity over all measurement periods, representing long-term improvement. Thus, we calculated the area under the curve of BPDSI values (BPDSI-AUC) including five assessment points six, 12, and 18 months (during treatment), as well as 24 and 30 months after the start of treatment (follow-up), and used this variable as the prediction target. The Kalman filter algorithm imputed missing data in the outcome variable. The Kalman filter imputed both missing values at the end of a time series and missing values between available values.

2.5. Moderators

The following measures were used to predict the reduction of BPD symptoms. All measures were included at the subscale level if available.

2.5.1. Work and Social Adjustment Scale (WSAS)

The WSAS (Mundt et al., 2002) is a self-report measure of perceived impairment in everyday tasks (e.g., work, leisure, or relationships). It consists of five items rated on a 0 (not at all) to 8 (very severely) scale that are summed to yield a single total score. The WSAS has shown to be reliable, valid, and change-sensitive in different patient samples (Thandi et al., 2017; Zahra et al., 2014).

2.5.2. DBT ways of coping checklist (DBT-WCCL)

The DBT-WCCL (Neacsu et al., 2010) is a self-report measure designed to assess functional and dysfunctional coping skills. The first subscale assesses functional coping behaviour (i.e., emotion regulation, mindfulness, crisis survival, reality acceptance, and interpersonal effectiveness). In contrast, the second subscale includes strategies that are considered dysfunctional (e.g., self-blame, substance use, avoidance, denial). The items are rated from 0 (never used) to 3 (regularly used) and refer to the past month. The developers could find good to excellent psychometric properties for the DBT-WCCL (Neacsu et al., 2010).

2.5.3. Brief symptom inventory (BSI)

The BSI (Derogatis & Melisaratos, 1983) - a shorter version of the Symptom Checklist-90-R - is an assessment tool designed to evaluate a broad range of psychological symptoms and distress. This scale consists of 53 items, which measure nine primary symptom dimensions (somaticisation, obsession-compulsion, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism) and three global indices of distress and has good psychometric properties (Derogatis & Melisaratos, 1983). The nine symptom dimensions were included as predictors.

2.5.4. Childhood trauma questionnaire (CTQ)

The CTQ (Bernstein et al., 2003) is a self-report instrument designed to assess the history of childhood abuse and neglect in both clinical and non-clinical populations with good psychometric properties (Wingenfeld et al., 2010). It consists of 28 items that cover five main dimensions: emotional, physical, and sexual abuse, as well as emotional and physical neglect. The items are rated from 1 (never true) to 5 (very often true).

2.5.5. Young Schema Questionnaire (YSQ)

The YSQ (Siegmund et al., 2011) is a self-report questionnaire designed to capture dysfunctional schemas that are, according to schema theory, deeply ingrained emotional and cognitive patterns that can develop as a result of adverse childhood experiences. The German version used in this study consists of 75 descriptive statements that are rated on a 6-point Likert scale (1 = very inaccurate to 6 = very accurate) and can be scored on the following schemas: abandonment, mistrust, emotional deprivation, social isolation, dependence, vulnerability to harm, enmeshment, failure to achieve, entitlement, insufficient self-control, subjugation, self-sacrifice, emotional inhibition and unrelenting standards. The YSQ has adequate internal consistency and good reliability (Baranoff et al., 2006).

In addition to these measures, the current medication status (medication vs. no medication) was included.

3. Statistics

3.1. Causal forest

The Causal Forest (CF) algorithm (Wager & Athey, 2018) is a machine learning method based on the random forest. It repeatedly splits data using potential moderator variables to maximise the difference in outcomes between two treatment groups. The importance of individual variables was determined using the “drop and refit” procedure (Bénard & Josse, 2023). This procedure involves fitting multiple CF models to the data, leaving out individual variables to determine the change in predictive accuracy when they are not included in the model. This method has been shown to produce more accurate variable importance scores than other available alternatives, both in simulation studies and real-world applications (Bénard & Josse, 2023). Higher importance scores indicate that a variable is more relevant in determining differences in treatment outcomes, while the importance of variables that are irrelevant for heterogeneity in treatment effects approaches zero. The variable selection process was performed in two steps. First, a CF

including all moderators was fitted, including a randomly generated variable as an additional predictor. The importance of this variable was used as a cutoff to select variables that contributed to the heterogeneity of treatment effects. Next, a CF model using only the retained moderators was fitted. CF was applied to the complete data set. Missing data in moderators were handled internally by the algorithm using the missing incorporated in attributes criterion (MIA) (Twaal et al., 2008).

3.2. Estimates of differential treatment effects

In personalised treatment selection, it is common to assess the usefulness of statistical decision rules for assigning treatments by calculating the impact of receiving the best treatment based on the model's predictions. Patients who are predicted to show the strongest effects under any of the treatments can be prioritised based on these predictions. We utilised the Target Operating Characteristic (TOC) to estimate the benefit of this prioritisation for a range of percentiles, providing estimates for the ATE if only a certain percentage of patients with the highest predicted treatment outcomes were treated. The rank-weighted average treatment effect (RATE) represents the area under the TOC curve. A RATE that is significantly different from zero indicates that the corresponding rule is effective in assigning patients to their optimal treatment.

3.3. Additional statistical tests

We used the final CF model to predict the differences in outcomes between DBT and ST for each patient in the dataset. This difference is commonly referred to as conditional average treatment effect (CATE). A predicted difference of zero indicates that, given the included covariates, a patient is expected to show equal outcomes in both options, whereas a significant deviation from zero suggests that a patient will have a better outcome under one of the options. We determined significant deviations from zero by calculating 95 % confidence intervals for individual predictions. Positive values indicate that a patient is likely to have a better outcome under DBT, while negative values suggest that a patient may benefit more from ST. Patients were classified as having received their optimal therapy when their predicted treatment effect significantly differed from zero in the direction expected for the therapy to which they were assigned.

We were also interested in whether patients for whom one of the two methods was predicted to be optimal also benefited on the original scale instead of the area under the curve. Thus, we fitted a piecewise linear mixed model that predicts the BPDSI total score from two continuous time variables (time during treatment, time during follow-up) and a binary variable indicating whether a patient received their optimal treatment (yes/no). We included random slopes and intercepts to account for differences in BPDSI baseline scores and to allow for individual differences in symptom reduction over time. Following the well-established dose effect in psychotherapy (Bone et al., 2021; Howard et al., 1986), we assumed that symptom reduction was greatest during the earlier stages of the active treatment phase and decreased in the follow-up. We applied a base-10 logarithm to the time variables.

This model helped us determine the benefit of receiving the optimal treatment on the actual scale of the primary outcome (i.e., the BPDSI total score) instead of the area under the curve used for the CF model, while still accounting for nonlinear trends.

4. Results

4.1. Sample characteristics

The data available for analysis included a total of $N = 164$ patients, with 83 patients who received an average of $M = 44.5$ ($SD = 18.10$) therapy sessions of DBT and 81 patients who received an average of $M = 46.7$ ($SD = 16.20$) ST sessions. The baseline sociodemographic and

clinical characteristics of the sample are displayed in Table 1. The available data decreased steadily over the course of treatment (from 80 % DBT and 85 % ST at 6 months to 61 % DBT and 63 % ST posttreatment) and during the follow-up phase (52 % DBT and 42 % ST at 30 months), as would be expected with this disorder and the long follow-up period. See main analysis for more details (Assmann et al., 2024).

4.2. Causal forest

Most predictors included in the CF model received a variable importance that was lower than the random variable. For the final model, the WSAS total score, CTQ sexual abuse and emotional neglect, BSI anxiety, and YSQ ‘failure to achieve’ were retained. Table 2 provides an overview of the relationships between the patient characteristics and the superiority of one of the two treatment protocols. Graphical representations of the nonlinear associations between individual predictors and conditional average treatment effect (CATE) estimates are shown in the Supplementary Material (Fig. S1).

On average, patients receiving ST had a 6.88 (95 % CI: 5.41; 12.29) points higher BPDSI-AUC, indicating no significant advantage of either DBT or ST when considering all points of assessment. However, the RATE was estimated at 17.10 (95 % CI: 6.96; 27.24), indicating a statistically significant benefit of assigning patients to their optimal treatment.

Table 1
Descriptive statistics at baseline.

Characteristic	Treatment condition		
	All (N=164)	DBT (n=83)	ST (n=81)
Age, years <i>M(SD)</i>	33.71 (10.61)	34.54 (11.15)	32.85 (10.01)
Gender <i>n</i> (%)			
male	33 (20.1)	17 (20.5)	16 (19.8)
female	130 (79.3)	66 (79.5)	64 (79.0)
non-binary	1 (0.6)	–	1 (1.2)
Relationship status <i>n</i> (%)			
Partner	82 (50.0)	41 (49.4)	41 (50.6)
No partner	82 (50.0)	42 (50.6)	40 (49.4)
Education level <i>n</i> (%)			
No/primary education	9 (5.5)	5 (6.0)	4 (4.9)
Lower secondary education	92 (56.1)	46 (55.4)	46 (56.8)
Upper secondary education	49 (29.9)	26 (31.3)	23 (28.4)
Tertiary education	14 (8.5)	6 (7.2)	8 (9.9)
Ethnic background <i>n</i> (%)			
German	151 (92.1)	78 (94.0)	74 (91.4)
Different	12 (7.3)	5 (6.0)	7 (8.6)
Work status <i>n</i> (%)			
Working	20 (12.2)	7 (8.4)	13 (16.0)
Studying	17 (10.4)	7 (8.4)	10 (12.3)
Homemaker	12 (7.3)	6 (7.2)	6 (7.4)
Disability Pension	41 (25.0)	22 (26.5)	19 (23.5)
Unable to work due to sick leave	44 (26.8)	25 (30.1)	19 (23.5)
Unemployed	15 (9.1)	8 (9.6)	7 (8.6)
Retirement pension	3 (1.8)	3 (3.6)	–
Other	12 (7.3)	5 (6.0)	7 (8.6)
BPD-severity			
BPDSI at Baseline <i>M(SD)</i>	32.80 (8.71)	33.19 (9.38)	32.39 (8.01)
Number of BPD criteria (SKID-II)	7.26 (1.26)	7.23 (1.32)	7.30 (1.21)
Comorbid Disorders <i>M (SD)</i>			
Number of comorbid SCID-I Diagnoses	3.99 (1.95)	4.12 (1.88)	3.85 (2.03)
Number of comorbid SCID-II Diagnoses (excl. BPD)	1.35 (1.12)	1.37 (1.06)	1.32 (1.18)
Psychiatric medication ¹ <i>n</i> (%)	126 (76.8)	64 (77.1)	62 (76.5)
Previous treatment <i>n</i> (%)	154 (93.9)	82 (98.8)	72 (88.9)
Previous psychotherapeutic treatment <i>n</i> (%)	151 (92.1)	80 (96.4)	71 (87.7)

Note. DBT: Dialectical Behaviour Therapy. ST: Schema Therapy.

Table 2
Overview of the effects of retained predictors for the final causal forest model.

Variable	Lower scores	Higher scores
WSAS	DBT > ST	DBT = ST
CTQ-SA	DBT > ST	ST > DBT
CTQ-EN	DBT > ST	DBT = ST
BSI-ANX	DBT > ST	DBT = ST
YSQ-Failure	DBT = ST	DBT > ST

Note. WSAS: Work and Social Adjustment Scale. CTQ-SA: Childhood trauma questionnaire, sexual abuse subscale. CTQ-EN: emotional neglect subscale. BSI-ANX: Brief Symptom Inventory Anxiety subscale. YSQ-Failure: Young Schema Questionnaire, ‘failure to achieve’ subscale. DBT: Dialectical Behaviour Therapy. ST: Schema Therapy. ‘DBT > ST’: better outcome under DBT. ‘ST > DBT’: better outcome under schema therapy. ‘DBT = ST’: no better outcome in either of the options. The predictors cannot be interpreted individually, but only in combination.

Overall, 23 patients (17.83 %) received their optimal treatment according to the CF model. As depicted in Fig. 1, there were only patients predicted to show better treatment outcomes when receiving DBT. While some patients had point predictions that showed greater improvement under ST, their confidence intervals did not indicate a significant difference from zero.

Fig. 2 summarises the course of BPDSI symptoms for patients who received their optimal treatment and those who did not. Between these groups, BPDSI scores were significantly different after treatment completion, but not during follow-up.

The linear mixed model, as illustrated in Table 3, showed a significant decrease in BPD symptoms over time in the treatment phase. When marginalising over all time points, the patients who received their optimal treatment had BPDSI scores that were on average 5.79 points (95 % CI: 0.10; 11.47) or 0.65 standard deviations (95 % CI: 0.01; 1.30) lower compared to those who did not receive their optimal treatment.

5. Discussion

In this study, we analysed data from the first randomised trial comparing DBT and ST in patients with BPD to examine which patient characteristics might be related to their differential effectiveness. The results suggest certain patient characteristics may be related to differential effectiveness of DBT and ST.

As shown in Table 2, patients with a pattern of the following characteristics: lower levels of everyday life impairment, less childhood sexual abuse and emotional neglect, lower level of anxiety, and higher scores on the schema ‘failure to achieve’ improved more if they were treated with DBT instead of ST. No variables were identified as being related to a relatively better outcome after ST compared to DBT treatment.

Symptom course over time shows that receiving the optimal treatment is associated with a significant better treatment outcome in the treatment phase until the post-treatment assessment. In the follow-up phase, the optimal treatment group stagnated whereas the non-optimal treatment group improved further. We found no evidence for a between-group difference in this phase. One reason for this might be the high proportion of missing values. Although we imputed missing values, every imputation of missing values represents an additional source of uncertainty in the parameter estimation (Rubin, 1987). It is also possible that the effect of personalised assignment to DBT is actually only temporary. This would also be a relevant finding, as a faster reduction in symptoms implies less suffering and lower costs, particularly in the case of severe borderline disorder studied in this sample.

The possibility to compare our results to previous studies is limited since DBT and ST have never been directly compared in a randomised clinical trial before. An ongoing multicenter trial comparing DBT and ST aims to analyse differential treatment effects to enable treatment selection for BPD patients (Wibbelink et al., 2022), but the study has not

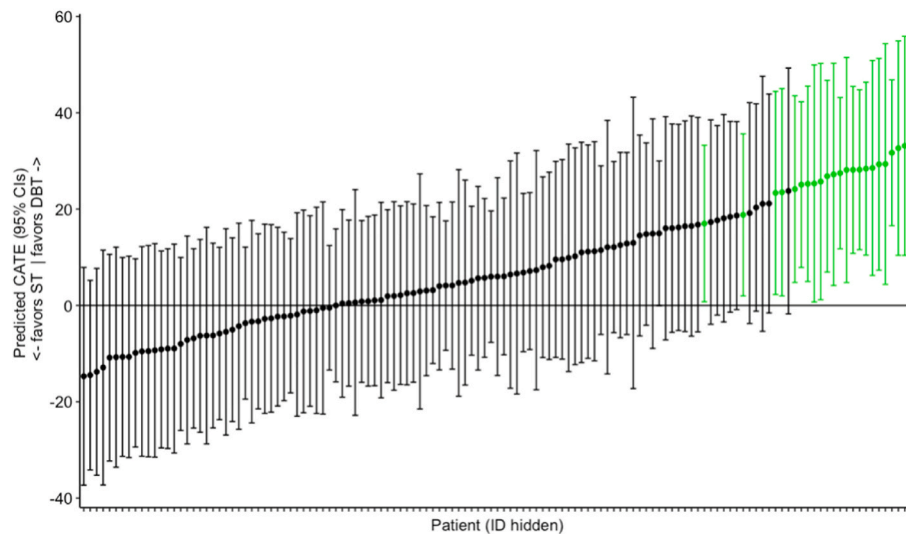


Fig. 1. Caterpillar plot showing differences in treatment effects for all included patients. ST: Schema Therapy. DBT: Dialectical-Behavioural Therapy. CATE: conditional average treatment effect. Values with 95 % confidence intervals not including zero were colored green. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

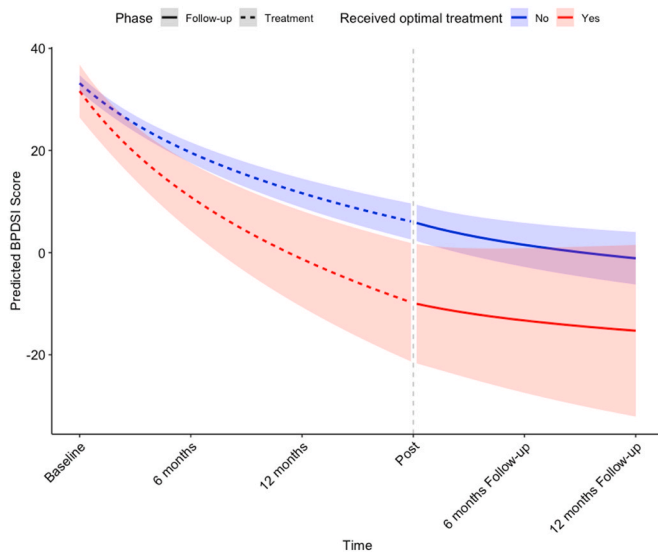


Fig. 2. Outcomes of patients depending on whether they were assigned to their optimal treatment, separated by treatment and follow-up phase. Shaded areas represent 95 % confidence intervals. BPDSI: Borderline Personality Disorder Severity Index.

been completed yet. Nevertheless, our results regarding DBT differ from those of Keefe et al. (2021) who found that more emotional abuse, depressive symptoms, dependent personality traits, and social maladjustment were characteristics associated with better outcomes after DBT compared to psychodynamic psychiatric management (PPM). Especially in terms of impairment and childhood maltreatment, these results contrast with our findings. However, the sample size ($N = 156$) of Keefe et al. (2021) was also too small to allow reliable clinical recommendations. This, along with the differences in interventions (ST vs. PPM), sample characteristics, and instruments, might explain the inconsistent results. Whereas the outcome of our analysis was the manifestation of BPD criteria, Keefe et al. (2021) used the Global Severity Index (GSI) of the Symptom Checklist-90, reflecting a broader assessment of psychopathology and dysfunction. Regarding the predictors, there was a certain overlap, but also clear differences between the analyses.

Our finding that patients reporting less sexual abuse and emotional neglect improved more in the DBT treatment during treatment might be explained by DBT’s focus on handling difficult emotions and tension in present situation, instead of working with childhood memories, which is one of the major foci of ST. The superiority of DBT for patients with less impairment is unexpected, as previous studies suggest that DBT might outperform other treatments for patients with higher impairment due to its structured skills training (Keefe et al., 2021; Sahin et al., 2018). Given the explorative character of our study, these results should be interpreted with great caution. This aligns well with previous findings showing the difficulty of identifying stable moderators of treatment outcomes in BPD (Herzog et al., 2020).

Of note, no pattern of moderators predicting a significantly better outcome with ST treatment could be identified.

Further research is needed to determine whether our analysis failed to identify patient characteristics related to a better outcome in ST due to our sample size or the limited number of predictor variables, or whether no such predictors exist. For example biological parameters, such as sleep quality, would also be an interesting addition to the moderators (Socci et al., 2024).

Although our results do not represent reliable recommendations for clinicians because of the small sample sizes and the explorative nature of our study, the additional symptom reduction of 26 % when comparing optimal and non-optimal treatment allocation might be considered as clinically relevant, underlining the need for future research.

This analysis is the first comparing DBT and ST using an innovative procedure to analyse differential treatment response. Further strengths of the PRO*BPD trial also include the high quality of outcome measures, the one-year follow-up period, and minimal exclusion criteria that allowed BPD patients with a wide range of heterogeneity to participate. The main limitation of our analysis is the relatively small sample size. With $N = 164$ patients, powered to detect medium effect sizes between groups, we can only gain an initial impression of the differential effectiveness of ST compared to DBT, but cannot develop a reliable treatment selection strategy. Clinical trials powered to detect medium effect sizes between groups are often underpowered for treatment selection using machine learning analyses (Lorenzo-Luaces et al., 2021; Luedtke et al., 2019). The omission of sex- and gender-based analyses due to gender imbalances and the limited sample size also restricts the generalizability of our results. Lastly, the small sample size prevented holding out a random subsample for cross-validation of the derived model. Therefore, the present results should be tested in an independent sample. External

Table 3
Results of the linear mixed model predicting BPDSI sum scores from time and optimal treatment allocation.

Parameter	Estimate	Std. beta	95 % CI		Std. 95 % CI		p
			Lower	Upper	Lower	Upper	
(Intercept)	31.12		29.43	32.61			<0.001
log ₁₀ (Time: treatment)	-19.59	-0.36	-22.28	-16.9	-0.41	-0.31	<0.001
log ₁₀ (Time: follow-up)	-6.48	-0.10	-9.51	-3.45	-0.15	-0.05	<0.001
Optimal treatment: Yes	-1.53	-0.13	-6.96	3.90	-0.57	0.32	0.5581
log ₁₀ (Time: treatment) x Optimal treatment: Yes	-10.35	-0.85	-19.56	-1.14	-1.61	-0.09	0.028
log ₁₀ (Time: follow-up) x Optimal treatment: Yes	1.54	0.02	-8.84	11.93	-0.14	0.19	0.771
$R^2_{\text{marginal}} = 0.22$ $R^2_{\text{conditional}} = 0.86$ $\tau_{ID} = 7.76$ $\tau_{\text{time: treatment}} = 11.17$ $\tau_{\text{time: follow-up}} = 12.32$							

Note. 95 % CI: 95 % confidence interval. Std. beta: standardised regression coefficients. 95 % CIs and p values were computed using a Wald t-distribution approximation. log₁₀: base-10 logarithm. R^2_{marginal} : variance in outcomes explained by fixed effects (i.e., time variables and optimal treatment assignment), $R^2_{\text{conditional}}$: variance in outcomes explained by fixed and random effects (i.e., individual differences in baseline and symptom reduction). τ_{ID} : standard deviation of intercepts. $\tau_{\text{time: treatment}}$: standard deviation of regression slopes for time variables. BPDSI: Borderline Personality Severity Index.

validation on an independent testing sample is necessary to assess the real-world applicability of the model by evaluating its generalisability to new populations. A systematic review of psychiatric prediction models found that three quarters used at least one internal validation, but only one in five were included external validation (Meehan et al., 2022). Models that have undergone proper external validation have often produced disappointing results (Chekroud et al., 2024; van Bronswijk, Brujniks, et al., 2021). Another limitation is the high proportion of missing data at the follow-up assessments. Further limitations of the PRO*BPD trial also apply to this analysis (Assmann et al., 2024).

6. Conclusions

This was the first study to analyse treatment heterogeneity in BPS patients treated with ST and DBT. Our results provide explanatory signals that BPD patients with a certain pattern of characteristics showed better acute treatment response to DBT, suggesting that treatment selection might be a promising approach for these patients. Replications in independent samples with larger sample sizes are required to confirm our results to develop reliable treatment selection strategies that can improve outcomes for BPD patients in clinical practice.

CRedit authorship contribution statement

Nele Assmann: Writing – original draft, Project administration, Investigation, Data curation, Conceptualization. **Tim Kaiser:** Writing – original draft, Methodology, Formal analysis. **Philipp Herzog:** Writing – review & editing, Investigation. **Arnoud Arntz:** Writing – review & editing, Methodology, Conceptualization. **Jan Philipp Klein:** Writing – review & editing. **Eva Fassbinder:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization. **Anja Schaich:** Writing – review & editing, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization.

Funding sources

The Pro*BPD study was supported by the Else-Kröner-Fresenius-Stiftung (2018_A152). Dr. Fassbinder obtained internal funding from the University of Lübeck (Einzelprojektförderung und Habilitationsförderung für Wissenschaftlerinnen, Sektion Medizin). Dr. Schaich obtained funding from the University of Lübeck (Habilitationsförderung für Wissenschaftlerinnen, Sektion Medizin). Funding bodies played no role in the design of the study, in the collection, analysis, and interpretation of data, in the writing of the manuscript, and in the decision to submit the manuscript for publication.

Declaration of competing interest

Prof. Arntz reported receiving grants from the Netherlands Organization for Health Research and Development and Netherlands Foundation for Mental Health, and receiving other grants outside the submitted work from Netherlands’ Organization for Scientific Research (NWO), Netherlands Organization for Health Research and Development (ZONMW), Stichting Achmea Gezondheidszorg, CZ Fonds, Stichting Volksbond Rotterdam, and Stichting tot Steun VCVGZ; receiving royalties (paid to the university) from Academic Press, American Psychological Association Press, Beltz, Bohn Stafleu van Loghum, Boom Uitgevers, Cambridge University Press, Context Press, Guilford, De Tijdstroom, Oxford University Press, SAGE Publications, Uitgeverij Nieuwezijds, Wiley; providing workshops and lectures on cognitive behavioural therapy (CBT), imagery rescripting, personality disorders, schema therapy, and small-scale research in clinical practice (remuneration to the university) for the BABCP, Bulgarian Association for CBT, Clinical Academic Group for Psychotherapy Denmark, Danish Competence Centre for Psychotherapy, EABCT, ECNP, ESSPD, Estonian CBT Association, German Psychosomatic Congress, GGZ InGeest, Greek CBT Association, ICCP, Institut für Schematherapie Frankfurt, ISC International, ISSPD, ISST, Jellinek, Kenniscentrum Persoonlijkheidsstoornissen, Leiden University Medical Center, Lemion, Moroccan Association of CBT, Norwegian Psychological Association, Parnassia/PsyQ, Polish Association for Cognitive and Behavioural Therapies, Portuguese Association of Behaviour Therapy, Psyflix, SCEM, Scuole APC-SPC-SICC-IGB-AIPC, Tunisian Association of CBT, Turkish Association for Cognitive & Behavioural Psychotherapies, Ukraine Association for CBT, Ukraine Institute for CBT, University of Bordeaux, VGCT, VST, WCBCT; supervising research at the mental health institute PsyQ (remuneration to the University of Amsterdam); and being chair of the board of the PDO foundation, North Holland postgraduate training institute (unpaid).

Dr. Assmann provided workshops on schema therapy and imagery rescripting (Institut for schema therapy Hamburg, Ausbildungsinstitut für Verhaltenstherapie und Verhaltensmedizin Hannover, Justus-Liebig-University Gießen, Förderverein für interdisziplinäre Sucht-und Drogenforschung), received personal fees from supervision in schema therapy and received grants from Lübeck University and from addisca gGmbH for an observational study (both outside the submitted work).

PD Dr. Fassbinder reported receiving grants for the PROBPD study from the Else Kröner-Fresenius-Stiftung and the University of Lübeck, and grants outside the submitted work from Addisca GmbH; receiving royalties from Beltz Verlag and Elsevier Books; receiving personal fees from supervision in schema therapy and group schema therapy and from workshops and presentations on CBT, imagery rescripting, personality

disorders, schema therapy, and behavioural activation for Ausbildungsinstitut für Verhaltenstherapie und Verhaltensmedizin Hannover, Arbeitsgemeinschaft Wissenschaftliche Psychotherapie Berlin, AWKV gGmbH, Ausbildungs- und Weiterbildungseinrichtung für klinische Verhaltenstherapie Kassel, Deutscher Psychotherapie Kongress, Deutsche Gesellschaft für Entspannungsmedizin, the DGPPN, GAIA AG, IPAM Marburg, IFT-Nord Institut für Therapie- und Gesundheitsforschung gemeinnützige GmbH Kiel, IPP Halle, Institut für Schematherapie Frankfurt, Institut für Schematherapie Hamburg, Institut für Schematherapie Köln, Institut für Schematherapie Berlin, Online Akademie für Psychotherapie, Oberberg Kliniken, Psychosomatisches Zentrum Eggenburg GmbH, Sophien- und Hufeland-Klinikum gGmbH Weimar, Universität Zürich, and the WCBC; and being co-chair of the Deutscher Fachverband für Verhaltenstherapie eV (unpaid).

Dr. Herzog reported that he has no conflict of interest to declare.

Prof. Jauch-Chara reported that she has no conflict of interest to declare.

Dr. Kaiser reported that he has no conflict of interest to declare.

Prof. Klein received funding for clinical trials (German Federal Ministry of Health, Servier), payments for presentations on psychological internet interventions (GAIA, Oberberg, Servier, Stillachhaus), consulting fees from developers and distributors of psychological internet interventions (all about me, Boehringer, Ethypharm, GAIA, medac, sympatient) payments for workshops and books (Beltz, Elsevier, Hogrefe and Springer) on psychiatry, psychosomatics and psychotherapy. He is past president of the CBASP network (now DsG-CBASP) and serves as vice chairman of the chapters "Digital Psychiatry" and "Psychotherapy" of the German Psychiatric Association (DGPPN).

Dr. Schaich received intramural from University of Lübeck (Habilitationförderung für Wissenschaftlerinnen), personal fees from supervision in schema therapy and imagery rescripting, payments for manuscripts (Thieme) and workshops (ISST Hamburg DZPG Marburg, IFKV Bad Dürkheim, DPKBerlin, DeGPT) on imagery rescripting and schema therapy, support for attending meetings and/or travel from GePs e.V. Hamburger Fellowship and Verbund Norddeutscher Universitäten (Impulse Forschung), funding of open access publication from Lübeck University and grants from addisca gGmbH for an RCT (outside the submitted work).

Acknowledgment

We dedicate this study and manuscript to Ulrich Schweiger. In the Department of Psychiatry and Psychotherapy in Lübeck, Ulrich Schweiger did not only introduce Dialectical Behavioural Therapy and Schema Therapy but numerous other modern methods of psychotherapy. He thus laid the foundation for this trial and supported it in all stages. He was involved in the conceptualisation of the study, supervision as well as analysis and interpretation of the data. Without him, this study would not have been possible. His untimely death in November 2022 left an irreplaceable gap. However, his passion for modern psychotherapy and his commitment to improving treatment for patients with severe mental disorders will continue to serve as an inspiration and guide.

Further, we wish to thank all patients and therapists participating in this trial as well as our research assistants and students for their engagement.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.brat.2025.104899>.

Data availability

Data will be made available on request.

References

- 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. <https://doi.org/10.1521/pedi.17.1.45.24053>
- Arntz, A., & van Genderen, H. (2009). *Schema therapy for borderline personality disorder*. Wiley.
- Assmann, N., Schaich, A., Arntz, A., Wagner, T., Herzog, P., Alvarez-Fischer, D., ... Fassbinder, E. (2024). The effectiveness of dialectical behaviour therapy compared to schema therapy for borderline personality disorder: A randomized clinical trial. *Psychotherapy and Psychosomatics*. <https://doi.org/10.1159/000538404>, 1–1.
- Baranoff, J., Oei, T. P., Cho, S. H., & Kwon, S.-M. (2006). Factor structure and internal consistency of the young schema questionnaire (short form) in Korean and Australian samples. *Journal of Affective Disorders*, *93*(1–3), 133–140. <https://doi.org/10.1016/j.jad.2006.03.003>
- Bénard, C., & Josse, J. (2023). Variable importance for causal forests: Breaking down the heterogeneity of treatment effects. *arXiv preprint arXiv:2308.03369*. <https://doi.org/10.48550/arXiv.2308.03369>
- Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., Stokes, J., Handelsman, L., Medrano, M., & Desmond, D. (2003). Development and validation of a brief screening version of the childhood trauma questionnaire. *Child Abuse & Neglect*, *27*(2), 169–190. [https://doi.org/10.1016/S0145-2134\(02\)00541-0](https://doi.org/10.1016/S0145-2134(02)00541-0)
- Bohus, M., Stoffers-Winterling, J., Sharp, C., Krause-Utz, A., Schmahl, C., & Lieb, K. (2021). Borderline personality disorder. *The Lancet*, *398*(10310), 1528–1540. [https://doi.org/10.1016/S0140-6736\(21\)00476-1](https://doi.org/10.1016/S0140-6736(21)00476-1)
- Bone, C., Delgado, J., & Barkham, M. (2021). A systematic review and meta-analysis of the good-enough level (GEL) literature. *Journal of Counseling Psychology*, *68*(2), 219. <https://doi.org/10.1037/cou0000521>
- Chekroud, A. M., Bondar, J., Delgado, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154–170. <https://doi.org/10.1002/wps.20882>
- Chekroud, A. M., Hawrilenko, M., Loho, H., Bondar, J., Gueorguieva, R., Hasan, A., Kambeitz, J., Corlett, P. R., Koutsouleris, N., Krumholz, H. M., Krystal, J. H., & Paulus, M. (2024). Illusory generalizability of clinical prediction models. *Science*, *383*(6679), 164–167. <https://doi.org/10.1126/science.adg8538>
- Cohen, Z. D., Delgado, J., & DeRubeis, R. J. (2021). Personalized treatment approaches. In W. L. M. Barkham, & L. G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behaviour change: 50th anniversary edition* (7th ed., pp. 673–703). John Wiley & Sons.
- D'Aurizio, G., Di Stefano, R., Succi, V., Rossi, A., Barlattani, T., Pacitti, F., & Rossi, R. (2023). The role of emotional instability in borderline personality disorder: A systematic review. *Annals of General Psychiatry*, *22*(1), 9. <https://doi.org/10.1186/s12991-023-00439-0>
- Derogatis, L. R., & Melisaratos, N. (1983). The brief symptom inventory: An introductory report. *Psychological Medicine*, *13*(3), 595–605. <https://doi.org/10.1017/S0033291700048017>
- DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The personalized advantage index: Translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS One*, *9*(1), Article e83875. <https://doi.org/10.1371/journal.pone.0083875>
- Fassbinder, E., Assmann, N., Schaich, A., Heinecke, K., Wagner, T., Sipos, V., ... Schweiger, U. (2018). PRO* BPD: Effectiveness of outpatient treatment programs for borderline personality disorder: A comparison of schema therapy and dialectical behaviour therapy: Study protocol for a randomized trial. *BMC Psychiatry*, *18*(1), 1–17. <https://doi.org/10.1186/s12888-018-1905-6>
- Fassbinder, E., Schweiger, U., Martius, D., Brand-de Wilde, O., & Arntz, A. (2016). Emotion regulation in schema therapy and dialectical behaviour therapy [methods]. *Frontiers in Psychology*, *7*. <https://doi.org/10.3389/fpsyg.2016.01373>
- Friedl, N., Krieger, T., Chevreul, K., Hazo, J. B., Holtzmann, J., Hoogendoorn, M., Kleiboer, A., Mathiasen, K., Urech, A., Riper, H., & Berger, T. (2020). Using the personalized advantage index for individual treatment allocation to blended treatment or treatment as usual for depression in secondary care. *Journal of Clinical Medicine*, *9*(2). <https://doi.org/10.3390/jcm9020490>
- 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*(2), 136–141. <https://doi.org/10.1016/j.paid.2010.03.023>
- Herzog, P., Feldmann, M., Voderholzer, U., Gärtner, T., Armbrust, M., Rauh, E., Doerr, R., Rief, W., & Brakemeier, E.-L. (2020). Drawing the borderline: Predicting treatment outcomes in patients with borderline personality disorder. *Behaviour Research and Therapy*, *133*, Article 103692. <https://doi.org/10.1016/j.brat.2020.103692>
- Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose–effect relationship in psychotherapy. *American Psychologist*, *41*(2), 159.
- Iliakis, E. A., Sonley, A. K. I., Ilagan, G. S., & Choi-Kain, L. W. (2019). Treatment of borderline personality disorder: Is supply adequate to meet public health needs? *Psychiatric Services*, *70*(9), 772–781. <https://doi.org/10.1176/appi.ps.201900073>
- Kaiser, T., & Herzog, P. (2023). Is personalized treatment selection a promising avenue in bpd research? A meta-regression estimating treatment effect heterogeneity in RCTs of BPD. *Journal of Consulting and Clinical Psychology*. <https://doi.org/10.1037/ccp000803>
- Keefe, J. R., Kim, T. T., DeRubeis, R. J., Streiner, D. L., Links, P. S., & McMain, S. F. (2021). Treatment selection in borderline personality disorder between dialectical

- behaviour therapy and psychodynamic psychiatric management. *Psychological Medicine*, 51(11), 1829–1837. <https://doi.org/10.1017/S0033291720000550>
- Kröger, C., Vonau, M., Kliem, S., Roepke, S., Kosfelder, J., & Arntz, A. (2013). Psychometric properties of the German version of the borderline personality disorder severity index-version IV. *Psychopathology*, 46(6), 396–403. <https://doi.org/10.1159/000345404>
- Linehan, M. (2015a). *DBT skills handouts and worksheets* (2 ed.). The Guilford Press.
- Linehan, M. (2015b). *DBT skills training manual* (2 ed.). The Guilford Press.
- Lorenzo-Luaces, L., Peipert, A., De Jesús Romero, R., Rutter, L. A., & Rodríguez-Quintana, N. (2021). Personalized medicine and cognitive behavioural therapies for depression: Small effects, big problems, and bigger data. *International Journal of Cognitive Therapy*, 14(1), 59–85. <https://doi.org/10.1007/s41811-020-00094-3>
- Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clinical Psychological Science*, 7(3), 445–461. <https://doi.org/10.1177/2167702618815466>
- Meehan, A. J., Lewis, S. J., Fazel, S., Fusar-Poli, P., Steyerberg, E. W., Stahl, D., & Danese, A. (2022). Clinical prediction models in psychiatry: A systematic review of two decades of progress and challenges. *Molecular Psychiatry*, 27(6), 2700–2708. <https://doi.org/10.1038/s41380-022-01528-4>
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The work and social adjustment scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*, 180(5), 461–464. <https://doi.org/10.1192/bjp.180.5.461>
- Neacsiu, A. D., Rizvi, S. L., Vitaliano, P. P., Lynch, T. R., & Linehan, M. M. (2010). The dialectical behaviour therapy ways of coping checklist: Development and psychometric properties. *Journal of Clinical Psychology*, 66(6), 563–582. <https://doi.org/10.1002/jclp.20685>
- Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.
- Sahin, Z., Vinnars, B., Gorman, B. S., Wilczek, A., Åsberg, M., & Barber, J. P. (2018). Clinical severity as a moderator of outcome in psychodynamic and dialectical behaviour therapies for borderline personality disorder. *Personal Disord*, 9(5), 437–446. <https://doi.org/10.1037/per0000276>
- Siegmund, A., R ath, D., Finck, C., Meves, A. M., Stoy, M., & Str ohle, A. (2011). Young-Schemafragebogen: Psychometrische Eigenschaften einer deutschen Kurzform. *Psychotherapies*, 16(2), 207–212.
- Socci, V., Festucci, F., Barlattani, T., Salfi, F., D'Aurizio, G., Rossi, R., Ferrara, M., Rossi, A., Pacitti, F., & Tempesta, D. (2024). Sleep quality and emotional reactivity in patients with borderline personality disorder. *Frontiers in Sleep*, 3, Article 1394979.
- Stoffers-Winterling, J. M., Storeb , O. J., Kongerslev, M. T., Faltinsen, E., Todorovac, A., Sedoc J rgensen, M., Sales, C. P., Edemann Callesen, H., Pereira Ribeiro, J., V llm, B. A., Lieb, K., & Simonsen, E. (2022). Psychotherapies for borderline personality disorder: A focused systematic review and meta-analysis. *The British Journal of Psychiatry*, 221(3), 538–552. <https://doi.org/10.1192/bjp.2021.204>
- Storeb , O. J., Stoffers-Winterling, J. M., V llm, B. A., Kongerslev, M. T., Mattivi, J. T., J rgensen, M. S., Faltinsen, E., Todorovac, A., Sales, C. P., Callesen, H. E., et al. (2020). Psychological therapies for people with borderline personality disorder. *Cochrane Database of Systematic Reviews*, (5) <https://doi.org/10.1002/14651858.CD012955.pub2>
- Thandi, G., Fear, N. T., & Chalder, T. (2017). A comparison of the work and social adjustment scale (WSAS) across different patient populations using rasch analysis and exploratory factor analysis. *Journal of Psychosomatic Research*, 92, 45–48. <https://doi.org/10.1016/j.jpsychores.2016.11.009>
- Tomko, R. L., Trull, T. J., Wood, P. K., & Sher, K. J. (2014). Characteristics of borderline personality disorder in a community sample: Comorbidity, treatment utilization, and general functioning. *Journal of Personality Disorders*, 28(5), 734–750. <https://doi.org/10.1521/pedi.2012.26.093>
- Twala, B. E. T. H., Jones, M. C., & Hand, D. J. (2008). Good methods for coping with missing data in decision trees. *Pattern Recognition Letters*, 29(7), 950–956. <https://doi.org/10.1016/j.patrec.2008.01.010>
- van Bronswijk, S. C., Bruijnicks, S. J. E., Lorenzo-Luaces, L., Derubeis, R. J., Lemmens, L., Peeters, F., & Huibers, M. J. H. (2021). Cross-trial prediction in psychotherapy: External validation of the personalized advantage index using machine learning in two Dutch randomized trials comparing CBT versus IPT for depression. *Psychotherapy Research*, 31(1), 78–91. <https://doi.org/10.1080/10503307.2020.1823029>
- van Bronswijk, S. C., DeRubeis, R. J., Lemmens, L., Peeters, F., Keefe, J. R., Cohen, Z. D., & Huibers, M. J. H. (2021). Precision medicine for long-term depression outcomes using the personalized advantage index approach: Cognitive therapy or interpersonal psychotherapy? *Psychol Med*, 51(2), 279–289. <https://doi.org/10.1017/s0033291719003192>
- Varadhan, R., & Seeger, J. (2013). Estimation and reporting of heterogeneity of treatment effects. In P. Velentgas, N. Dreyer, P. Nourjah, S. Smith, & M. Torchia (Eds.), *Developing a protocol for observational comparative effectiveness research: A user's guide*. Agency for Healthcare Research and Quality (US). <https://www.ncbi.nlm.nih.gov/books/NBK126188/>.
- Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228–1242. <https://doi.org/10.1080/01621459.2017.1319839>
- Wibbelink, C. J. M., Arntz, A., Grasman, R., Sinnaeve, R., Boog, M., Bremer, O. M. C., ... Kamphuis, J. H. (2022). Towards optimal treatment selection for borderline personality disorder patients (BOOTS): A study protocol for a multicenter randomized clinical trial comparing schema therapy and dialectical behaviour therapy. *BMC Psychiatry*, 22(1), 89. <https://doi.org/10.1186/s12888-021-03670-9>
- Wingenfeld, K., Spitzer, C., Mensebach, C., Grabe, H., Hill, A., Gast, U., Schlosser, N., H pp, H., Beblo, T., & Driessen, M. (2010). Childhood trauma questionnaire (CTQ): Erste befunde zu den psychometrischen kennwerten. *Psychother psych med. Psychotherapie Psychosomatik Medizinische Psychologie*, 60, 442–450. <https://doi.org/10.1055/s-0030-1247564>
- Woodbridge, J., Reis, S., Townsend, M. L., Hobby, L., & Grenyer, B. F. (2021). Searching in the dark: Shining a light on some predictors of non-response to psychotherapy for borderline personality disorder. *PLoS One*, 16(7), Article e0255055. <https://doi.org/10.1371/journal.pone.0255055>
- Young, J., Klosko, S., & Weishaar, M. (2003). *Schema therapy - A practitioner's guide*. The Guilford Press.
- Zahra, D., Qureshi, A., Henley, W., Taylor, R., Quinn, C., Pooler, J., Hardy, G., Newbold, A., & Byng, R. (2014). The work and social adjustment scale: Reliability, sensitivity and value. *International Journal of Psychiatry in Clinical Practice*, 18(2), 131–138. <https://doi.org/10.3109/13651501.2014.894072>