

## Invited Review

\*Current address: Trubendorffer Addiction Care, Amsterdam, The Netherlands

†Current address: Department of Clinical Child and Family Studies, Utrecht University, The Netherlands

**Cite this article:** Arntz A, Mensink K, Cox WR, Verhoef REJ, van Emmerik AAP, Rameckers SA, Badenbach T, Grasman RPPP (2023). Dropout from psychological treatment for borderline personality disorder: a multilevel survival meta-analysis. *Psychological Medicine* **53**, 668–686. <https://doi.org/10.1017/S0033291722003634>

Received: 29 April 2018

Revised: 28 October 2022

Accepted: 2 November 2022

First published online: 1 December 2022

**Key words:**






Borderline personality disorder; dropout; meta-analysis; psychotherapy; treatment retention

**Author for correspondence:**

Arnoud Arntz,

E-mail: [A.R.Arntz@uva.nl](mailto:A.R.Arntz@uva.nl)

# Dropout from psychological treatment for borderline personality disorder: a multilevel survival meta-analysis

Arnoud Arntz<sup>1</sup> , Kyra Mensink<sup>1,\*</sup>, Wouter R. Cox<sup>1</sup> , Rogier E. J. Verhoef<sup>1,†</sup> , Arnold A. P. van Emmerik<sup>1</sup>, Sophie A. Rameckers<sup>1</sup> , Theresa Badenbach<sup>1</sup> and Raoul P. P. Grasman<sup>2</sup> 

<sup>1</sup>Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands and <sup>2</sup>Department of Psychological Methods, University of Amsterdam, Amsterdam, The Netherlands

**Abstract**

**Background.** Dropout from psychotherapy for borderline personality disorder (BPD) is a notorious problem. We investigated whether treatment, treatment format, treatment setting, substance use exclusion criteria, proportion males, mean age, country, and other variables influenced dropout.

**Methods.** From Pubmed, Embase, Cochrane, Psycinfo and other sources, 111 studies (159 treatment arms,  $N = 9100$ ) of psychotherapy for non-forensic adult patients with BPD were included. Dropout per quarter during one year of treatment was analyzed on participant level with multilevel survival analysis, to deal with multiple predictors, nonconstant dropout chance over time, and censored data. Multiple imputation was used to estimate quarter of drop-out if unreported. Sensitivity analyses were done by excluding DBT-arms with deviating push-out rules.

**Results.** Dropout was highest in the first quarter of treatment. Schema therapy had the lowest dropout overall, and mentalization-based treatment in the first two quarters. Community treatment by experts had the highest dropout. Moreover, individual therapy had lowest dropout, group therapy highest, with combined formats in-between. Other variables such as age or substance-use exclusion criteria were not associated with dropout.

**Conclusion.** The findings do not support claims that all treatments are equal, and indicate that efforts to reduce dropout should focus on early stages of treatment and on group treatment.

**Introduction**

Psychological treatment of borderline personality disorder (BPD) is usually considered as being highly complex, with treatment discontinuation before a significant improvement has been reached as one of the most challenging problems. Early reports documented treatment dropout rates higher than 50% within 6 months of traditional psychotherapy (Gunderson et al., 1989; Skodol, Buckley, & Charles, 1983; Waldinger & Gunderson, 1984). The high dropout rate among patients with BPD is generally viewed as related to their complex psychopathology, including impulsivity, anger problems, and difficulties in establishing trusting relationships. High dropout risk is problematic given the high levels of disfunctioning, high suicide risk, and high societal costs associated with BPD (Lieb, Zanarini, Schmahl, Linehan, & Bohus, 2004; van Asselt, Dirksen, Arntz, & Severens, 2007; Wagner et al., 2022). It is also demotivating for therapists, who have to invest a lot in the treatment of difficult patients, and are confronted with many patients that end treatment prematurely. Moreover, premature treatment discontinuation constitutes a threat for the cost-effectiveness of the intensive and costly interventions for BPD, that are often available only for a limited number of BPD-patients. Understandably, one of the aims of specialized psychotherapies such as dialectical behavior therapy (DBT), transference-focused psychotherapy (TFP), mentalization-based treatment (MBT) and schema therapy (ST) (the ‘big-four’), that were developed since the late eighties of the previous century, was therefore the reduction of treatment dropout.

Attempts to create comprehensive quantitative overviews of dropout rates of different treatments of BPD (e.g. Barnicot, Katsakou, Marougka, & Priebe, 2011) have been limited by a number of factors. First, treatments and studies of treatments vary widely in the time period they cover, making a simple meta-analytic approach of risk of dropout without taking the duration of treatment into account senseless. Second, the standard meta-analytic approach to only include randomized clinical trials (RCTs) severely limits the comparison possibilities between treatment approaches, as many have not been directly compared. Third, many studies that have been published are based on a non-RCT design, and disregarding them, while

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

methodologically sound from one point of view, seriously limits the comprehensiveness in terms of the available evidence and the type of treatments that can be studied. Fourth, BPD treatments have been studied in different settings (inpatient, day treatment, outpatient) and in different formats (individual, group, and combined) which raises the question how these variables are related to treatment retention. Fifth, meta-regression offers the opportunity for a multivariate analysis of putative factors predicting treatment dropout.

Various factors on different levels have been suggested to relate to treatment dropout. First, patient characteristics, such as male gender and younger age (Arntz, Stupar-Rutenfrans, Bloo, van Dyck, & Spinhoven, 2015; Crawford et al., 2009; Edlund et al., 2002; McMurrin, Huband, & Overton, 2010). Second, treatment characteristics, such as treatment model, format (i.e. group, individual, and combined group-individual), and setting (outpatient, day-treatment and inpatient). As to treatment models, for the 'big-four' treatments (DBT, ST, MBT, TFP) superior treatment retention has been claimed. As to format, group therapy tends to have higher dropout rates than individual therapy, which might relate to practical issues (less agenda flexibility) as well as psychological factors (e.g. groups might be threatening for patients) (MacNair & Corazzini, 1994; Yalom, 1966). The present authors are not aware of any claims as to setting, but it seems an important factor to investigate. Third, there are socio-economic factors that might be related to treatment retention (e.g. the general difference between European public health care *v.* the limited availability of mental health care for poor people in the US might influence dropout from studies; Edlund et al., 2002; Gaglia, Essletzbichler, Barnicot, Bhatti, & Priebe, 2013; McMain et al., 2009; Priebe et al., 2012).<sup>†</sup> Lastly, study design factors might be related to dropout. RCTs have a higher methodological status than open trials, and perhaps the latter report lower dropout rates than RCTs, suggesting biases in reporting dropout rates in open trials. Moreover, studies use different exclusion criteria and the in- *v.* exclusion of different levels of substance use disorders might be especially important for treatment retention, with more lenient criteria perhaps being associated with higher dropout.<sup>2</sup>

The aim of the present meta-analysis was therefore to study dropout from psychological treatments for BPD, taking into account as much data as possible, including data from non-RCTs, investigating various factors that might influence treatment dropout. We chose survival analysis as this approach is suitable for analyzing dropout data from observational periods of varying lengths, while allowing the use of multiple covariates, and we chose a multilevel approach so that studies could be combined in one analysis and different predictors could be tested. More specifically, we aimed to address the following questions:

- (1) How do different psychological treatments compare in terms of treatment retention?
- (2) Is setting (inpatient, day-treatment, outpatient) associated with treatment retention?
- (3) Is treatment format (individual, group, combined individual-group) associated with treatment retention?
- (4) Is gender composition of the study sample associated with dropout?
- (5) Is the sample's mean age associated with treatment retention?

<sup>†</sup>The notes appear after the main text.

- (6) Is study quality associated with treatment retention?
- (7) Are methodological aspects of studies, like trial design (RCT, non-RCT, open trial), and distinction *v.* non-distinction of treatment and study dropouts, associated with dropout?
- (8) Is country where the study was conducted related to treatment retention?
- (9) Does type of exclusion of substance use disorders relate to treatment retention?
- (10) Has treatment retention improved over the years that we study BPD treatments?

Initially, we also aimed to investigate the association of suicidality, comorbidity, educational level, and unemployment with treatment retention. However, too many studies did not present (suitable) data on these variables, so that we had to disregard this aim.

## Methods

### Guidelines for meta analyses

For this meta-analysis, we followed the PRISMA Guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009) and the American Psychological Association MARS Guidelines (APA, 2008). However, we did not preregister the meta-analysis.

### Identification and selection of studies

A database search was done in Pubmed, Embase, Cochrane, and Psycinfo, on 21 June 2013; and was repeated on 4 February 2015 and January 31 and 15 June 2022. Appendix A (appendices can be found in the supplementary material) provides the search terms. In total, 2997 records were retrieved. In addition, reference lists of reviews, meta-analyses, and other manuscripts were checked, and one submitted ms. was obtained with permission of the authors, which yielded another 880 records. The following criteria were used by three independent judges (from PC, AvE, AA, RV) to select studies for inclusion in the meta-analysis. In case of disagreement, a decision was reached through consensus.

#### Inclusion criteria:

- (1) a study of psychological treatment for BPD: RCT, open trial, case series, cohort study.
- (2) adult patients (age  $\geq 18$ ) with primary diagnosis of BPD according to DSM-III, DSM-III-R or DSM-IV (-Tr) criteria.

#### Exclusion criteria:

- (1) 'double diagnoses': study focuses exclusively on a specific combination of two diagnoses, e.g. BPD and eating disorder; BPD and opioid dependence. The reason was that such studies have biased sampling from the BPD population; and that treatments are modified to the double diagnosis.
- (2) single case studies: in contrast to consecutive case series studies, there is little guarantee that reporting is not biased (i.e. the case was only reported if a success).
- (3) mixed PD – samples; unless separate statistics on the BPD subsample are given. A tolerance of 10% was allowed: at least 90% had to meet full BPD diagnosis. Thus a study was excluded if more than 10% did not meet full BPD-diagnosis (in case of mixed samples). Authors of

studies published after 2000 were asked for statistics of the BPD-subsample.

- (4) treatments that consist of subsets of techniques or modules which are clearly not intended to be complete treatments: i.e. incomplete parts of treatments. However, tests of protocols intended to be complete treatments but missing usual ingredients of the protocol are included and are marked as such (e.g. 'reduced DBT', also labeled 'DBT-min', for DBT treatments without a specific ingredient).
- (5) treatment modules that are explicitly additions to treatments, i.e. psycho-education; courses; specific skills training, as these are not complete treatments (for example: STEPPS; psycho-education). One study tested a complete STEPPS treatment by adding protocolized individual sessions to group, and hence was included.
- (6) forensic populations, as these require specific forms of treatment, and effects and dropout are difficult to generalize outside the forensic context.
- (7) no dropout data reported, as without data on dropout, the study could not be included in the analysis (in case there were unclarity, authors were contacted if published after 2000).

Note that studies might have treatments that meet selection criteria as well as treatments that do not meet them. E.g. some studies on training add-on's had a TAU comparison condition. In such cases the TAU condition data was included, if it passed the selection criteria.

If an English abstract was available and survived the initial screening, language was not an eligibility criterion, and non-English, non-Dutch or non-German papers were translated into English for further scrutiny. [Figure 1](#) shows the flowchart of the study selection. Appendix B gives an overview of the characteristics of the studies included, appendix C their references. As a result of the criterion that diagnoses should be based on DSM-III or later editions, the earliest included study was published in 1990.

### Treatment definition

We initially classified treatments into 17 categories (*n*'s refer to the total sample sizes): DBT (*n* = 3916); reduced DBT (*n* = 716); ST (*n* = 539); MBT (*n* = 448); TFP (*n* = 163); cognitive-behavioral therapy (CBT; *n* = 258); psychodynamic therapy (PsyDyn; *n* = 723); cognitive-analytic therapy (CAT; *n* = 61); interpersonal therapy (IPT; *n* = 60); client-centered therapy (CCT; *n* = 44); structural clinical management (SCM; *n* = 63); general psychiatric management (GPM; *n* = 90); therapeutic community (TherCom; *n* = 78); community treatment by experts (CTBE; *n* = 101); treatment-as-usual (TAU; *n* = 728); Dynamic Deconstructive Psychotherapy (DDP; *n* = 42); and 'mixed' (*n* = 1070), treatments combining different approaches such as CBT with psychodynamic, or DBT with TFP, as well as treatment arms consisting of individually allocated specialized treatments. Following the developer of DBT (Linehan et al., 2015), a DBT-treatment was classified as full DBT if it included four standard DBT components [group skills training, individual coaching, outside session telephone crisis support, therapist consultation (in case of inpatient DBT, outside session crisis support was assumed)]. If one component was not present, the treatment was classified as reduced DBT. Some studies deliberately tested reduced DBT.

We next reduced the number of treatment categories by collapsing specified treatment categories with *n* < 100 together with the mixed category into a 'specified others' category (*N* = 1432). The 'specified others' category thus consisted of psychotherapies that had at least some adjustment to BPD, but individually had a *n* < 100. CTBE was distinguished from TAU as it is viewed as an optimized variant of TAU, and thus constitutes a more stringent comparison condition than TAU (e.g. Linehan et al., 2006).

### Coding of methodological quality (risk of bias) of studies

Included studies were assessed for risk of bias by evaluating nine design criteria. These criteria were based on Cuijpers, van Straten, Bohlmeijer, Hollon, and Andersson (2010) and slightly modified to accommodate the BPD treatment outcome literature, including non-RCTs. Specifically, we evaluated for each treatment arm whether: (1) the BPD diagnosis was made using semi-structured diagnostic interviews such as the SCID-II (First, Gibbon, Spitzer, Williams, & Benjamin, 1997) [0 = no or unknown, 1 = yes, but with inadequate or unknown inter-rater reliability (IRR), 2 = yes, with adequate IRR]; (2) a treatment manual was used (0 = no or unknown, 1 = yes, but treatment manual is unpublished, 2 = yes, with published treatment manual); (3) therapists were trained either specifically for the study or in a general training (0 = no or unknown, 1 = no or unknown, but therapists are clearly experts, 2 = yes); (4) treatment integrity was checked (0 = no or unknown, 1 = yes, by supervision, 2 = yes by independent raters); (5) the study was randomized (0 = no or unknown, 1 = yes, but randomization was partly violated, 2 = yes); (6) if applicable, whether randomization was independent and adequately concealed (0 = no or unknown, 1 = either independent or adequately concealed, 2 = both independent and adequately concealed); (7) if applicable, whether assessment interviews were conducted by independent or blind assessors (0 = no or unknown, 1 = yes, independent but not blind, 2 = yes, blind); and (8) whether and how dropout was reported (0 = no, 1 = yes, but no distinction between types of dropout, 2 = yes, with adequate distinction between types of dropout). If a study investigated DBT, ST, TFP or MBT, criterion 2 was coded as 2 (i.e. using a published treatment manual).

Following calibration exercises on a subset of included studies, the remaining studies were independently rated on these criteria by different pairs from a total of five coders. Interrater agreement of the initial ratings was assessed using two-way mixed, absolute agreement, average-measures intra-class correlations (ICCs). The ICC per item ranged from 0.86 to 0.97, with a mean and median of 0.90. The ratings were summarized into a mean score (range 0–2; internal consistency 0.69, ICC 0.95) for each study's arm (see Appendix B). Note that treatment arms per study could get different ratings, e.g. in RCTs that compared a manualized therapy to a non-manualized TAU. The study arm's quality score was used as covariate in the analyses.

### Coding of dropout and other characteristics

Two raters (from KM, WC, SR, TB) independently coded treatment characteristics and dropout per quarter. In case of disagreement, the issue was resolved through discussion. In several cases authors were contacted by email to clarify issues. Dropout per quarter was derived from the article's text (methods, results, discussion sections), flow diagram, or survival curve. When quarter of dropout was not clear from the report, authors of studies

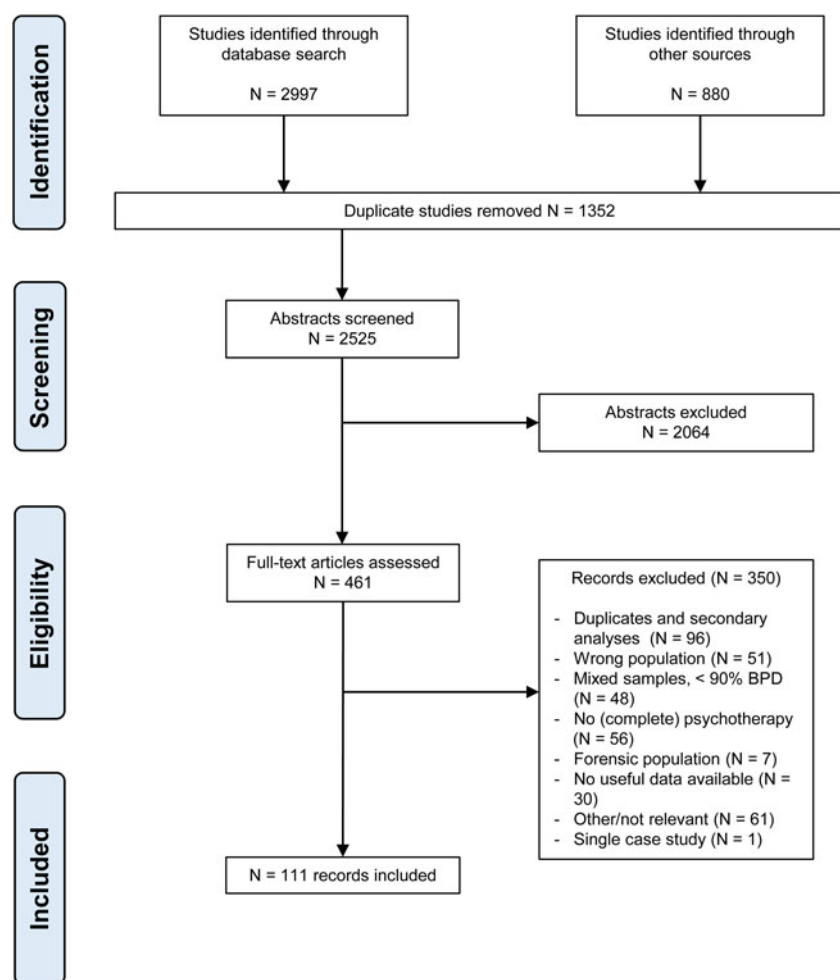


Fig. 1. Flowchart of study selection.

published after 2000 were emailed but not all were able or willing to provide details (Appendix D). Most studies did not distinguish between ‘dropouts’ and ‘pushouts’, the first being based on patients’ decisions, the latter on therapists’ decisions or protocol rule. We therefore couldn’t distinguish between these types of dropout. Moreover, most studies did not report details about reasons for dropout, thus for these studies we took all treatment dropouts into account. However, when reasons clearly not related to treatment were reported, these cases were not considered as treatment dropout. Type of exclusion of substance use was coded as no exclusion, clinical detox needed excluded, substance dependence excluded, substance abuse excluded, and unclear. Country of the study was classified in four categories: Europe, USA, Australia/Canada/New Zealand, and emerging (China/Iran/Mexico).

### Statistical analysis

A multilevel survival analysis was used to analyze treatment discontinuation over time, using a random effects approach by adding study as random factor to the model. This method allows for testing of multiple predictors whilst controlling for other variables, and for distinguishing between individual cases and studies on separate levels. Moreover, the method can handle censored data, which is necessary as study length varies, and cases ‘disappear’ when they stop treatment prematurely. Study served as

random factor in the analysis, representing random variation between studies. This approach models the included studies as a random sample from a population of studies (hence, from the BPD population, treatment centers, therapists) and allows generalization of the findings (e.g. Hedges & Vevea, 1998). The statistical test of this random factor is reported in the results section. Although multilevel continuous time survival methods have been developed, they are not currently accessible through widely available software. Furthermore, they tend to suffer from numerical instability (Eager & Roy, 2017), and require exact dropout times for each patient which is not always available. Therefore we used multilevel survival analysis with quarter as time period. As survival chance might differ per quarter, quarter was entered as factor in the analysis. Interval censored survival analysis of Generalized Linear Mixed Models of SPSS version 28 was used (International Business Machines, 2021), which uses a binomial distribution with a complementary log-log link. We used Restricted Maximum Likelihood estimation with the Satterthwaite method for defining degrees of freedom in the  $t$  tests of the fixed effects coefficients (Luke, 2017).

For the analyses the numbers of cases (dis)continuing treatment in the specific quarter were reconstructed on the basis of the reports of the studies. In case the study period was not equal to a complete number of quarters (e.g. study length was 2.5 quarters), we estimated the dropout for the last quarter of the study, assuming constant survival chance for that quarter.



For 34 studies (45 treatment arms) dropout was reported, but not in enough detail to reconstruct dropout per quarter [for seven of these studies only a part of the dropout development had to be estimated (in 11 arms)]. For these we estimated dropout per quarter based on the general development of treatment dropout based on studies reporting dropout per quarter. Note that the reported total dropout formed the basis of each estimation, i.e. the estimated dropouts per quarter sum up to the total number of dropouts reported. Available data showed clear evidence for a smaller retention rate in the first quarter than in later quarters, with a gradual increase in retention in later quarters. This time-dependence was best described by a logistic time model and this model was therefore used to estimate dropout per quarter which was subsequently used in the multiple imputation (MI) model (Appendix E). Appendix D gives an overview of treatment dropout per quarter per study arm, including the estimated dropout per quarter.

Because the estimation of dropout per quarter for studies not reporting this detail underestimates variance in retention rate per quarter, we used a MI strategy to deal with this. Twenty datasets were created with, for studies with incomplete details, random varying dropout numbers based on a binomial distribution of the numbers showing a particular dropout pattern, with the numbers of treatment completers as well as the total dropout number held constant. Appendix F provides details of this MI procedure. For studies not reporting mean age and/or proportion male participants, these variables were also estimated by MI. For model selection, the resulting 20 datasets were analyzed in one GLMM survival analysis, with for dropout-pattern of each set a weight of  $n/20$ , with  $n$  = number of participants of the pattern, so that the total sample size of the 20 combined sets was equal to the observed sample size.<sup>3</sup>

For model selection, we first entered all covariates as main effects in the fixed part and then applied stepwise deletion of covariates with significance level  $>0.05$ . For the remaining covariates it was next tested whether the interaction with quarter was significant. Both the full model with all main effects and the final models (with significant main effects, and with significant interactions with quarter) are presented. Note that because of model selection, the  $p$  values are indicative. For the final models, significant effects of categorical variables were further tested by deviation contrasts, which test the difference of a category with the overall mean. This way, the number of comparisons is not too large (e.g. with 10 treatment models, there are 10 deviation contrasts compared to 45 pairwise comparisons). Deviation contrasts test what is generally most relevant when large numbers of conditions are investigated, i.e. which conditions differ from the general picture. For the final step of testing the selected model, we used a conventional MI procedure with Rubin's rule for estimating means, their s.e.'s, and the deviation contrasts and their  $t$  tests and  $p$  values (this procedure turned out to be more conservative than the procedure used for model selection, which ensures that not too optimistic levels of significance are reported).

The following variables were initially entered as covariates: quarter, study design (RCT, open trial, nonrandomized controlled), dropout type (treatment dropout *v.* no distinction made between treatment and study dropout), medication policy (prescribed *v.* nonprescribed medication), treatment model, treatment format (individual, group, combined), treatment offered in addition to TAU (yes/no), setting (inpatient, outpatient, day treatment), dropout imputed in case of insufficient details in study report (yes/no), proportion males, mean age, methodological

quality, publication year, exclusion of substance disorders, and country group (Europe, USA, Australia/Canada/New Zealand, emerging).

Survival curves were constructed from the estimated means from the fixed part, controlled for the indicated covariates. The period of investigation was limited to the first year, as very few studies had a study length beyond one year, which led to estimation problems in the statistical analysis of longer periods.

The random effect  $I^2$  of study was derived from the random effect estimate:  $I^2 = \text{random effect} / (1 + \text{random effect}) \times 100\%$ . Egger's test was used to test whether precision was associated with treatment retention by adding  $1/\sqrt{N}$  as covariate to the fixed part of the final model, with  $N$  = sample size of the study arm. Lastly, funnel plots were constructed by plotting residuals of the final analysis against study precision (i.e. the s.e. of the observed treatment retention at quarter  $i$ , with  $\text{s.e.}_i = \sqrt{(p_i(1-p_i)/N_i)}$ , with  $p_i$  = retention proportion in quarter  $i$ , and  $N_i$  = sample size of quarter  $i$ . In case of low dropout,  $<17\%$ , the Agresti–Coul approximation was used by defining  $p_i = (n_i + 2)/(N_i + 4)$  (Agresti & Coull, 1998).

In addition to the prespecified analysis described above, a sensitivity analysis was done excluding DBT treatment arms that used different pushout rules than the DBT protocol prescribes.

### Selected studies

There were 111 studies (159 treatment arms, total  $N = 9100$ ) that met inclusion criteria and reported treatment dropout data. Table 1 gives an overview of these studies; appendix B shows the characteristics per study arm. In short, sample size per arm varied from  $N = 5$  to  $N = 1423$  (mean  $N = 57.2$ , median  $N = 33$ ); 54 studies investigated DBT [ $N = 4632$  ( $N = 3916$  full DBT;  $N = 716$  reduced DBT)]; 11 studies investigated solely (a) reduced form(s) of DBT, two both full and reduced forms of DBT]; 12 ST ( $N = 539$ ); 5 TFP ( $N = 163$ ); 8 MBT ( $N = 448$ ); 8 CBT ( $N = 258$ ); 25 TAU ( $N = 728$ ); 11 psychodynamic psychotherapy ( $N = 723$ ); 3 CAT ( $N = 61$ ); 3 IPT ( $N = 60$ ); 2 CCT ( $N = 44$ ); 1 SCM ( $N = 63$ ); 1 GPM ( $N = 90$ ); 8 mixed approaches, using combinations of models ( $N = 824$ ) or a range of specified therapies (1;  $N = 246$ ); 2 CTBE ( $N = 101$ ); 2 Therapeutic Communities ( $N = 78$ ); 2 DDP ( $N = 42$ ). Study-arms varied in length from one quarter (32,  $N = 3407$ ) up to 3 years (3,  $N = 94$ ), with 32 spanning 2 quarters ( $N = 1648$ ), 10 3 quarters ( $N = 269$ ), 58 one year ( $N = 2251$ ), 27 longer than one year ( $N = 1525$ ). Over all studies, the mean of the study's mean age of patients was 31.02 (s.d. 4.31; median 31.30; range 20.40–40.10), seven studies did not report age (for the analysis, missings were handled by MI). The mean proportion male patients was 0.150 (s.d. 0.130; median 0.138; range 0–0.560), eight studies did not report gender composition (for the analysis, missings were handled by MI). Year of publication varied from 1990 to 2022 (one study was submitted and labeled as 2015), with mean 2011 and median 2011. Most studies investigated outpatient treatment (79% of arms). Combined individual-group therapies were most often investigated (57% of arms). RCTs were the most common (52.8% of arms). Most studies distinguished between treatment and study dropout (80.5% of arms). Most studies used an intent-to-treat approach (57.2% of arms), though a completers analysis only was not uncommon. Seven studies investigated psychotherapy delivered in addition to TAU (5% of arms), and five studies had a prescribed medication policy. The majority of studies (and participants) came from Europe ( $N = 6458$ ), followed by USA (1544), Australia/Canada/

**Table 1.** Number of studies and sample sizes by treatment and study characteristics

	# studies	N	# studies	N	# studies	N
Setting <sup>a</sup>	Outpatient		Day-treatment		Inpatient	
	83	4938	15	1158	17	3004
Format <sup>a</sup>	Individual + Group		Individual		Group	
	74	6595	43	1652	13	853
Study design	Open Trial		RCT		Non-randomized controlled	
	58	5008	44	3397	9	695
Add-on to TAU <sup>a</sup>	Not added to TAU		Added to TAU			
	110	8853	7	247		
Medication	Not prescribed		Prescribed			
	106	8946	5	154		
Dropout type	Treatment dropout distinguished from Study dropout		Study and treatment dropout not distinguished			
	90	7775	21	1325		
Dropout per Quarter <sup>b</sup>	Completely reported		Partially reported		Not reported	
	87	7232	7	483	20	1385
Country group	Europe		USA		Australia/Canada/NewZealand	
	48	4781	27	1089	6	418
					418	341

<sup>a</sup>Note that some studies investigated multiple settings/formats/models. Hence, #studies >111 in these rows.

<sup>b</sup>Note that some studies had complete dropout reports for one arm, but not or partially reported for another arm. Hence, #studies >111 in this row.

New Zealand ( $N=757$ ), and emerging countries ( $N=341$ ) (Table 1). As to substance abuse related exclusion, 62 study-arms did not report exclusion of substance related disorders ( $N=2985$ ), 33 excluded only when a clinical detox was necessary ( $N=3202$ ), 40 excluded substance dependence ( $N=2170$ ), 20 substance abuse ( $N=623$ ), and 4 ( $N=120$ ) were unclear about this. All these variables were used as covariates.

## Results

### All studies included

The upper part of Table 2 presents the results of the fixed part of the initial model with all main effects. The middle part presents the results after stepwise deletion of predictors with  $p > 0.05$ , the lower part results with significant interactions added.

### Initial model

Study as random factor was significant,  $z = 4.825$ ,  $p < 0.001$  ( $I^2 = 100 \times .135/1.135 = 11.9\%$ ). In the full model, quarter was significant, with growing treatment retention over time. Using imputation for number of dropouts per quarter was associated with less retention ( $\beta = -0.171$ ,  $p = 0.032$ ). Treatment category was significant (details: see final model). The trend in format was related to group having lower treatment retention than individual and combined formats.

### Intermediate model

Only the main effects of quarter, dropout per quarter imputation, format, and treatment category survived the backward deletion (Table 2).

### Final model

Of the interactions only that between treatment category and quarter was significant. The lower part of Table 2 shows the final model. The random effect of study was significant,  $z = 5.219$ ,  $p < 0.001$  ( $I^2 = 100 \times .110/1.110 = 9.9\%$ ). Egger's test was significant, with lower precision associated with less treatment retention. However, only 3.7% of the variance was associated with imprecision ( $r^2 = F/(F + df)$ ). Follow-up contrasts are reported in Table 3. Quarter was significant, with the lowest retention chance in the first quarter, and gradual increase in retention in later quarters. (Fig. 2 shows the estimated retention chances for these effects).

Pure or predominantly group treatments had significantly less than average treatment retention. Studies reporting not enough details to infer dropout per quarter were associated with less retention. The significant main effect of treatment model and the treatment model by quarter interaction were related to the following specific effects. ST had a significantly higher treatment retention in all quarters. MBT had a significantly higher treatment retention than average in Quarter 1 and 2, Specified Others in Quarter 1, effects that disappeared in later quarters. CTBE had lower treatment retention in Quarter 1 and 2, CBT in Quarter 1, and reduced DBT in Quarter 3. Figure 3 shows the (cumulative) survival curves per treatment model (3a, left panel) and format (3b, right panel). After one year, the (unweighted) average retention was 57%, with CTBE showing considerably lower (28%) and ST considerably higher treatment retention (78%).

### Funnel plot

Figure 4 presents the funnel plot over all treatments arms and quarters (Appendix G presents funnel plots per treatment arm). Note that each value represents a residual of a specific quarter of a specific arm of a study. There were 23 residuals lying outside

**Table 2.** *F*-tests of predictors of treatment retention of the fixed part of the initial and final models

Model	<i>F</i>	df1	df2	<i>p</i>
Initial model				
Corrected model	7.041	33	176	<0.001
Year of publication	1.940	1	69	0.17
Methodological quality of study's arm	<0.001	1	116	0.99
Trial type	0.135	2	74	0.87
Dropout type distinguished	0.604	1	66	0.44
TAU added to treatment	0.727	1	601	0.39
<b>Dropout per quarter to be imputed*</b>	<b>4.723</b>	<b>1</b>	<b>129</b>	<b>0.032</b>
Setting (inpatient, day treatment, outpatient)	1.485	2	256	0.23
Treatment Format (Individual, Group, Combined)	2.723	2	1397	0.066
Medicine prescribed by study	0.774	1	81	0.38
Substance use exclusion	0.539	4	70	0.71
Country group (Europe, USA, Aus/Can/NZ, emerging)	0.140	3	67	0.94
Mean age of study's arm	0.582	1	486	0.45
Proportion males of study's arm	0.159	1	346	0.69
<b>Treatment category*</b>	<b>5.952</b>	<b>9</b>	<b>955</b>	<b>&lt;0.001</b>
<b>Quarter*</b>	<b>50.305</b>	<b>3</b>	<b>23 546</b>	<b>&lt;0.001</b>
Intermediate model (Main effects after backwards deletion)				
Corrected model	15.281	15	1230	<0.001
<b>Dropout per quarter to be imputed*</b>	<b>4.868</b>	<b>1</b>	<b>128</b>	<b>0.029</b>
<b>Treatment format (Individual, Group, Combined)*</b>	<b>3.607</b>	<b>2</b>	<b>988</b>	<b>0.027</b>
<b>Treatment category*</b>	<b>6.533</b>	<b>9</b>	<b>1655</b>	<b>&lt;0.001</b>
<b>Quarter*</b>	<b>51.505</b>	<b>3</b>	<b>23 564</b>	<b>&lt;0.001</b>
Final main and interaction effects model				
Corrected model	7.308	42	5000	<0.001
<b>Dropout per quarter to be imputed*</b>	<b>4.867</b>	<b>1</b>	<b>125</b>	<b>0.029</b>
<b>Treatment format (Individual, Group, Combined)*</b>	<b>3.954</b>	<b>2</b>	<b>1011</b>	<b>0.019</b>
<b>Treatment category*</b>	<b>5.353</b>	<b>9</b>	<b>2120</b>	<b>&lt;0.001</b>
<b>Quarter*</b>	<b>37.623</b>	<b>3</b>	<b>23 537</b>	<b>&lt;0.001</b>
<b>Treatment category × quarter*</b>	<b>2.849</b>	<b>27</b>	<b>23 537</b>	<b>&lt;0.001</b>
<b>Egger's test (1/√<i>N</i>)*</b>	<b>5.996</b>	<b>1</b>	<b>158</b>	<b>0.015</b>

\*Significant effects are printed in **bold**.

the 95% CI, which is 4.96% of the 463 residuals – thus less than the 5% that can be expected given the 95% CI. Nine of the outliers (to the left) were related to more actual dropouts than predicted by the GLMM survival model: three DBT (Barnicot & Crawford, 2019, Quarter 4; Fitzpatrick, Bailey, & Rizvi, 2020, Quarter 2; Sinnaeve, van den Bosch, Hakkaart-van Roijen, & Vansteelandt, 2018, Quarter 1); two TAU (Soler et al., 2009, Quarter 1; Verheul et al., 2003, Quarter 1); one psychodynamic (Löffler-Stastka, Ponocny-Seliger, Meißel, & Springer-Kremser, 2006, Quarter 1); one CTBE (Doering et al., 2010, Quarter 1); one specified other (Chanen et al., 2022, Quarter 1), and one CBT (Morey, Lowmaster, & Hopwood, 2010, Quarter 1). Seven of 9 were from Quarter 1. Fourteen outliers (to the right) were related to less actual dropouts than predicted, all from relatively more precise observations: four DBT (Fitzpatrick, 2020, Quarter

1; Sinnaeve, 2018, Quarter 3; Verheul et al., 2003, Quarter 4; Walton, Bendit, Baker, Carter, & Lewin, 2020, Quarter 1); one MBT (Barnicot & Crawford, 2019, Quarter 3); five TAU (Bos, van Wel, Appelo, & Verbraak, 2010, Quarter 1; Carter, Willcox, Lewin, Conrad, & Bendit, 2010, Quarter 2; Kleindienst et al., 2011, Quarter 1; Majdara, Rahimian-Boogar, Talepasand, & Gregory, 2021, Quarter 1; Priebe et al., 2012, Quarter 1), one CTBE (Linehan et al., 2006, Quarter 1); two Specified Other (Chanen et al., 2022, Quarter 1 & 4); and one CBT (Cottraux et al., 2009, Quarter 1). Again, most outliers were from Quarter 1 (9/14). Two outpatient DBT study-arms had outliers of different signs (at different quarters; Fitzpatrick 2020; Sinnaeve 2018), indicating rather the timing of dropout than the cumulative dropout was diverting from the model. Given the heterogeneous character of TAU, it is understandable that relatively many outliers (7/23)

**Table 3.** Nominal predictors: Retention chances and follow-up contrasts of the final model (MI on complete study set)

Predictor	Transformed scale				Estimated survival chance per quarter			
	Pooled mean	Rubin's s.e.	Deviation contrast		Pooled mean	95% CI		
			<i>t</i>	<i>p</i>		Lower	Upper	
<b>Dropout per quarter imputed*</b>								
<b>No*</b>	<b>0.803</b>	<b>0.049</b>	<b>2.101</b>	<b>0.038</b>	<b>0.893</b>	<b>0.869</b>	<b>0.914</b>	
<b>Yes*</b>	<b>0.657</b>	<b>0.064</b>	<b>-2.101</b>	<b>0.038</b>	<b>0.855</b>	<b>0.818</b>	<b>0.887</b>	
<b>Treatment format*</b>								
Individual	0.807	0.054	1.706	0.09	0.894	0.867	0.917	
Combined individual-group	0.784	0.052	1.602	0.11	0.888	0.862	0.912	
<b>Group*</b>	<b>0.598</b>	<b>0.077</b>	<b>-2.724</b>	<b>0.007</b>	<b>0.838</b>	<b>0.791</b>	<b>0.879</b>	
<b>Quarter*</b>								
<b>Quarter 1*</b>	<b>0.475</b>	<b>0.048</b>	<b>-9.244</b>	<b>&lt;0.001</b>	<b>0.800</b>	<b>0.769</b>	<b>0.829</b>	
Quarter 2	0.696	0.054	-1.020	0.31	0.865	0.835	0.893	
<b>Quarter 3*</b>	<b>0.843</b>	<b>0.060</b>	<b>3.095</b>	<b>0.002</b>	<b>0.902</b>	<b>0.873</b>	<b>0.927</b>	
<b>Quarter 4*</b>	<b>0.905</b>	<b>0.063</b>	<b>4.581</b>	<b>&lt;0.001</b>	<b>0.916</b>	<b>0.888</b>	<b>0.939</b>	
<b>Treatment category*</b>								
DBT	0.670	0.060	-1.084	0.28	0.858	0.824	0.889	
DBTReduced	0.553	0.095	-1.930	0.054	0.824	0.764	0.877	
<b>ST*</b>	<b>1.049</b>	<b>0.071</b>	<b>5.000</b>	<b>&lt;0.001</b>	<b>0.942</b>	<b>0.916</b>	<b>0.962</b>	
TFP	0.705	0.104	-0.280	0.78	0.868	0.808	0.916	
<b>MBT*</b>	<b>0.908</b>	<b>0.081</b>	<b>2.580</b>	<b>0.010</b>	<b>0.916</b>	<b>0.880</b>	<b>0.945</b>	
TAU	0.805	0.067	1.339	0.18	0.893	0.859	0.922	
PsychoDynamic	0.769	0.079	0.563	0.57	0.884	0.842	0.920	
<b>CTBE*</b>	<b>0.320</b>	<b>0.127</b>	<b>-3.699</b>	<b>&lt;0.001</b>	<b>0.748</b>	<b>0.658</b>	<b>0.829</b>	
<b>Specified other*</b>	<b>0.827</b>	<b>0.062</b>	<b>2.034</b>	<b>0.042</b>	<b>0.898</b>	<b>0.868</b>	<b>0.924</b>	
CBT	0.691	0.121	-0.358	0.72	0.864	0.793	0.920	
<b>Treatment category × quarter*</b>								
Quarter 1	DBT	0.431	0.064	-0.702	0.48	0.785	0.743	0.826
	DBTReduced	0.423	0.094	-0.562	0.57	0.783	0.719	0.841
	<b>ST*</b>	<b>0.872</b>	<b>0.086</b>	<b>4.989</b>	<b>&lt;0.001</b>	<b>0.908</b>	<b>0.868</b>	<b>0.941</b>
Quarter 1	TFP	0.309	0.134	-1.412	0.16	0.744	0.649	0.830
	<b>MBT*</b>	<b>1.066</b>	<b>0.121</b>	<b>5.453</b>	<b>&lt;0.001</b>	<b>0.945</b>	<b>0.899</b>	<b>0.975</b>
Quarter 1	TAU	0.448	0.073	-0.426	0.67	0.791	0.742	0.836
	PsychoDynamic	0.554	0.098	0.891	0.37	0.824	0.762	0.879
	<b>CTBE*</b>	<b>-0.221</b>	<b>0.161</b>	<b>-4.832</b>	<b>&lt;0.001</b>	<b>0.551</b>	<b>0.443</b>	<b>0.667</b>
	<b>Specified Other*</b>	<b>0.607</b>	<b>0.070</b>	<b>2.142</b>	<b>0.033</b>	<b>0.840</b>	<b>0.798</b>	<b>0.878</b>
Quarter 1	CBT	0.263	0.128	-1.854	0.06	0.728	0.637	0.812
	Quarter 2	DBT	0.629	0.069	-0.984	0.33	0.847	0.805
	DBTReduced	0.566	0.119	-1.108	0.27	0.828	0.752	0.892
	<b>ST*</b>	<b>0.957</b>	<b>0.093</b>	<b>2.917</b>	<b>0.004</b>	<b>0.926</b>	<b>0.886</b>	<b>0.956</b>
Quarter 2	TFP	0.675	0.154	-0.152	0.88	0.860	0.766	0.930
	<b>MBT*</b>	<b>0.959</b>	<b>0.112</b>	<b>2.615</b>	<b>0.009</b>	<b>0.926</b>	<b>0.877</b>	<b>0.961</b>
	TAU	0.811	0.093	1.390	0.17	0.895	0.847	0.933

(Continued)



Table 3. (Continued.)

Predictor	Transformed scale				Estimated survival chance per quarter		
	Pooled mean	Rubin's s.e.	Deviation contrast		Pooled mean	95% CI	
			<i>t</i>	<i>p</i>		Lower	Upper
PsychoDynamic	0.701	0.106	0.050	0.96	0.867	0.806	0.916
<b>CTBE*</b>	<b>0.202</b>	<b>0.182</b>	<b>-2.970</b>	<b>0.003</b>	<b>0.706</b>	<b>0.575</b>	<b>0.826</b>
Specified Other	0.798	0.078	1.445	0.15	0.891	0.851	0.925
CBT	0.662	0.209	-0.189	0.85	0.856	0.724	0.946
Quarter 3							
DBT	0.700	0.081	-1.753	0.08	0.866	0.820	0.906
<b>DBTReduced*</b>	<b>0.514</b>	<b>0.166</b>	<b>-2.088</b>	<b>0.037</b>	<b>0.812</b>	<b>0.701</b>	<b>0.901</b>
<b>ST*</b>	<b>1.233</b>	<b>0.111</b>	<b>3.610</b>	<b>&lt;0.001</b>	<b>0.968</b>	<b>0.937</b>	<b>0.986</b>
TFP	0.905	0.173	0.392	0.70	0.915	0.828	0.969
MBT	0.809	0.107	-0.331	0.74	0.894	0.838	0.937
TAU	0.946	0.120	0.922	0.36	0.924	0.869	0.962
PsychoDynamic	0.827	0.131	-0.127	0.90	0.898	0.829	0.948
CTBE	0.703	0.216	-0.704	0.48	0.867	0.733	0.954
Specified Other	0.890	0.088	0.584	0.56	0.912	0.872	0.945
CBT	0.901	0.263	0.250	0.80	0.915	0.770	0.984
Quarter 4							
DBT	0.919	0.094	0.163	0.87	0.919	0.876	0.951
DBTReduced	0.708	0.192	-1.075	0.28	0.869	0.752	0.948
<b>ST*</b>	<b>1.134</b>	<b>0.107</b>	<b>2.170</b>	<b>0.030</b>	<b>0.955</b>	<b>0.920</b>	<b>0.978</b>
TFP	0.930	0.170	0.161	0.87	0.921	0.838	0.971
MBT	0.798	0.108	-1.028	0.30	0.891	0.834	0.936
TAU	1.017	0.131	0.893	0.37	0.937	0.882	0.972
PsychoDynamic	0.995	0.143	0.682	0.50	0.933	0.870	0.972
CTBE	0.596	0.213	-1.574	0.12	0.837	0.697	0.936
Specified Other	1.012	0.099	1.157	0.25	0.936	0.896	0.965
CBT	0.939	0.268	0.142	0.89	0.922	0.780	0.987

Significant effects in **bold**. \*  $p < 0.05$ .

came from the TAU category. Taken together, the number of outliers is in the expected range, but whereas outliers indicating underestimation of treatment retention were at the higher precision level, outliers indicating overestimation of retention were at a more medium precision level. This is in line with the results of Egger's test. Note that no study had quarters with residuals systematically outside the 95% CI.

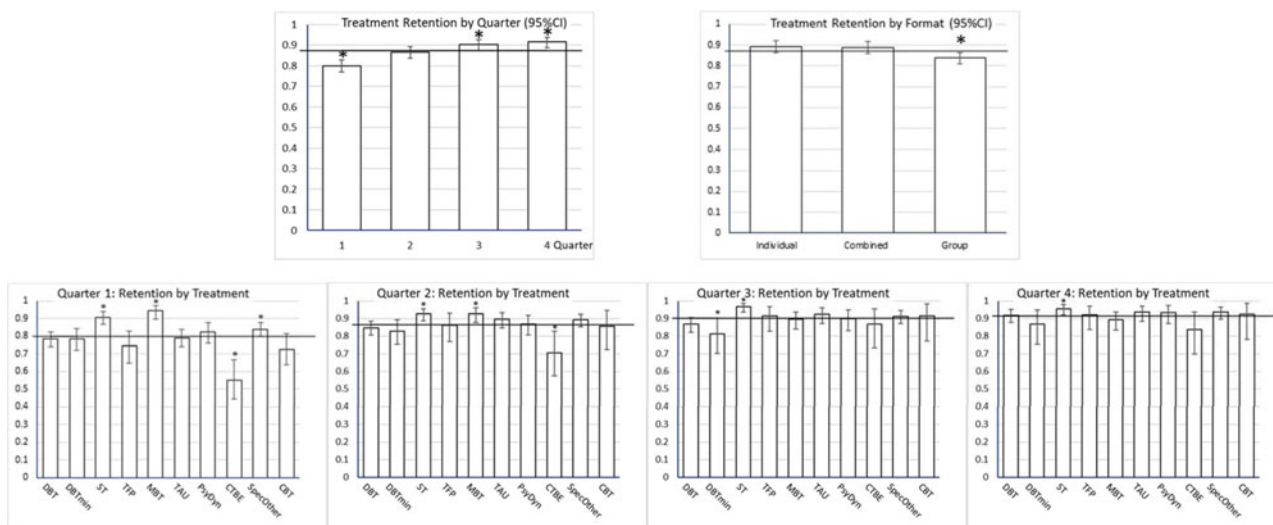
#### Sensitivity analysis: DBT studies with deviating pushout rules excluded

Two British DBT studies used a pushout rule deviating from the rule as formulated in the DBT protocol: participants were pushed out when they missed any consecutive series of 4 sessions, for instance 2 skills group and 2 individual coaching sessions within 2 weeks; whereas the original guideline is 4 consecutive sessions of either group or individual (Barnicot & Gaglia, personal communication, 24 September 2016). The more stringent rule used in the two studies (Gaglia et al., 2013; Priebe et al., 2012) seems related to relatively high dropout (Appendix D). We therefore repeated the analyses with the DBT arms of these studies excluded. The initial model, before backward deletion, was highly similar to

the one based on the complete study set, except that treatment format was significant (Table 4). The random effect of study was significant,  $z = 4.889$ ,  $p < 0.001$  ( $I^2 = 100 \cdot 146/1.146 = 12.7\%$ ). Backward deletion resulted in the same set of predictors (quarter, treatment, treatment format and dropout imputation) as in the primary analysis (Table 4). For the final model only the treatment by quarter interaction was added, Table 4. Egger's test was significant, explaining 3.9% of the variance, Table 4. The random effect of study was significant,  $z = 5.286$ ,  $p < 0.001$  ( $I^2 = 100 \cdot 120/1.120 = 10.7\%$ ). Figure 5 shows the fixed effects of the final model, Table 5 the statistics of the deviation contrasts.

The results of the deviation contrasts can be described as follows.

- (1) *Quarter*. As in the primary analysis, the first quarter had the lowest retention, with later quarters showing increasing levels of treatment retention.
- (2) *Dropout imputation per quarter* was significantly related to less retention.
- (3) *Treatment models*. As in the primary analysis, ST and MBT showed generally higher and CTBE less treatment retention than average. As to the treatment by quarter interaction, the



**Fig. 2.** Treatment retention proportion per quarter (with 95%CI) as estimated in the complete dataset. The horizontal line is the average treatment retention, to which the estimated effects are compared (deviation contrasts). Significant effects ( $p < 0.05$ ) indicated by \*. Upper left panel: treatment retention by quarter, showing increasing retention with time. Upper right panel: treatment retention by treatment format, showing significantly less retention in group treatment. Lower panels: treatment retention by treatment types and quarter. In all quarters, ST had significantly higher treatment retention than average. In quarters 1 and 2 MBT had significantly higher retention, CTBE significantly less, than average. Reduced DBT (DBTmin) had significantly less retention in quarter 3.

results were mostly similar compared to the primary analysis, with the exception that reduced DBT did no longer show lower treatment retention in quarter 3. **Figure 6a** shows the (cumulative) survival curves per treatment. After one year, the (unweighted) average retention was about 57%, with CTBE showing considerably lower (28%), MBT and ST higher treatment retention (70%, 77%).

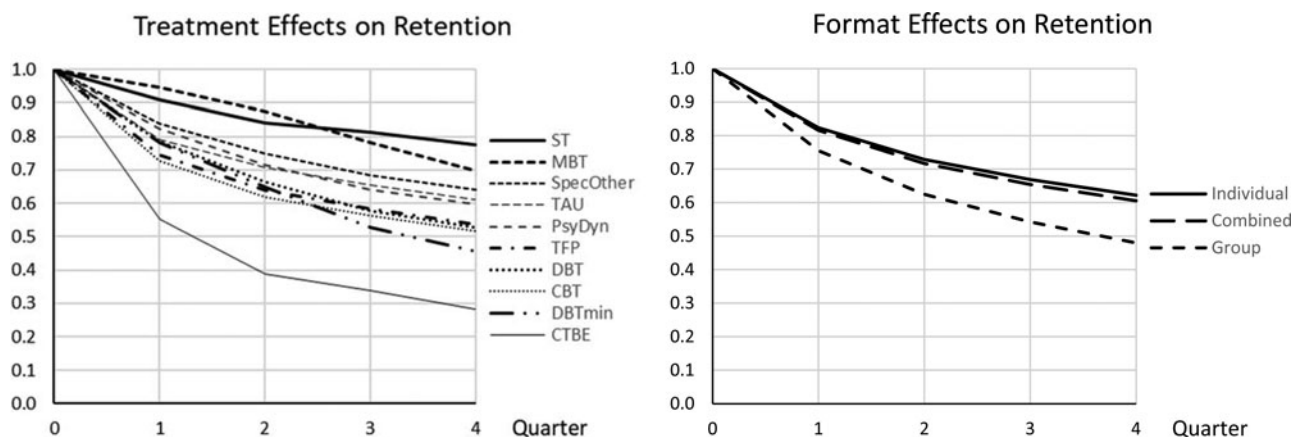
- (4) *Treatment format.* Individual treatment had significantly higher and group significantly lower than average treatment retention, combined individual-group format in between the other two. The relationship was approximately linear: the stronger the group component, the lower treatment retention was. **Figure 6b** shows the survival curves for the three formats over 1 year. At 1 year, the retention estimate was 66.5% for individual, 60% for combined, and 48% for group format.

**Figure 7** depicts the funnel plot of the residuals of the final analysis of the reduced study set (see Appendix H for funnel

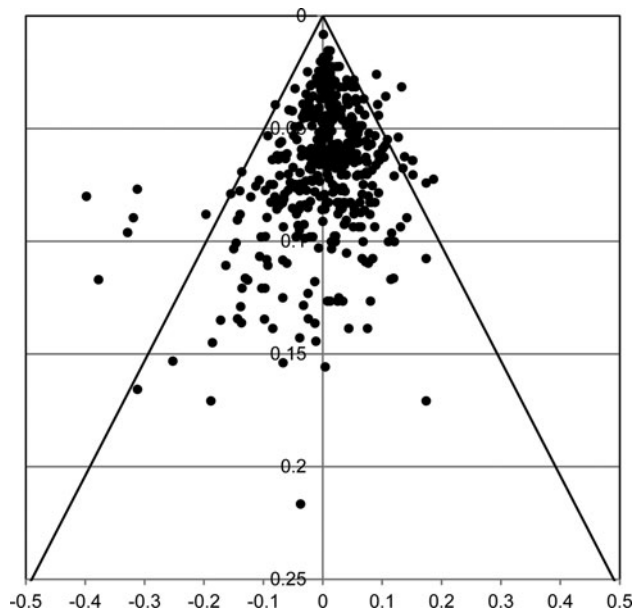
plots per treatment category). There were 24 outliers out of a total of 455 residuals (5.3%). Most outliers were the same as in the full data analysis, however one disappeared (Priebe 2012, TAU, Q1) and two additional positive residuals emerged (Barnicot & Crawford, 2019, MBT, Q4; Sachdeva, Goldman, Mustata, Deranja, & Gregory, 2013, TAU, Q1). Most were from quarter 1 (16/24).

**Discussion**

Premature treatment discontinuation is a well-known phenomenon in the treatment of BPD. It has not only plagued health care of BPD-patients for years, but has also motivated the development of specialized approaches, like the ‘big-4’, that were designed to (among other things) reduce treatment dropout. We used a meta analytic approach to study treatment retention in psychological therapies for BPD, and tested various factors that might be associated with treatment retention. We found



**Fig. 3.** Retention curves for 4 quarters for the complete data set. (a) (left). Cumulative treatment retention over 4 quarters depicted with survival curves for the 10 treatment models. Over 1 year CTBE had considerable less treatment retention, while ST and MBT had considerable more. (b) (right). Cumulative treatment retention over 4 quarters depicted with survival curves for the 3 treatment formats. Over 1 year group formats had considerable less treatment retention than the other two.



**Fig. 4.** Funnel plot of 463 residuals of the final GLMM survival analysis ( $x$ -axis = residual;  $y$ -axis = study precision per quarter). Residuals were the differences between observed and estimated survival proportions. To the left residuals related to more actual dropouts in a quarter than predicted by the model, to the right residuals related to less actual dropouts than predicted by the model. There were 23 (4.96%) residuals outside the 95% CI.

evidence for superior treatment retention in MBT and ST. CTBE showed very poor treatment retention, mainly in the first two quarters of treatment. Specified others showed somewhat higher retention than average in Quarter 1, CBT lower retention in Quarter 1 and reduced DBT in Quarter 3 (both in one of the two analyses). All other treatment categories did not differ significantly from the average. We did not find any evidence that on average over the last 32 years treatment retention improved, nor did we find evidence that gender or patients' age had any influence. No effect of treatment setting was detected, with no significant differences between inpatient, outpatient and day-treatment settings. However, there was evidence for an effect of treatment format, with group therapy having higher dropout than other formats. Interestingly, when the two DBT arms that used a more stringent pushout rule than the original DBT protocol were excluded from the analysis, individual therapy had the lowest and group therapy the highest dropout, with mixed individual-group approaches having average retention. Study design (RCT, open trial, nonrandomized controlled), design quality, dropout type (treatment dropout *v.* no distinction made between treatment and study dropout), prescribed medication policy, treatment offered in addition to TAU, country group, and type of exclusion of substance use, had no significant effects on dropout rates. However, insufficient details about timing of treatment dropouts (necessitating MI to estimate dropout per quarter) had a significant effect in that these studies were associated with less treatment retention. By including this study characteristic as a covariate we controlled for this effect. Last, but not least, we found that it is especially the first quarter of treatment during which dropout manifests itself.

Egger's test and Funnel plots indicated that less precise studies, such as studies with a small sample size, were associated with less treatment retention. Note that this is opposite to what might be expected in case of (publication) bias, where less precision

would be expected to be associated with overly optimistic findings, i.e. higher treatment retention. Moreover, though significant, the effect was small (<4% explained variance). The number of residuals exceeding a magnitude that could be expected on the basis of a 95% CI was around the to be expected 5%. Most of the extreme residuals came from the first Quarter and from treatment categories that were heterogeneous, such as TAU, CTBE, and specified others. With seven exceptions of DBT and one of MBT, none of the 'Big-four' residuals was excessive. Thus, treatment dropout could be estimated fairly precisely for these four specialized psychotherapies.

Excluding the DBT-arms of the Priebe and Gaglia studies as was done in the sensitivity analysis might yield more trustworthy results than an analysis including these arms. Interestingly, an effect of individual therapy format appeared in this analysis, indicating that the stronger the individual component of the treatment, the higher the retention rate is. A recent RCT compared predominantly group to combined individual-group treatment and also found more dropout from the predominantly group format, supporting a causal interpretation of format (Arntz et al., 2022).

Our study differs from previous studies documenting dropout from psychological treatment for BPD (e.g. Barnicot et al., 2011; Iliakis, Ilagan, & Choi-Kain, 2021; Stoffers-Winterling et al., 2012; Storebø et al., 2020). For instance, in contrast to the Cochrane meta-analysis by Stoffers-Winterling et al. (2012) and Storebø et al. (2020), we included all kinds of designs, investigated the development of retention over time (and not accumulated over time), included multiple predictors (i.e. meta-regression), and based the analysis on individual cases (thus, in a sense our study was a 'mega-analysis', although for some predictors we did not have individual values). In contrast to the approach chosen in the Cochrane analysis, our approach allowed comparison between treatment models and formats. Although the meta-analysis by Barnicot et al. (2011) distinguished between treatments with a duration shorter than one year *v.* longer treatments, the survival analysis approach we used was more fine-grained in its modeling of the development of dropout over time. Moreover, we found that in addition to time (quarter), treatment format and treatment model were important predictors. Our findings suggest that the substantial between-study heterogeneity found by Barnicot et al. (2011) and Iliakis et al. (2021) can be explained for an important part by these variables. Note that the pooled completion rates found by Barnicot et al. (2011) are similar to those we found.

### Clinical implications

The results suggest some important implications for clinical practice. First, as most dropout takes place in the first quarter, it is pivotal to give attention at the start of treatment to factors that influence treatment engagement. Although this has been acknowledged in some specialized models (e.g. the 'contract phase' in TFP), the present results do not support that all attempts are successful. More research is needed to understand why patients tend to dropout so early in treatment and how treatment can be made more acceptable, especially in the early phase. Second, pure individual treatment has superior retention, whereas the larger the group part in treatment is, the lower treatment retention is. Factors like practical (e.g. agenda) problems, but also group-dynamics might play a role here. Although mixed individual-group models seem to do better, it is the pure

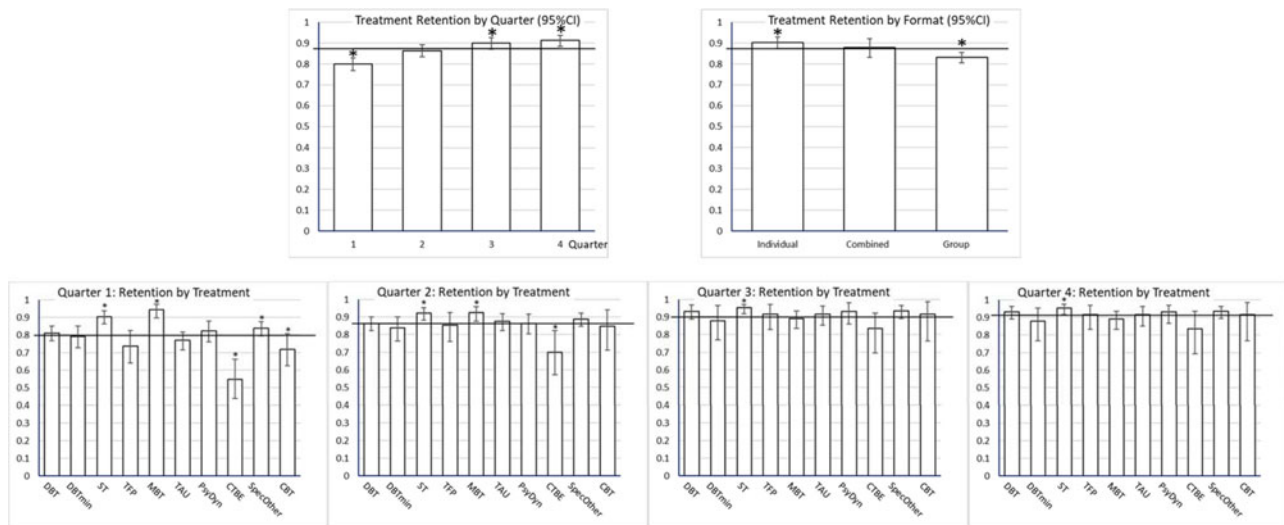
**Table 4.** *F*-tests of predictors of treatment retention of the fixed part of the initial, intermediate, and final models, without DBT-arms of Priebe and Gaglia studies

Model	<i>F</i>	df1	df2	<i>p</i>
Initial model				
Corrected model	6.753	33	179	<0.001
Year of publication	1.753	1	70	0.19
Methodological quality of study's arm	0.081	1	120	0.78
Trial type	0.454	2	75	0.64
Dropout type distinguished	0.495	1	67	0.48
TAU added to treatment	0.602	1	632	0.44
<b>Dropout per quarter to be imputed*</b>	<b>5.705</b>	<b>1</b>	<b>137</b>	<b>0.018</b>
Setting (inpatient, day treatment, outpatient)	1.646	2	273	0.20
<b>Treatment format (Individual, Group, Combined)</b>	<b>3.344</b>	<b>2</b>	<b>1278</b>	<b>0.036</b>
Medicine prescribed by study	0.985	1	80	0.32
Substance use exclusion	0.535	4	70	0.71
Country group (Europe, USA, Aus/Can/NZ, emerging)	0.266	3	68	0.85
Mean age of study's arm	0.077	1	539	0.78
Proportion males of study's arm	0.067	1	375	0.80
<b>Treatment category*</b>	<b>4.975</b>	<b>9</b>	<b>987</b>	<b>&lt;0.001</b>
<b>Quarter*</b>	<b>47.817</b>	<b>3</b>	<b>22 986</b>	<b>&lt;0.001</b>
Intermediate model (Main Effects after backwards deletion)				
Corrected model	14.614	15	1335	<0.001
<b>Dropout per quarter to be imputed*</b>	<b>5.528</b>	<b>1</b>	<b>136</b>	<b>0.020</b>
<b>Treatment format (Individual, Group, Combined)*</b>	<b>4.718</b>	<b>2</b>	<b>944</b>	<b>0.009</b>
<b>Treatment category*</b>	<b>6.194</b>	<b>9</b>	<b>1754</b>	<b>&lt;0.001</b>
<b>Quarter*</b>	<b>49.073</b>	<b>3</b>	<b>23 004</b>	<b>&lt;0.001</b>
Final main and interaction effects model				
Corrected model	7.007	42	4441	<0.001
<b>Dropout per quarter to be imputed*</b>	<b>5.648</b>	<b>1</b>	<b>131</b>	<b>0.019</b>
<b>Treatment format (Individual, Group, Combined)*</b>	<b>4.947</b>	<b>2</b>	<b>959</b>	<b>0.007</b>
<b>Treatment category*</b>	<b>4.811</b>	<b>9</b>	<b>2248</b>	<b>&lt;0.001</b>
<b>Quarter*</b>	<b>36.769</b>	<b>3</b>	<b>22 977</b>	<b>&lt;0.001</b>
Treatment category × quarter*	2.757	27	22 977	<0.001
<b>Egger's test (1/√(N))*</b>	<b>6.453</b>	<b>1</b>	<b>159</b>	<b>0.012</b>

individual treatment that has the highest treatment retention. The tendency to provide more therapies in group formats in an attempt to reduce delivery costs might thus result in higher personal and societal costs associated with dropout. More research is needed to understand what underlies dropout from group treatment, what can be done to prevent this, and, perhaps, how patients can be better matched to group, combined, or individual treatment. Better understanding of factors that are involved in dropout from groups is even more important when patients will learn about the higher dropout risk from group treatment, as this might increase resistance against groups. As not all patients drop out from groups, and the treatments with the highest retention, MBT and ST, involve group components (for ST the recent models) there might be more factors involved than the group modality as such. For example, the degree of structure and

hence safety in the group probably plays an important role in preventing dropout. On the other hand, the format effect should not be underestimated: even for ST the difference between individual (87% retention over 1 year) and group (73% retention) is large. Pending research that will help to better personalize the matching of treatment format to patients, a clinical recommendation might be to take resistance to group treatment seriously and consider individual treatment for those that indicate to be too distrustful, inhibited, or easily provoked (in anger or aggression) to participate in a group. In other words, it is suggested to explore with the patient whether a group format is a good match with the patient instead of mechanically putting everyone in a group treatment for reasons of efficiency.

The finding that more than 30 years of research has not led to a general improvement of treatment retention is disappointing.



**Fig. 5.** Treatment retention proportion per quarter (with 95% CI) as estimated in the reduced dataset, without DBT-arms with deviating pushout rules. The horizontal line is the average treatment retention, to which the estimated effects are compared (deviation contrasts). Significant effects ( $p < 0.05$ ) indicated by \*. Upper left panel: treatment retention by quarter, showing increasing retention with time. Upper right panel: treatment retention by treatment format, illustrating significantly less retention in group and more in individual treatment. Lower panels: treatment retention by treatment types and quarter. In all quarters, ST had significantly higher treatment retention than average. In quarters 1 and 2 MBT had significantly higher retention, CTBE significantly less, than average. In Quarter 1, specified others had significantly more and CBT less retention than average.

However, during this period new specialized treatment models were developed and tested, and the present evidence suggests that some, but not all of them, might actually prevent premature treatment discontinuation. The results therefore do not support claims that all (specialized) approaches are equal (e.g. Paris, 2010). If dropout systematically differs between treatment approaches, there must be specific factors that account for this. In the absence of equivalence trials (or equivalence meta-analysis), claims that treatments are equivalent, typically based on nonsignificant differences between treatments, were premature anyway. The present findings cast serious doubt on such claims, at least with regard to treatment retention.

The finding that CTBE has such a high dropout rate is remarkable, especially because CTBE has been framed as a superior variant of TAU. One interpretation is that patients agree to participate in a trial involving CTBE in the hope to get the experimental and not the control treatment, and drop out if they do not get the preferred treatment. However, one would then expect a similar finding with TAU, which was not the case. Possibly it is the strong focus on addressing difficult issues of the patient by CTBE therapists using traditional, confrontational psychotherapeutic strategies that makes the treatment difficult to tolerate.

More traditional psychodynamic treatment than for instance TFP and MBT did not show inferior treatment retention. This is surprising, given early findings that psychodynamic psychotherapy with BPD-patients was associated with high dropout (Gunderson et al., 1989; Skodol et al., 1983; Waldinger & Gunderson, 1984). These early studies however used a wider definition of BPD, than that of the DSM-III and later editions, and thus might have included more difficult patients. Moreover, the presently included studies of psychodynamic treatments might have investigated (successful) adaptations of psychodynamic psychotherapy, e.g. involving a more supportive and less neutral stance of therapists, that differ from the more traditional versions that were investigated in early studies.

The pooled treatment retention in DBT was strongly influenced by two studies using a stricter pushout rule than the

DBT protocol prescribes. Excluding the two studies from the analysis led to an increase in estimated treatment retention in DBT (estimated survival chance over 1 year increased from 53% to 59%). Taken together, the present data indicate that the original DBT protocol, and not more stringent rules should be followed when it comes to prevent treatment dropout.

The finding that ST has low dropout rates is in line with previous observations, both in BPD and in other PDs. For instance, RCTs for non-borderline PDs have also found superior treatment retention in ST compared to other treatments (Bamelis, Evers, Spinhoven, & Arntz, 2014; Bernstein et al., 2021). What might explain the high treatment retention in ST? First, patients tend to appreciate the therapeutic relationship in ST higher than in comparison conditions, and higher appreciation is associated with treatment retention (Bamelis et al., 2014; Spinhoven, Giesen-Bloo, van Dyck, Kooiman, & Arntz, 2007). Second, qualitative research into patients' perspectives suggests a number of elements in ST that are particularly appreciated by patients: (i) the schema mode model, helping patients to better understand and control their problems; (ii) the therapeutic relationship: which is experienced as more personal, directive and caring than in other models; and (iii) specific ST techniques, notably experiential techniques, such as imagery rescripting, which are reported as particularly helpful (de Klerk, Abma, Bamelis, & Arntz, 2017). Third, in contrast to what was found with other treatments (Katsakou et al., 2012), patients did not report a too narrow focus of ST (de Klerk et al., 2017). Thus, it seems that the ST model meets the needs of patients quite well. ST for BPD has now been studied in 12 studies, in 7 countries, by different research groups. Clearly, this still limited database calls for further studies that will help to clarify whether ST indeed is characterized by a high treatment retention.

ST was not the only treatment model showing superior treatment retention: in the first two quarters of treatment, this was also shown by MBT. However, the data indicate that the initial high retention chance in MBT is not maintained in later phases of treatment, whereas ST continued to show the highest retention



**Table 5.** Nominal Predictors: Retention chances and follow-up contrasts of the final model [reduced study set (without DBT arms from Priebe 2012 and Gaglia et al., 2013)]

Predictor	Transformed scale				Estimated survival chance per quarter			
	Pooled mean	Rubin's s.e.	Deviation contrast		Pooled mean	95% CI		
			<i>t</i>	<i>P</i>		Upper	Lower	
<b>Dropout per quarter imputed*</b>								
<b>No*</b>	<b>0.808</b>	<b>0.050</b>	<b>2.266</b>	<b>0.025</b>	<b>0.894</b>	<b>0.869</b>	<b>0.916</b>	
<b>Yes*</b>	<b>0.645</b>	<b>0.066</b>	<b>-2.266</b>	<b>0.025</b>	<b>0.851</b>	<b>0.813</b>	<b>0.886</b>	
<b>Treatment format*</b>								
<b>Individual*</b>	<b>0.850</b>	<b>0.056</b>	<b>2.612</b>	<b>0.009</b>	<b>0.904</b>	<b>0.877</b>	<b>0.927</b>	
Combined individual-group	0.750	0.054	0.666	0.51	0.880	0.851	0.905	
<b>Group*</b>	<b>0.579</b>	<b>0.078</b>	<b>-2.988</b>	<b>0.003</b>	<b>0.832</b>	<b>0.784</b>	<b>0.875</b>	
<b>Quarter*</b>								
<b>Quarter 1*</b>	<b>0.473</b>	<b>0.049</b>	<b>-9.134</b>	<b>&lt;0.001</b>	<b>0.799</b>	<b>0.767</b>	<b>0.829</b>	
Quarter 2	0.692	0.055	-1.033	0.30	0.864	0.833	0.892	
<b>Quarter 3*</b>	<b>0.841</b>	<b>0.062</b>	<b>3.112</b>	<b>0.002</b>	<b>0.901</b>	<b>0.872</b>	<b>0.927</b>	
<b>Quarter 4*</b>	<b>0.899</b>	<b>0.064</b>	<b>4.473</b>	<b>&lt;0.001</b>	<b>0.914</b>	<b>0.885</b>	<b>0.938</b>	
<b>Treatment category*</b>								
DBT	0.753	0.064	0.459	0.65	0.880	0.846	0.910	
DBTReduced	0.591	0.097	-1.454	0.15	0.836	0.776	0.887	
<b>ST*</b>	<b>1.033</b>	<b>0.073</b>	<b>4.801</b>	<b>&lt;0.001</b>	<b>0.940</b>	<b>0.913</b>	<b>0.961</b>	
TFP	0.689	0.105	-0.416	0.68	0.864	0.802	0.913	
<b>MBT*</b>	<b>0.910</b>	<b>0.082</b>	<b>2.667</b>	<b>0.008</b>	<b>0.917</b>	<b>0.880</b>	<b>0.946</b>	
TAU	0.726	0.071	0.005	1.00	0.873	0.834	0.907	
PsychoDynamic	0.765	0.081	0.548	0.58	0.883	0.840	0.919	
<b>CTBE*</b>	<b>0.311</b>	<b>0.128</b>	<b>-3.176</b>	<b>&lt;0.001</b>	<b>0.745</b>	<b>0.654</b>	<b>0.827</b>	
Specified other	0.818	0.063	1.898	0.06	0.896	0.865	0.923	
CBT	0.664	0.123	-0.568	0.57	0.857	0.783	0.916	
<b>Treatment category × quarter*</b>								
Quarter 1	DBT	0.510	0.068	0.572	0.57	0.811	0.768	0.851
	DBTReduced	0.453	0.096	-0.209	0.83	0.793	0.728	0.850
<b>ST*</b>	<b>0.858</b>	<b>0.087</b>	<b>4.811</b>	<b>&lt;0.001</b>	<b>0.905</b>	<b>0.863</b>	<b>0.939</b>	
TFP	0.294	0.135	-1.517	0.13	0.738	0.643	0.826	
<b>MBT*</b>	<b>1.067</b>	<b>0.121</b>	<b>5.463</b>	<b>&lt;0.001</b>	<b>0.945</b>	<b>0.899</b>	<b>0.975</b>	
TAU	0.385	0.076	-1.331	0.18	0.770	0.718	0.818	
PsychoDynamic	0.554	0.010	0.905	0.37	0.824	0.761	0.879	
<b>CTBE*</b>	<b>-0.232</b>	<b>0.162</b>	<b>-4.865</b>	<b>&lt;0.001</b>	<b>0.548</b>	<b>0.439</b>	<b>0.663</b>	
<b>Specified other*</b>	<b>0.598</b>	<b>0.071</b>	<b>2.006</b>	<b>0.045</b>	<b>0.838</b>	<b>0.795</b>	<b>0.876</b>	
<b>CBT*</b>	<b>0.243</b>	<b>0.130</b>	<b>-1.986</b>	<b>0.048</b>	<b>0.721</b>	<b>0.628</b>	<b>0.807</b>	
Quarter 2	DBT	0.696	0.073	0.059	0.95	0.865	0.824	0.901
	DBTReduced	0.601	0.121	-0.771	0.44	0.839	0.763	0.901
	<b>ST*</b>	<b>0.942</b>	<b>0.093</b>	<b>2.788</b>	<b>0.005</b>	<b>0.923</b>	<b>0.882</b>	<b>0.954</b>
	TFP	0.659	0.155	-0.238	0.81	0.855	0.760	0.927
	<b>MBT*</b>	<b>0.964</b>	<b>0.113</b>	<b>2.699</b>	<b>0.007</b>	<b>0.927</b>	<b>0.878</b>	<b>0.962</b>

(Continued)

Table 5. (Continued.)

Predictor	Transformed scale				Estimated survival chance per quarter		
	Pooled mean	Rubin's s.e.	Deviation contrast		Pooled mean	95% CI	
			<i>t</i>	<i>P</i>		Upper	Lower
TAU	0.741	0.097	0.571	0.57	0.877	0.824	0.921
PsychoDynamic	0.696	0.107	0.046	0.96	0.866	0.803	0.916
<b>CTBE*</b>	<b>0.194</b>	<b>0.182</b>	<b>-2.994</b>	<b>0.003</b>	<b>0.703</b>	<b>0.572</b>	<b>0.824</b>
Specified other	0.789	0.079	1.367	0.17	0.889	0.848	0.923
CBT	0.637	0.212	-0.304	0.76	0.849	0.713	0.943
Quarter 3							
DBT	0.809	0.088	-0.368	0.71	0.894	0.849	0.930
DBTReduced	0.557	0.168	-1.789	0.07	0.825	0.715	0.911
<b>ST*</b>	<b>1.216</b>	<b>0.112</b>	<b>3.459</b>	<b>&lt;0.001</b>	<b>0.966</b>	<b>0.933</b>	<b>0.985</b>
TFP	0.888	0.174	0.302	0.76	0.912	0.823	0.967
MBT	0.811	0.108	-0.291	0.77	0.895	0.838	0.938
TAU	0.861	0.126	0.172	0.86	0.906	0.842	0.951
PsychoDynamic	0.822	0.133	-0.155	0.88	0.897	0.827	0.948
CTBE	0.695	0.217	-0.731	0.46	0.865	0.730	0.953
Specified other	0.880	0.088	0.473	0.64	0.910	0.868	0.943
CBT	0.869	0.267	0.121	0.90	0.908	0.757	0.982
Quarter 4							
DBT	0.998	0.101	1.034	0.30	0.934	0.892	0.963
DBTReduced	0.752	0.193	-0.796	0.43	0.880	0.766	0.955
<b>ST*</b>	<b>1.119</b>	<b>0.108</b>	<b>2.067</b>	<b>0.039</b>	<b>0.953</b>	<b>0.916</b>	<b>0.977</b>
TFP	0.914	0.170	0.098	0.92	0.918	0.832	0.969
MBT	0.800	0.109	-0.947	0.34	0.892	0.834	0.937
TAU	0.919	0.137	0.152	0.88	0.918	0.853	0.962
PsychoDynamic	0.988	0.144	0.672	0.50	0.932	0.868	0.972
CTBE	0.588	0.213	-1.583	0.11	0.835	0.694	0.935
Specified other	1.004	0.100	1.128	0.26	0.935	0.894	0.964
CBT	0.908	0.272	0.035	0.97	0.916	0.766	0.985

Significant effects in **bold**. \*  $p < 0.05$ .

chance per quarter. Nevertheless, cumulative treatment retention over 1 year was also relatively high in MBT (70% *v.* 78% in ST)

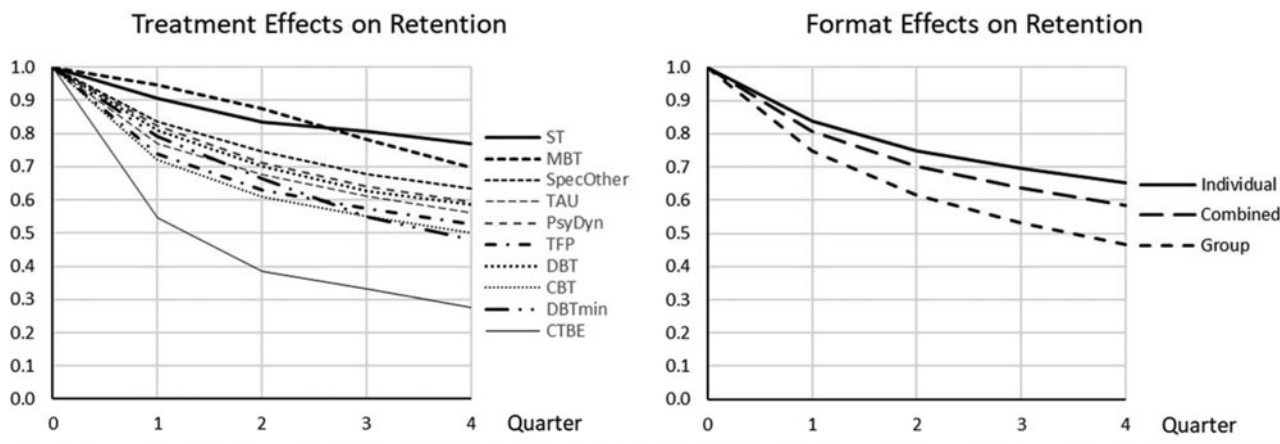
### Limitations

A number of limitations of the present meta-analysis should be considered. First, a meta-analysis is not an RCT – hence differences in populations and sites might influence findings. This might especially be a problem when not all treatment approaches (including treatment models, formats and settings) are directly compared in RCT's. On the other hand, we did not find evidence that the study design (RCT *v.* non-RCT *v.* open trial) influenced results. Moreover, not enough direct comparisons between treatments and formats are available, and it would be a virtually impossible task to compare 16 treatment models each provided in 3 formats, each delivered in 3 settings, thus 144 arms to each other in multiple RCTs (i.e. 10 296 comparisons). The current analysis provides a statistical summary, not a proof, of what

studies into psychological treatment of BPD found with respect to treatment retention. With all the problems that are inherent to a meta-analysis, the current study indicates what is associated with treatment retention, and what not, and thus informs clinical practice and researchers.

Second, estimation of (parts of) the development of dropout over time was necessary in many studies. Although we used MI as an appropriate statistical strategy to deal with this, it is always better to base the analysis on the original detailed survival lengths. Future trials should follow guidelines such as the CONSORT statement (Schulz, Altman, & Moher, 2010) and report treatment retention in sufficient detail.

Third, we could not always distinguish between treatment and study dropout. Although we did not find a significant effect of not distinguishing between study and treatment dropout, future studies should follow guidelines of reporting by distinguishing between dropout types (e.g. CONSORT-guidelines, Schulz et al., 2010).



**Fig. 6.** Retention curves for 4 quarters for the reduced data set (sensitivity analysis). (a) (left). Cumulative treatment retention over 4 quarters depicted with survival curves for the 10 treatment models, estimated from the reduced data set, without DBT-arms with deviant pushout rules. Over 1 year CTBE had considerable less treatment retention, while ST and MBT had considerable more. (b) (right). Cumulative treatment retention over 4 quarters depicted with survival curves for the 3 treatment formats, estimated from the reduced data set, without DBT-arms with deviant pushout rules. Over 1 year group formats had considerable less treatment retention, while individual had considerably more treatment retention than average. The combined format was in between.

Fourth, only two studies had a CTBE arm, limiting the generalizability of the CTBE findings, though the two studies were from different continents and reported remarkably similar findings. Nevertheless, the high dropout rates from CTBE are alarming, and questions both the ‘expert’ level of the therapists and the suitability of CTBE as comparison condition in RCTs.

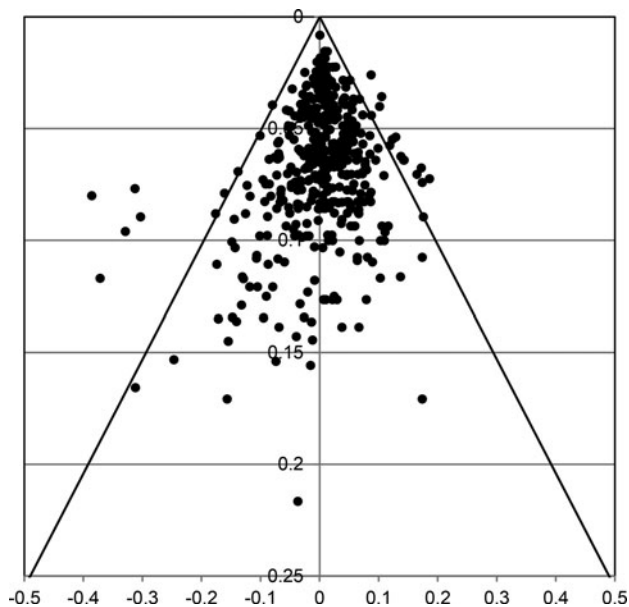
Fifth, some treatment models, such as DBT, have rules about pushouts, while others have not. Such rules are defined by the treatment protocol however, and are based on the model underlying the protocol. Thus, we considered the effects of such rules on treatment retention as inherent to these models. Future

research should test whether changes in such rules lead to changes in treatment retention, or not.

Sixth, the influence of covariates such as gender and age could only be assessed on an aggregated level. This not only reduced the power of our statistical tests of these variables, but also has the risk that effects are overseen as within-study relationships might not be detected when means of studies are analyzed (Simpson’s paradox). Analysis of individual data would solve this problem, but requires data sharing between researchers. Another issue that might limit the validity of the results on the influence of gender is the average 15% of male participants in the data. This is very low compared to prevalence estimates from the general population, that generally show equal prevalence in men and women (Torgersen, 2012). Samples from mental health centers generally show a dominance of female patients (about 75%, American Psychiatric Association, 2013). The difference in gender proportion might be related to the use of structured interviews by lay interviewers in epidemiological research (*v.* semi-structured interviews by clinicians in clinical samples), gender differences in help seeking behavior, and higher numbers of men with BPD in addiction treatment centers and forensic institutes. But even compared to the approximately 25% prevalence of men in clinical samples the 15% in the present data is low, possibly related to the habit to recruit female patients only by some researchers (e.g. Linehan et al., 2015). Thus, the finding that mean age and mean proportion of male patients did not relate to treatment retention should be interpreted with caution.

Seventh, we collapsed treatment models with sample size less than 100 into one category. The resulting specialized others category is quite heterogeneous, and conclusions about specific treatment models within this category cannot be drawn. The relatively high retention in Quarter 1 suggests that there might be promising treatments in this category.

Eighth, keeping patients in treatment that will not profit from it, is not cost-effective: the resources can better be allocated to those that will profit. Thus, high treatment retention as such is not necessarily good. On the other hand, the (cost-)effectiveness of treatments is limited by premature dropouts, as dropout limits the potential overall (cost)effectiveness. High treatment retention does not necessarily imply high effectiveness, however in a recent



**Fig. 7.** Funnel plot of 455 residuals of the final GLMM survival analysis (*x*-axis = residual; *y*-axis = study precision per quarter) of the reduced data set. Residuals were the differences between observed and estimated survival proportions. To the left residuals related to more actual dropouts in a quarter than predicted by the model, to the right residuals related to less actual dropouts than predicted by the model. There were 24 (5.3%) residuals outside the 95% CI.

meta-analysis on effectiveness using a similar multilevel approach as the current study we found evidence for relatively high effect sizes of ST and MBT (and lower effectiveness of TAU and CTBE; Rameckers et al., 2021).

Ninth, the team that performed the present meta-analysis was led by the first author, an ST expert, which raises the question whether allegiance effects influenced the findings. To prevent such effects, study selection was done by different combinations of 3 from 4 experts and although the first author co-selected studies, others were the majority, and none of them had an affiliation with ST. The first author did not assess quality nor coded dropout data and other study characteristics, and none of the coders had an ST-affiliation. The analyses were collaboratively done by the last author, a statistician not affiliated with any treatment, and the first author. Lastly, the data are given in the appendices so that other researchers can check them and conduct independent analyses.

## Conclusions

Although the current findings are not a definitive proof that treatments models and formats differ in treatment retention, they are provocative and help to stimulate further research into improving treatment retention for BPD patients. Contradicting the popular ‘Dodo bird verdict’ that all treatments are equal, the findings suggest that they are not when it comes to treatment retention. More specifically, CTBE does not seem to be a good idea, whereas ST and MBT seem to do a better job than other treatments. Individual treatment seems to protect against dropout, whereas group treatment in particular might be a risk factor for premature treatment ending. Some factors thought to be predictive of dropout were not supported, and it might not be recommended to use them for treatment selection.

**Supplementary material.** Appendices are provided in the supplementary material which can be found at <https://doi.org/10.1017/S0033291722003634>.

**Acknowledgements.** Thanks are due to Dr Pim Cuijpers for help with the literature search and study selection, to Dr Stephan Doering for raising the hypothesis that country might influence dropout, and to Dr Martin Bohus for raising the hypothesis that different exclusion policies of substance use disorders might influence dropout. Thanks are also due to the students that helped with the project in various phases.

**Financial support.** This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

**Conflict of interest.** Dr Arnaud Arntz contributed to the development and testing of CBT and ST. He wrote about ST and CBT in books and chapters, and occasionally gives workshops and training. Financial remunerations go to the university to support research.

## Notes

<sup>1</sup> Thanks are due to Dr. Stephan Doering for raising this hypothesis during a discussion at the ESSPD conference in Vienna, 8–10 September 2016.

<sup>2</sup> Thanks are due to Dr. Martin Bohus for raising this hypothesis during a discussion at the ESSPD conference in Vienna, 8–10 September 2016.

<sup>3</sup> SPSS 28.0 does not offer the possibility to run automated pooling of MI sets with the GLMM module, nor does it offer output of the covariances of the fixed effects. Therefore, pooling of fixed effects consisting of multiple levels of predictors (e.g., quarter, treatment, format) of the 20 MI sets with a conventional approach was not possible. The chosen strategy to base model selection on GLMM analyses with MI set as additional level is less conservative than a conventional MI strategy, based on Rubin’s rule. Thus, after model selection the final model might lead to overestimation of effects and their significance.

We therefore analyzed the t-tests of the deviation contrasts with a conventional MI analysis using Rubin’s rule for estimating pooled means and s.e.’s (for these the covariances are not required). This guarantees accurate estimates of means, s.e.’s, and contrasts of interest. The final conclusions of the analyses are based on these contrasts, and not on the overall *F*-tests.

## References

- Agresti, A., & Coull, B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, 52(2), 119–126.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in psychology: Why do we need them? *What might they be?* *American Psychologist*, 63, 839–851. doi:10.1037/0003-066X.63.9.839
- Arntz, A., Jacob, G. A., Lee, C. W., Brand-de Wilde, O. M., Fassbinder, E., Harper, R. P., ... Farrell, J. M. (2022). Effectiveness of predominantly group schema therapy and combined individual and group schema therapy for borderline personality disorder: A randomized clinical trial. *JAMA Psychiatry*, 79(4), 287–299. doi:10.1001/jamapsychiatry.2022.0010
- Arntz, A., Stupar-Rutenfrans, S., Bloo, J., van Dyck, R., & Spinhoven, P. (2015). Prediction of treatment discontinuation and recovery from borderline personality disorder: Results from an RCT comparing schema therapy and transference focused psychotherapy. *Behaviour Research and Therapy*, 74, 60–71. doi:10.1016/j.brat.2015.09.002
- Bamelis, L. L. M., Evers, S. M. A. A., Spinhoven, P., & Arntz, A. (2014). Results of a multicentered randomized controlled trial of the clinical effectiveness of schema therapy for personality disorders. *American Journal of Psychiatry*, 171(3), 305–322. doi:10.1176/appi.ajp.2013.12040518
- Barnicot, K., & Crawford, M. (2019). Dialectical behaviour therapy v. mentalisation-based therapy for borderline personality disorder. *Psychological Medicine*, 49(12), 2060–2068. doi:10.1017/S0033291718002878
- Barnicot, K., Katsakou, C., Marougka, S., & Priebe, S. (2011). Treatment completion in psychotherapy for borderline personality disorder—a systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, 123(5), 327–338. doi:10.1111/j.1600-0447.2010.01652.x
- Bernstein, D. P., Keulen-de Vos, M., Clercx, M., De Vogel, V., Kersten, G. C., Lancel, M., ... Arntz, A. (2021). Schema therapy for violent PD offenders: A randomized clinical trial. *Psychological Medicine*, 1–15. doi:10.1017/S0033291721001161
- Bos, E. H., van Wel, E. B., Appelo, M. T., & Verbraak, M. J. (2010). A randomized controlled trial of a Dutch version of systems training for emotional predictability and problem solving for borderline personality disorder. *The Journal of Nervous and Mental Disease*, 198(4), 299–304. doi:10.1097/NMD.0b013e3181d619cf
- Carter, G. L., Willcox, C. H., Lewin, T. J., Conrad, A. M., & Bendit, N. (2010). Hunter DBT project: Randomized controlled trial of dialectical behaviour therapy in women with borderline personality disorder. *Australian and New Zealand Journal of Psychiatry*, 44(2), 162–173. doi:10.3109/00048670903393621
- Chanen, A. M., Betts, J. K., Jackson, H., Cotton, S. M., Gleeson, J., Davey, C. G., ... Mccutcheon, L. (2022). Effect of 3 forms of early intervention for young people with borderline personality disorder: The MOBY randomized clinical trial. *JAMA Psychiatry*, 79(2), 109–119. doi:10.1001/jamapsychiatry.2021.3637
- Cottraux, J., Boutitie, F., Milliery, M., Genouihlac, V., Yao, S. N., Mollard, E., ... Gueyffier, F. (2009). Cognitive therapy versus Rogerian supportive therapy in borderline personality disorder. *Psychotherapy and Psychosomatics*, 78(5), 307–316. doi:10.1159/000229769
- Crawford, M. J., Price, K., Gordon, F., Jossen, M., Taylor, B., Bateman, A., ... Moran, P. (2009). Engagement and retention in specialist services for people with personality disorder. *Acta Psychiatrica Scandinavica*, 119(4), 304–311. doi:10.1111/j.1600-0447.2008.01306.x
- Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated:



- A meta-analysis of study quality and effect size. *Psychological Medicine*, 40(2), 211–223. doi:10.1017/S0033291709006114
- de Klerk, N., Abma, T. A., Bamelis, L. L., & Arntz, A. (2017). Schema therapy for personality disorders: A qualitative study of patients' and therapists' perspectives. *Behavioural and Cognitive Psychotherapy*, 45(1), 31–45. doi:10.1017/S1352465816000357
- Doering, S., Hörz, S., Rentrop, M., Fischer-Kern, M., Schuster, P., Benecke, C., ... Buchheim, P. (2010). Transference-focused psychotherapy v. treatment by community psychotherapists for borderline personality disorder: randomised controlled trial. *The British Journal of Psychiatry*, 196(5), 389–395. doi:10.1192/bjp.bp.109.070177
- Eager, C., & Roy, J. (2017). Mixed Effects Models are Sometimes Terrible. arXiv: 1701. 04858 [stat]. Retrieved July 3, 2017, from <https://arxiv.org/pdf/1701.04858.pdf>.
- Edlund, M. J., Wang, P. S., Berglund, P. A., Katz, S. J., Lin, E., & Kessler, R. C. (2002). Dropping out of mental health treatment: Patterns and predictors among epidemiological survey respondents in the United States and Ontario. *American Journal of Psychiatry*, 159(5), 845–851. doi:10.1176/appi.ajp.159.5.845
- 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.
- Fitzpatrick, S., Bailey, K., & Rizvi, S. L. (2020). Changes in emotions over the course of dialectical behavior therapy and the moderating role of depression, anxiety, and posttraumatic stress disorder. *Behavior Therapy*, 51(6), 946–957. doi:10.1016/j.beth.2019.12.009
- Gaglia, A., Essletzbichler, J., Barnicot, K., Bhatti, N., & Priebe, S. (2013). Dropping out of dialectical behaviour therapy in the NHS: The role of care coordination. *The Psychiatrist Online*, 37(8), 267–271. doi:10.1192/pb.bp.112.041251
- Gunderson, J. G., Prank, A. F., Ronningstam, E. F., Wachter, S., Lynch, V. J., & Wolf, P. J. (1989). Early discontinuance of borderline patients from psychotherapy. *The Journal of Nervous and Mental Disease*, 177(1), 38–42. doi:10.1097/00005053-198901000-00006
- Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, 3(4), 486–504.
- Iliakis, E. A., Ilagan, G. S., & Choi-Kain, L. W. (2021). Dropout rates from psychotherapy trials for borderline personality disorder: A meta-analysis. *Personality Disorders: Theory, Research, and Treatment*, 12(3), 193–206. doi:10.1037/per0000453
- International Business Machines (IBM) (2021). *IBM SPSS Statistics for Windows, Version 28.0*. Armonk, NY: IBM Corp.
- Katsakou, C., Marougka, S., Barnicot, K., Savill, M., White, H., Lockwood, K., & Priebe, S. (2012). Recovery in borderline personality disorder (BPD): A qualitative study of service users' perspectives. *PLoS ONE*, 7(5), e36517. doi:10.1371/journal.pone.0036517
- 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. doi:10.1521/pedi.2011.25.4.432
- Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M. M., & Bohus, M. (2004). Borderline personality disorder. *The Lancet*, 364(9432), 453–461. doi:10.1016/S0140-6736(04)16770-6
- Linehan, M. M., Comtois, K. A., Murray, A. M., Brown, M. Z., Gallop, R. J., Heard, H. L., ... Lindenboim, N. (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*, 63(7), 757–766. doi:10.1001/archpsyc.63.7.757
- Linehan, M. M., Comtois, K. A., Murray, A. M., Brown, M. Z., Gallop, R. J., Heard, H. L., ... Lindenboim, N. (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*, 63(7), 757–766. doi:10.1001/archpsyc.63.7.757
- Linehan, M. M., Korslund, K. E., Harned, M. S., Gallop, R. J., Lungu, A., Neacsiu, A. D., ... Murray-Gregory, A. M. (2015). Dialectical behavior therapy for high suicide risk in individuals with borderline personality disorder: A randomized clinical trial and component analysis. *JAMA Psychiatry*, 72(5), 475–482. doi:10.1001/jamapsychiatry.2014.3039
- Löffler-Stastka, H., Ponocny-Seliger, E., Meißel, T., & Springer-Kremser, M. (2006). Gender aspects in the planning of psychotherapy for borderline personality disorder. *Wiener klinische Wochenschrift*, 118(5), 160–169. doi:10.1007/s00508-006-0573-6
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49, 1494–1502. doi:10.3758/s13428-016-0809-y
- MacNair, R. R., & Corazzini, J. G. (1994). Client factors influencing group therapy dropout. *Psychotherapy: Theory, Research, Practice, Training*, 31, 352–362. doi:10.1037/H0090226
- Majdara, E., Rahimian-Boogar, I., Talepasand, S., & Gregory, R. J. (2021). Dynamic deconstructive psychotherapy in Iran: A randomized controlled trial with follow-up for borderline personality disorder. *Psychoanalytic Psychology*, 38(4), 328–335. doi:10.1177/0003065119891390
- McMain, S. F., Links, P. S., Gnam, W. H., Guimond, T., Cardish, R. J., Korman, L., & Streiner, D. L. (2009). A randomized trial of dialectical behavior therapy versus general psychiatric management for borderline personality disorder. *American Journal of Psychiatry*, 166(12), 1365–1374. doi:10.1176/appi.ajp.2009.09010039
- 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(3), 277–287. doi:10.1016/j.cpr.2009.12.002
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. doi: 10.1371/journal.pmed.1000097
- Morey, L. C., Lowmaster, S. E., & Hopwood, C. J. (2010). A pilot study of manual-assisted cognitive therapy with a therapeutic assessment augmentation for borderline personality disorder. *Psychiatry Research*, 178(3), 531–535. doi:10.1016/j.psychres.2010.04.055
- Paris, J. (2010). Effectiveness of different psychotherapy approaches in the treatment of borderline personality disorder. *Current Psychiatry Reports*, 12(1), 56–60. doi:10.1007/s11920-009-0083-0
- Priebe, S., Bhatti, N., Barnicot, K., Bremner, S., Gaglia, A., Katsakou, C., ... Zinkler, M. (2012). Effectiveness and cost-effectiveness of dialectical behaviour therapy for self-harming patients with personality disorder: A pragmatic randomised controlled trial. *Psychotherapy and Psychosomatics*, 81(6), 356–365. doi:10.1159/000338897
- Priebe, S., Bhatti, N., Barnicot, K., Bremner, S., Gaglia, A., Katsakou, C., ... Zinkler, M. (2012). Effectiveness and cost-effectiveness of dialectical behaviour therapy for self-harming patients with personality disorder: A pragmatic randomised controlled trial. *Psychotherapy and Psychosomatics*, 81(6), 356–365. doi:10.1159/000338897
- Rameckers, S. A., Verhoef, R. E., Grasman, R. P., Cox, W. R., van Emmerik, A. A., Engelmoer, I. M., & Arntz, A. (2021). Effectiveness of psychological treatments for borderline personality disorder and predictors of treatment outcomes: A multivariate multilevel meta-analysis of data from all design types. *Journal of Clinical Medicine*, 10(23), 5622. doi:10.3390/jcm10235622
- Sachdeva, S., Goldman, G., Mustata, G., Deranja, J. E., & Gregory, R. J. (2013). Naturalistic outcomes of evidence-based therapies for borderline personality disorder at a university clinic: A quasi-randomized trial. *Journal of the American Psychoanalytic Association*, 61(3), 578–584. doi:10.1177/0003065113490637
- Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, c332. doi:10.1136/bmj.c332
- Sinnaeve, R., van den Bosch, L. M. C., Hakkaart-van Roijen, L., & Vansteelandt, K. (2018). Effectiveness of step-down versus outpatient dialectical behaviour therapy for patients with severe levels of borderline personality disorder: A pragmatic randomized controlled trial. *Borderline Personality Disorder and Emotion Dysregulation*, 5(1), 12. doi:10.1186/s40479-018-0089-5
- Skodol, A. E., Buckley, P., & Charles, E. (1983). Is there a characteristic pattern to the treatment history of clinical outpatients with borderline personality disorder? *The Journal of Nervous and Mental Disease*, 171(7), 405–410. doi:10.1097/00005053-198307000-00003
- Soler, J., Pascual, J. C., Tiana, T., Cebrià, A., Barrachina, J., Campins, M. J., ... Pérez, V. (2009). Dialectical behaviour therapy skills training compared to standard group therapy in borderline personality disorder: A 3-month



- randomised controlled clinical trial. *Behaviour Research and Therapy*, 47(5), 353–358. doi:10.1016/j.brat.2009.01.013
- Spinhoven, P., Giesen-Bloo, J., van Dyck, R., Kooiman, K., & Arntz, A. (2007). The therapeutic alliance in schema-focused therapy and transference-focused psychotherapy for borderline personality disorder. *Journal of Consulting and Clinical Psychology*, 75(1), 104–115. doi:10.1037/0022-006X.75.1.104
- Stoffers-Winterling, J. M., Völlm, B. A., Rucker, G., Timmer, A., Huband, N., & Lieb, K. (2012). Psychological therapies for people with borderline personality disorder. *Cochrane Database of Systematic Reviews* 2012(8), Art.No.:CD005652, 1–259. 10.1002/14651858.CD005652.pub2
- Storebø, O. J., Stoffers-Winterling, J. M., Völlm, B. A., Kongerslev, M. T., Mattivi, J. T., Jørgensen, M. S., ... Simonsen, E. (2020). Psychological therapies for people with borderline personality disorder. *Cochrane Database of Systematic Reviews*, 5, CD012955. doi:10.1002/14651858.CD012955.pub2
- Torgersen, S. (2012). Epidemiology. In T. A. Widiger (Ed.), *The Oxford handbook of personality disorders* (pp. 186–205). Oxford/New York: Oxford University Press.
- van Asselt, A. D. I., Dirksen, C. D., Arntz, A., & Severens, J. L. (2007). The cost of borderline personality disorder: Societal cost of illness in BPD-patients. *European Psychiatry*, 22(6), 354–361. doi:10.1016/j.eurpsy.2007.04.001
- 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. *The British Journal of Psychiatry*, 182(2), 135–140. doi:10.1192/bjp.182.2.135
- Wagner, T., Assmann, N., Köhne, S., Schaich, A., Alvarez-Fischer, D., Borgwardt, S., ... Fassbinder, E. (2022). The societal cost of treatment-seeking patients with borderline personality disorder in Germany. *European Archives of Psychiatry and Clinical Neuroscience*, 272, 741–752. doi:10.1007/s00406-021-01332-1
- Waldinger, R. J., & Gunderson, J. G. (1984). Completed psychotherapies with borderline patients. *American Journal of Psychotherapy*, 38(2), 190–202. doi:10.1176/appi.psychotherapy.1984.38.2.190
- Walton, C. J., Bendit, N., Baker, A. L., Carter, G. L., & Lewin, T. J. (2020). A randomised trial of dialectical behaviour therapy and the conversational model for the treatment of borderline personality disorder with recent suicidal and/or non-suicidal self-injury: An effectiveness study in an Australian public mental health service. *Australian and New Zealand Journal of Psychiatry*, 54(10), 1020–1034. doi:10.1177/0004867420931164
- Yalom, I. D. (1966). A study of group therapy dropouts. *Archives of General Psychiatry*, 14(4), 393–414. doi:10.1001/archpsyc.1966.01730100057008