The status of psychodynamic psychotherapy as an empirically supported treatment for common mental disorders – an umbrella review based on updated criteria

Falk Leichsenring[1,2], Allan Abbass[3], Nikolas Heim[4], John R. Keefe[5], Steve Kisely[6], Patrick Luyten[7,8], Sven Rabung[9], Christiane Steinert[1,4]

[1]Department of Psychosomatics and Psychotherapy, University of Giessen, Giessen, Germany; [2]Department of Psychosomatics and Psychotherapy, University of Rostock, Rostock, Germany; [3]Dalhousie University, Halifax, NS, Canada; [4]International Psychoanalytic University, Berlin, Germany; [5]Department of Psychiatry and Behavioral Sciences, Albert Einstein College of Medicine, Bronx, New York, NY, USA; [6]School of Medicine, University of Queensland, Brisbane, QLD, Australia; [7]Faculty of Psychology and Educational Sciences, University of Leuven, Leuven, Belgium; [8]Research Department of Clinical, Educational and Health Psychology, University College London, London, UK; [9]Department of Psychology, University of Klagenfurt, Klagenfurt, Austria

To assess the current status of psychodynamic therapy (PDT) as an empirically supported treatment (EST), we carried out a pre-registered systematic umbrella review addressing the evidence for PDT in common mental disorders in adults, based on an updated model for ESTs. Following this model, we focused on meta-analyses of randomized controlled trials (RCTs) published in the past two years to assess efficacy. In addition, we reviewed the evidence on effectiveness, cost-effectiveness and mechanisms of change. Meta-analyses were evaluated by at least two raters using the proposed updated criteria, i.e. effect sizes, risk of bias, inconsistency, indirectness, imprecision, publication bias, treatment fidelity, and their quality as well as that of primary studies. To assess the quality of evidence we applied the GRADE system. A systematic search identified recent meta-analyses on the efficacy of PDT in depressive, anxiety, personality and somatic symptom disorders. High quality evidence in depressive and somatic symptom disorders and moderate quality evidence in anxiety and personality disorders showed that PDT is superior to (inactive and active) control conditions in reducing target symptoms with clinically meaningful effect sizes. Moderate quality evidence suggests that PDT is as efficacious as other active therapies in these disorders. The benefits of PDT outweigh its costs and harms. Furthermore, evidence was found for long-term effects, improving functioning, effectiveness, cost-effectiveness and mechanisms of change in the aforementioned disorders. Some limitations in specific research areas exist, such as risk of bias and imprecision, which are, however, comparable to those of other evidence-based psychotherapies. Thus, according to the updated EST model, PDT proved to be an empirically-supported treatment for common mental disorders. Of the three options for recommendation provided by the updated model (i.e., "very strong", "strong" or "weak"), the new EST criteria suggest that a strong recommendation for treating the aforementioned mental disorders with PDT is the most appropriate option. In conclusion, PDT represents an evidence-based psychotherapy. This is clinically important since no single therapeutic approach fits all psychiatric patients, as shown by the limited success rates across all evidence-based treatments.

Key words: Psychodynamic therapy, psychotherapies, empirically supported treatments, evidence-based medicine, depressive disorders, anxiety disorders, personality disorders, somatic symptom disorders

(World Psychiatry 2023;22:286–304)

More than 20 years ago, criteria for empirically supported psychotherapeutic treatments (ESTs) were first proposed[1,2]. These criteria suggested that at least two randomized controlled trials (RCTs) from independent research groups were required to demonstrate that a manual-guided treatment was superior to control conditions, or as efficacious as an already established treatment, in a specific mental disorder[1].

However, concerns were raised about those criteria. They included the exclusive focus on symptom improvement while neglecting psychosocial functioning, the limited generalizability of results from research settings to clinical practice, the neglect of design flaws and researcher allegiance, and the fact that only two RCTs were required to demonstrate efficacy[3]. Furthermore, an independent empirical re-evaluation of the studies included in the American Psychological Association's database of ESTs found replicability and power estimates to be low across almost all ESTs[4]. Some ESTs rated as having "strong" evidence according to the model failed to outperform their "modest" counterparts with regard to efficacy[4].

As a result, a new model for ESTs has been proposed, taking these concerns into account[3]. This model requires a focus on: a) systematic (quantitative) reviews (meta-analyses) rather than individual studies, b) study quality, c) clinical significance in addition to statistical significance, d) long-term outcomes in addition to short-term efficacy, e) functional or other health-related outcomes in addition to symptom improvement, f) generalization to non-research settings, g) de-emphasizing categorical diagnoses and emphasizing syndromes and diagnostically complex patients, and h) mechanisms of psychopathology and therapeutic change[3].

For candidate treatments, the new EST model suggests the use of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system by an expert committee, to assess the quality of evidence and the degree to which benefits exceed potential harms[5-8]. The original GRADE system allows to rate the evidence as "high quality", "moderate quality", "low quality" or "very low quality"[5-9]. If there are differences in ratings of evidence between primary (critical) outcomes and other outcomes (e.g., side effects or costs), GRADE regards efficacy outcomes as the most important on most occasions, and suggests that guideline panels can base their rating of the quality of evidence exclusively on data on efficacy[7].

For high quality evidence, the new EST model requires a wide range of studies with no major limitations, small heterogeneity and narrow confidence intervals (CIs)[3]. These recommendations differ considerably from the original approach of the GRADE

group, which considered "one or more well-designed RCTs yielding consistent directly applicable results" as required for high quality evidence[7]. Moderate quality evidence is defined by the updated EST model as "a few" studies, of which some have limitations, but no major flaws, with some variation between studies or a wide CI for the summary estimate[3]. Again, this recommendation differs from the original approach of the GRADE group, which defined moderate quality evidence for RCTs in terms of "important" limitations[7]. Low quality evidence was originally restricted by the GRADE group as referring to observational studies and only occasionally to RCTs with multiple serious limitations[7], whereas the newly proposed EST criteria define low quality evidence as referring to "studies" (no specification if RCTs or observational studies) with major flaws, or where there are important variations between studies and very wide CIs for the summary estimate[3].

In a next step, the original GRADE system results in "strong" or "weak" recommendations for a treatment[5-7]. In the new EST model, a third category was introduced, i.e. a "very strong" recommendation[3]. Additional contextual factors may increase or decrease the GRADE recommendations (e.g., comparative efficacy to other treatments, evidence for mechanisms of change, evidence for efficacy in minorities or across various patient sub-populations)[3].

Based on the original EST criteria[1], the empirical status of psychodynamic therapy (PDT) has been assessed in several reviews[10-14]. The revised EST criteria, however, have not yet been applied to studies available for PDT. Nevertheless, as pointed out by the Task Force on Promotion and Dissemination of Psychological Procedures, it is critical to investigate whether PDT fulfills the updated criteria, "if this clinically verified treatment is to survive in today's market"[15]. For this reason, we carried out an umbrella review of meta-analyses of PDT in common mental disorders in adults applying the revised EST criteria.

METHODS

Details of the procedures were described in a study protocol[16], which was also pre-registered (PROSPERO: CRD42022342350).

The authors of this review fulfil the criteria proposed by the new EST model[3], that is: a) a broad range of documented expertise, b) disclosure of actual and potential conflicts of interest (see supplementary information), c) maintaining a climate of openness, d) using clearly defined procedures and methods as described in the study protocol.

Definition of psychodynamic psychotherapy

PDT includes a family of psychotherapeutic approaches having in common a focus on the identification of recurring patterns of relating to the self and others (including the therapeutic relationship) and of expression of emotion, the exploration of defensive patterns, and the discussion of past experiences that have an impact on the person's present experiences[17].

PDT operates on a supportive-interpretive continuum[17]. The use of more interpretive or supportive interventions depends on the person's needs and mental capacities[17-19]. While interpretive interventions enhance the person's insight about repetitive conflicts sustaining his/her problems, supportive interventions aim to strengthen psychosocial abilities ("ego-functions") that are currently not accessible to the person.

Similarity

The treatments included in a meta-analysis are required to show sufficient similarity[20], a criterion adopted by the new EST model[3]. For many variants of PDT, the commonalities in techniques have been shown to outweigh the differences, allowing for the development of unified protocols[21-25]. Unified psychodynamic protocols focus on shared ingredients or mechanisms, representing a "mechanism-oriented approach"[21-23]. An analogous approach has been developed in the area of cognitive behavior therapy (CBT)[26,27]. Indeed, the updated EST model encourages a focus on core dimensions of pathology and treatments which may reduce "the EST movement's reliance on a large number of treatment manuals" and lead to a much simpler and "more practitioner-friendly system"[3]. For each mental disorder included in this umbrella review, we tested whether the applied PDT techniques showed sufficient similarity.

Conditions being studied

The following mental disorders in adults, defined according to the ICD or DSM, were eligible for inclusion in this umbrella review: depressive disorders, anxiety disorders, trauma- and stressor-related disorders, dissociative disorders, obsessive-compulsive disorder, eating disorders, somatic symptom disorders, attention-deficit/hyperactivity disorder, substance related disorders, personality disorders, bipolar disorder, schizophrenia spectrum disorders. In addition, complex mental disorders – defined as chronic disorders, highly comorbid disorders, and disorders associated with personality disorders – were also included.

Inclusion criteria

Following the revised EST criteria[3], when evaluating PDT efficacy, we focused on meta-analyses of RCTs in common mental disorders in adults published in the past two years. Older reviews were only included if they provided data not available in more recent reviews, e.g. results on specific domains such as functioning.

Meta-analyses were included which tested PDT against a control condition (e.g., waiting list, treatment-as-usual, TAU; pill or psychological placebo), or against pharmacotherapy or another form of psychotherapy[3]. Results were evaluated per comparison condition, divided into all controls, active controls (e.g., TAU, enhanced

TAU, low intensity therapy), and active therapies (e.g., pharmacotherapy or another form of psychotherapy). If several meta-analyses for one disorder were available, we included the largest, that is the one encompassing most RCTs.

Furthermore, systematic reviews focusing on mechanisms of change of PDT were evaluated, including both RCTs and open studies if they showed all characteristics of RCTs (e.g., treatment manuals, valid assessment of disorder and outcome) with the exception of not including a control condition. In addition, as suggested by the new EST model, effectiveness studies carried out under real-world conditions, as well as cost-effectiveness studies, were evaluated[3]. We included individual studies if no systematic reviews were available for a specific area of research, or if a recent study was not included in available systematic reviews.

Outcomes

As primary (critical) outcome, we used effect sizes in disorder-specific target symptoms post-therapy assessed by validated scales. In addition to statistical significance, clinical significance of effect sizes was assessed. If presented by the authors of the included meta-analyses, data of high-quality studies and data corrected for publication bias or outliers were preferably included. Through this paper, a negative effect size indicates superiority of PDT.

Secondary (important but not critical) outcomes assessed, if available, were adverse events, improvement in functioning, effectiveness under real-world conditions, cost-effectiveness, and impact on minorities.

Search for studies

We searched PubMed and PsycINFO and individual records of the Cochrane Library for systematic reviews, meta-analyses and individual RCTs on the efficacy of PDT in common mental disorders in adults published between 2012 and December 2022. We aimed to focus on meta-analyses published in the past two years, as required by the new EST criteria. However, we allowed inclusion of older reviews providing results not included in more recent ones.

Search terms were (meta-analy* or metaanaly*) and ("psychodynamic therapy" or "dynamic therapy" or "psychoanalytic therapy" or "psychodynamic psychotherapy" or "dynamic psychotherapy" or "psychoanalytic psychotherapy"). Additionally, a regularly updated comprehensive list containing RCTs of psychodynamic treatments was consulted (researchgatenet/publication/317335876) and a hand search in journal papers and textbooks was carried out. Studies on face-to-face and Internet PDT were included, as well as studies on individual and group therapy.

Furthermore, we searched for systematic reviews and individual studies on mechanisms of change in PDT, for effectiveness studies carried out under real-world conditions, and for studies on cost-effectiveness of PDT[3]. For these purposes, additional search terms were "mechanisms of change", "curative factors", "process-outcome", "cost-effectiveness", and "health economic analysis".

At least two reviewers independently screened the results of the database search for relevant meta-analyses and individual studies. If the title and abstract of a paper contained sufficient information to determine that it did not meet the inclusion criteria specified above, the paper was excluded. In a next step, full texts of all studies possibly relevant for inclusion were retrieved. Disagreements about the inclusion of a meta-analysis or study were solved by consensus or by consulting a third expert. The search results for meta-analyses were documented in a PRISMA flow chart (see Figure 1).

Data extraction

A data extraction form was used to retrieve details of included meta-analyses. A similar form was used for individual studies. At least two authors independently extracted the results: type of disorder, number of included RCTs, number of participants, risk of bias, effect sizes, 95% CI, heterogeneity, and adverse events. Discrepancies were solved by consensus. These procedures were applied to all ratings, including assessment of risk of bias, treatment fidelity, quality of meta-analyses and GRADE. We contacted the authors of the included meta-analyses for additional information.

Quality of meta-analyses and primary studies

For rating the quality of included meta-analyses, we applied the Joanna Briggs Institute Critical Appraisal Checklist for Systematic Reviews and Research Syntheses[28]. We used the first nine items which refer to quality, complemented by item 12 of AMSTAR 2 ("Was the impact of risk of bias in individual studies on results of the meta-analysis taken into account?")[29] and an additional item addressing whether the meta-analysis was preregistered.

If data on risk of bias were not reported in the included meta-analyses, we rated this risk for the included studies using the four criteria of the Cochrane Risk of Bias Tool[30] (adequate random sequence generation, allocation concealment, blinding of assessors and/or use of self-report measures only, and use of intent-to-treat analysis).

As to the quality of primary studies, we used ratings based on the Randomized Controlled Trial Psychotherapy Quality Rating Scale (RCT-PQRS)[31]. Treatment fidelity was assessed following criteria proposed by the new EST model[3] (i.e., use of treatment manuals, experienced/qualified therapists, monitoring of treatment during the trial, and empirical assessment of treatment integrity). The quality of studies on mechanisms of change was evaluated as proposed by Crits-Christoph and Connolly Gibbons[32].

Data synthesis

We used the criteria of the new EST model[3] to evaluate the empirical status of PDT in each of the mental disorders. Following

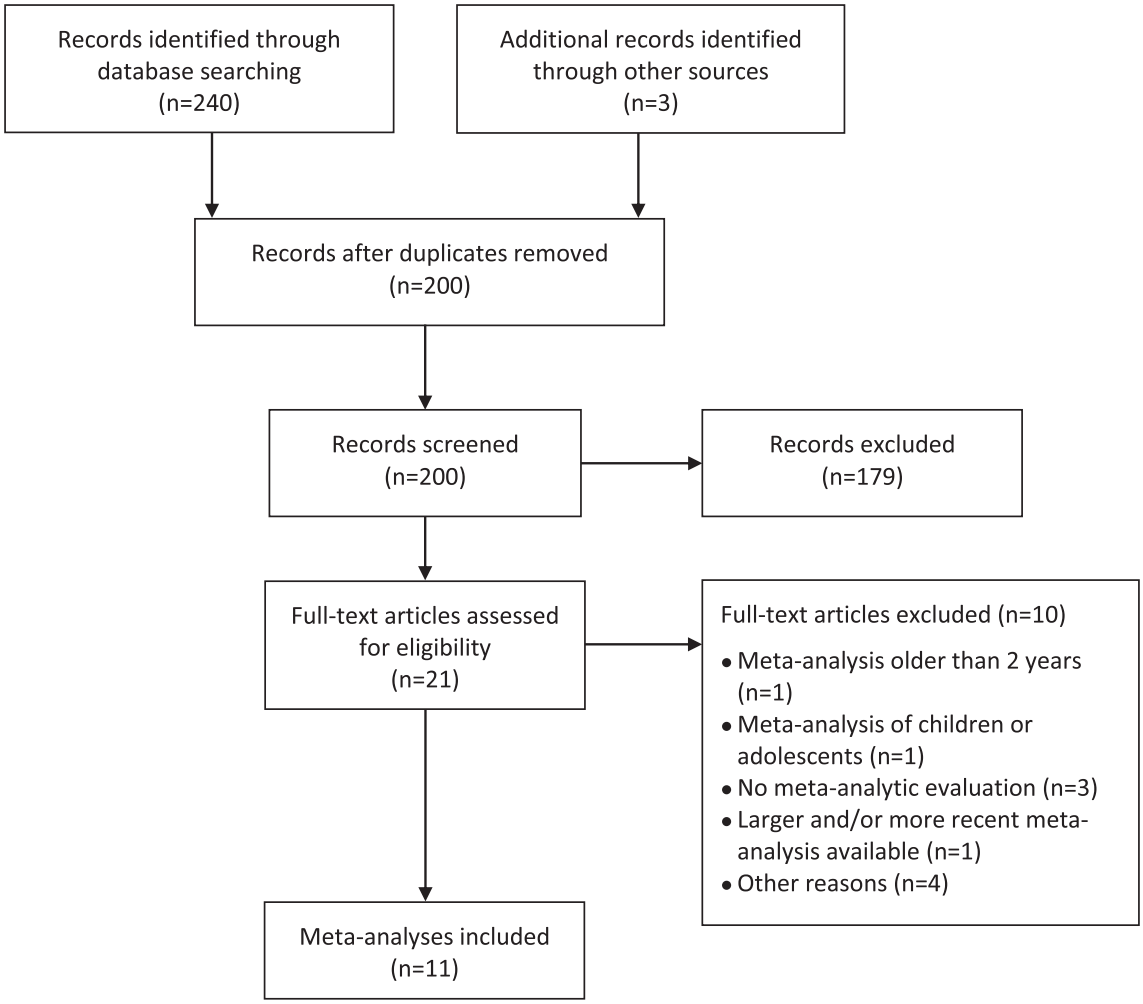

Figure 1 PRISMA flow chart

GRADE, we first identified critical (primary) and important (secondary) outcomes[33]. As critical outcomes, we defined disorder-specific symptom severity at treatment termination in comparison to control conditions or to active therapies. As important outcomes, we defined treatment effects at follow-ups, improvements in functioning, costs, frequency of adverse events, and data on mechanisms of change.

In a second step, following GRADE, we rated the quality of evidence for each outcome, taking risk of bias, inconsistency, imprecision, indirectness and publication bias into account[7]. As to inconsistency, we regarded I^2 values of 25%, 50% and 75% as indicating low, moderate or high heterogeneity[34], and took low and moderate heterogeneity as indicating no serious inconsistency. Indirectness encompassed deviations in patients, outcomes or treatments from those of interest, as well as indirect comparisons (e.g., comparing A and B with placebo without directly comparing A and B)[7,35]. As to risk of bias, a GRADE rating of high quality evidence could only be achieved if more than 50% of the studies were at low risk with regard to random sequence generation, allocation concealment, blinding of assessors (or use of self-report instruments only) and completeness of data (intention-to-treat analysis)[35]. For imprecision, we followed the GRADE guidelines, which suggest that the effect size needs to be statistically significant and the total sample size has to exceed the "optimal information size" (OIS), that is the sample size required to detect a clinically meaningful effect size with a power of 0.80 at α=0.05[36,37]. We also tested whether the recommendations would differ if the upper or lower boundaries of the CIs represented the truth.

We finally graded the evidence and assessed the strengths of treatment recommendations[3,8].

RESULTS

The initial search yielded 243 hits (see Figure 1). In total, eleven meta-analyses were included. Four recent meta-analyses addressing the efficacy of PDT fulfilled the inclusion criteria, referring to depressive, anxiety, personality and somatic symptom disorders[38-40]. Two older reviews fulfilling inclusion criteria which assessed the efficacy of PDT were also included[41,42]. Further included meta-analyses ad-

dressed the efficacy of Internet-delivered PDT[43], the efficacy of adding short-term PDT to antidepressants in depression[44], the effectiveness of PDT under real-world conditions[45], the mechanisms of change[46], and the quality of RCTs of PDT and CBT[47].

Depressive disorders

The eligible recent meta-analysis assessing the efficacy of PDT for depressive disorders, in comparison with control conditions or active therapies, included 27 RCTs (N=3,163 patients)[38]. There was sufficient similarity in the applied techniques to assume that the different studies tested the same essential treatment[21,22].

Efficacy of PDT vs. control conditions

PDT was found to be superior to all control conditions in improving depressive symptoms, with a medium effect size (g=–0.58, 95% CI: –0.33 to –0.83, n=12, I^2=63%, N=1,017) and no evidence for publication bias (see Table 1). Compared to waiting list controls only, the effect size was large (g=–1.14, 95% CI: –1.66 to –0.62, n=3, N=115), while it was medium compared to active controls (g=–0.51, 95% CI: –0.68 to –0.35, I^2=26%, n=9, N=945).

The effect size of –0.58 in comparison to all control conditions corresponds to a difference in success rates of about 33%, or a number needed to treat of about 3[48], clearly exceeding the threshold of a clinically significant effect size of d=±0.24 proposed by Cuijpers et al[49]. This is also true for the effect size of –0.51 achieved in comparison to active controls, which also compares favorably to those found in the largest meta-analyses of psychotherapy (0.31) and pharmacotherapy (0.30) for depressive disorders in comparison to TAU or placebo[50-52].

The above-mentioned threshold of a clinically significant effect (d=±0.24)[49] resulted in an OIS of 432[53]. For PDT vs. all control conditions, the effect size was significant and the sample size exceeded the OIS (N=1,017 > 432), thus indicating no serious imprecision. The lower boundary of the CI (–0.83) represents a large effect size, and the upper boundary (–0.33) exceeds –0.24, thus representing a small but still clinically meaningful effect size. For PDT vs. active controls, there was no serious imprecision (N=945 > 432), the lower boundary of the CI (–0.68) representing a medium to large effect size, while the upper boundary (–0.35) exceeded –0.24, thus representing a small but still clinically meaningful effect size. The width of the CI for comparison with controls is similar to other active therapies, such as CBT vs. TAU (see supplementary information).

There were no indications of serious indirectness with regard to patients, treatment outcomes, or comparisons.

Efficacy of PDT vs. active therapies

Compared to other active therapies, PDT did not differ significantly on the primary outcome, i.e., severity of depression (g=–0.01, 95% CI: –0.34 to 0.32, n=20, N=2,335). Heterogeneity was high (I^2=90%), due to one outlier[54]. When this was removed, heterogeneity was reduced to a moderate level (g=0.10, 95% CI: –0.06 to 0.26, I^2=62%, n=19, N=2,154). Correction for publication bias in the reduced sample did not affect the results (g=–0.03, 95% CI: –0.23 to 0.17, I^2=73%) (see Table 1).

The corrected effect size was not significant, and the sample size exceeded the OIS (N=2,154 > 432), thus indicating no serious imprecision. The CI of the corrected effect size did not exceed ±0.24, indicating no clinically meaningful difference in efficacy compared to other active therapies. Both the upper and the lower boundary of the CI represent small, clinically not meaningful effect sizes. Heterogeneity was moderate.

In follow-ups ranging from 2 to 55 months, the difference between PDT and active therapies remained insignificant (g=–0.01, 95% CI: –0.31 to 0.29, n=10, I^2=71%). After removing one outlier[55], heterogeneity was reduced (g=0.08, 95% CI: –0.14 to 0.30, I^2=50%, n=9, N=1,096). The effect size was below 0.24 and the sample size exceeded the OIS (N=693 > 432), thus indicating no serious imprecision. Both the upper and the lower boundary of the CI represent a small effect size. The upper boundary, however, exceeded 0.24.

Quality measures

The quality of the eligible meta-analysis was found to be good, with 10 out of the 11 relevant items fulfilled. The quality of primary studies, as assessed by the RCT-PQRS, was sufficient (total score ≥ 24) for most studies (74%).

As to treatment fidelity, most of the 27 included studies used a treatment manual (87%), included experienced/qualified therapists (91%), monitored the treatment during the trial by supervision (59%), and assessed treatment integrity empirically (57%).

Adequate random sequence generation, allocation concealment, blinding of assessors (or use of only self-report measures) and intention-to-treat analysis were applied in 56%, 48%, 74% and 67% of the studies, respectively, indicating that most studies showed a low risk of bias (see also supplementary information). The corresponding values for the comparison with all control conditions were 54%, 54%, 54%, and 85%; those for the comparison with active controls only were 67%, 78%, 78% and 100%; and those for the comparison with active therapies were 50%, 40%, 75% and 55%, respectively.

Secondary outcomes

Several studies covered in the recent included meta-analysis[38] reported data on tolerability, detecting no or only a few adverse events.

Improvement in functioning was not assessed in that meta-analysis. An earlier meta-analysis[42] found PDT to be superior to control conditions with regard to improving quality of life, with a medium effect size (d=–0.49, 95% CI: –0.73 to –0.24, n=3, I^2=0%, N=293), while there was no difference compared to other psychotherapies in improving interpersonal functioning, either post-

Table 1 Synthesis of evidence profiles for psychodynamic therapy (PDT) in depressive disorders

Outcomes	Quality assessment								Summary of findings		Quality of evidence
	Quality of systematic review (QSR)	Quality of studies (QS) (RCT-PQRS ≥ 24)	Treatment fidelity (FI)	GRADE				Publication bias	N patients (both arms)	Effect size (95% CI)	GRADE QSR, QS, FI Total rating
				Risk of bias	Inconsistency	Indirectness	Imprecision				
PDT vs. all controls											
Severity of depressive symptoms (critical) 12 RCTs	10/11	74%	+	No	No	No	No	Undetected	1,017	–0.58 (–0.33, –0.83) I^2=63%	++++ +++ High
PDT vs. active controls											
Severity of depressive symptoms (critical) 9 RCTs	10/11	100%	+	No	No	No	No	Undetected	945	–0.51 (–0.68, –0.35) I^2=26%	++++ +++ High
Functioning (important) 3 RCTs	9/11	100%	+	No	No	No	Yes	Not assessed	293	–0.49 (–0.73, –0.24) I^2=0%	+++– +++ Moderate
PDT vs. active therapies											
Severity of depressive symptoms (critical) 19 RCTs (1 outlier excluded)	9/11	68%	+	Yes	No	No	No	Corrected	2,154	–0.03 (–0.23, 0.17) I^2=73%	–+++ +++ Moderate
Follow-up depressive symptoms (important) 9 RCTs	9/11	82%	+	Yes	No	No	No	Undetected	1,096	0.08 (–0.14, 0.30) I^2=50%	–+++ +++ Moderate
Functioning (important) 5 RCTs	9/11	60%	+	Yes	No	No	Yes	Not assessed	408	0.05 (–0.23, 0.34) I^2=40%	–++– +++ Low
Follow-up functioning (important) 4 RCTs	9/11	75%	+	Yes	No	No	Yes	Not assessed	288	–0.15 (–0.70, 0.40) I^2=74%	–++– +++ Low

RCTs – randomized controlled trials; RCT-PQRS – Randomized Controlled Trial Psychotherapy Quality Rating Scale; GRADE – Grading of Recommendations Assessment, Development, and Evaluation

therapy (d=0.05, 95% CI: –0.23 to 0.34, n=5, I^2=40%, N=408) or at follow-up (d=–0.15, 95% CI: –0.70 to 0.40, n=4, I^2=74%, N=288). Using d=±0.24 and N=432 as an OIS resulted in rating of some imprecision in these estimates (N=293, 408, 288 < 432). Risk of bias was low for random sequence generation (100% of studies), allocation concealment (100%), blinding of outcomes (67%) and completeness of data (100%) (see Table 1).

A recent meta-analysis on effectiveness of routinely delivered psychotherapies[45] found large pre-post effect sizes in depression outcomes (d=0.96, 95% CI: 0.88-1.04), with no differences between CBT and PDT (d=–0.07 in favor of PDT). These results were corroborated by a recent effectiveness study on PDT in chronic depression, which found a large effect size (d=–0.90) in comparison to a waiting list condition[56]. As suggested by one RCT, PDT may be a cost-effective intervention in treatment-resistant depressive disorders as compared to TAU[57].

One RCT found gender and racial/ethnic minority status to moderate outcome, with PDT being more efficacious in minority men (primarily African-American) compared to pharmacotherapy and pill placebo[58]. Another RCT conducted in a community setting, including about 50% of patients who identified as a racial/ethnic minority, found PDT to be as efficacious as CBT[59].

Further results

A meta-analysis found PDT combined with antidepressants to be more efficacious than antidepressants with or without brief supportive therapy, with a significant but small effect size post-therapy (g=–0.26; standard error, SE=0.10, p=0.01) and a medium effect size at follow-up (g=–0.50, SE=0.10, p=0.001)[44]. Adequate random sequence generation, allocation concealment, blinding of assessors (or use of only self-report measures) and intention-to-treat analysis were applied in 100%, 100%, 71% and 86% of the studies, respectively.

A meta-analysis of Internet-delivered PDT[43] reported a medium effect size compared to controls in depression outcomes (g= –0.46, 95% CI: –0.73 to –0.19, I^2=23%, n=5, N=359), with two outliers excluded. Risk of bias was low for most studies, and no publication bias was found.

GRADE

According to the results presented above, PTD achieved medium effect sizes compared to both all control conditions (g= –0.58) and active controls (g=–0.51) in the reduction of depressive symptoms, and a small clinically not meaningful effect size compared to other active therapies (g=–0.03). There were no serious indications of inconsistency, indirectness, imprecision or publication bias in critical outcomes (see Table 1). Most studies (74%) showed acceptable quality as assessed by the RCT-PQRS. Treatment fidelity was sufficient for most studies. The quality of the meta-analysis was rated as good. Furthermore, there was a relatively wide range of studies (n=27), with moderate heterogeneity and CIs indicating enough precision. The benefits outweighed the costs and harms, as required by GRADE[6,60].

For comparisons with all controls and active controls, most studies showed a low risk of bias, suggesting high quality evidence (see also supplementary information). For the comparison with active therapies, risk of bias was low in most studies for masking and completeness of data, but not for random sequence generation and allocation concealment (see Table 1 and supplementary information). The GRADE guidelines recommend to be conservative with regard to rating down the quality of evidence[61]. Thus, the review panel decided to downgrade the evidence for PDT vs. active therapies by one level, rating the evidence as moderate, whereas the quality of evidence for PDT vs. all controls and active controls only in depression was rated as high for critical outcomes (see Table 1).

Anxiety disorders

The eligible recent meta-analysis assessing the efficacy of PDT for anxiety disorders, in comparison with control conditions or active therapies, comprised 17 RCTs (N=1,798), including agoraphobia with and without panic disorder, panic disorder, social anxiety disorder, generalized anxiety disorder, and post-traumatic stress disorder (PTSD)[38]. There was sufficient similarity in the applied techniques to assume that the different studies tested the same essential treatment[21,23,24].

Efficacy of PDT vs. control conditions

PDT was found to be superior to all control conditions in reducing anxiety symptoms, with a large effect size (g=–0.94, 95% CI: –1.55 to –0.33, n=7, I^2=78%, N=565). Removing one outlier[62] reduced heterogeneity to a moderate level (g=–0.72, 95% CI: –1.06 to –0.37, n=6, I^2=43%, N=479) (see Table 2). There was no evidence for publication bias. Effect sizes did not significantly differ if control conditions included an active element vs. waiting list alone (p=0.401). For comparison with active controls, only three small RCTs were available; PDT yielded a medium effect size, but the CI was wide (g=–0.64, 95% CI: –1.14 to –0.14, n=3, N=86).

The reported effect size of –0.72 in comparison to all control conditions corresponds to a difference in success rates of 38% or a number needed to treat of 2.6[48]. Thus, it can be considered as clinically meaningful. This is also true for the effect size of –0.64 achieved in comparison to active controls.

We used d=±0.25 as a conservative estimate for a minimum clinically meaningful effect size, similar to the proposed effect size of d=±0.24 for depression, resulting in an OIS of 398[53]. The effect size achieved by PDT in comparison to controls was statistically significant, and the sample size exceeded the OIS (479 > 398), indicating no serious imprecision. The lower boundary of the CI represented a large effect size, while the upper boundary exceeded –0.25, a still clinically meaningful effect size. The width of the CI for comparison with controls is similar to other active therapies, such as CBT vs. TAU or placebo (see supplementary information).

Table 2 Synthesis of evidence profiles for psychodynamic therapy (PDT) in anxiety disorders

Outcomes	Quality assessment								Summary of findings		Quality of evidence
	Quality of systematic review (QSR)	Quality of studies (QS) (RCT-PQRS ≥ 24)	Treatment fidelity (FI)	GRADE				Publication bias	N patients (both arms)	Effect size (95% CI)	GRADE QSR, QS, FI Total rating
				Risk of bias	Inconsistency	Indirectness	Imprecision				
PDT vs. all controls											
Anxiety symptoms (critical) 6 RCTs (1 outlier excluded)	10/11	33%	+	Yes	No	No	No	Undetected	479	–0.72 (–1.06, –0.37) I^2=43%	–+++ +–+ Moderate
PDT vs. active controls											
Anxiety symptoms (critical) 3 RCTs	10/11	33%	–	Yes	No	No	Yes	Not assessed	86	–0.64 (–1.14, –0.14) I^2=0%	–++– +–– Low
PDT vs. active therapies											
Anxiety symptoms (critical) 14 RCTs (1 outlier excluded)	10/11	64%	+	Yes	No	No	No	Undetected	1,196	0.06 (–0.11, 0.23) I^2=45%	–+++ +++ Moderate
Interpersonal functioning (important) 3 RCTs	8/11	100%	+	No	Yes	No	Yes	Not assessed	512	–0.03 (–1.19, 1.14) I^2=77%	+–+– +++ Low
Short-term follow-up anxiety symptoms (important) 9 RCTs (1 outlier excluded)	10/11	56%	+	Yes	No	No	No	Corrected	914	–0.03 (–0.25, 0.19) I^2=46%	–+++ +++ Moderate
Long-term follow-up anxiety symptoms (important) 4 RCTs (1 outlier excluded)	10/11	100%	+	No	No	No	No	Not assessed	617	0.00 (–0.20, 0.20) I^2=17%	++++ +++ High

RCTs – randomized controlled trials; RCT-PQRS – Randomized Controlled Trial Psychotherapy Quality Rating Scale; GRADE – Grading of Recommendations Assessment, Development, and Evaluation

There were no indications of serious indirectness with regard to patients, treatment outcomes, or comparisons.

Efficacy of PDT vs. active therapies

Compared to other active therapies, PDT was not significantly different in anxiety outcomes (g=–0.01, 95% CI: –0.21 to 0.20, n=15, I^2=60%, N=1,242). Excluding one potential outlier[63] reduced heterogeneity (g=0.06, 95% CI: –0.11 to 0.23, n=14, I^2=45%, N=1,196). Evidence for publication bias was not found. There were no significant differences in effect sizes achieved by PDT vs. active therapies in generalized anxiety disorder compared to other anxiety disorders (p=0.181), panic disorder (p=0.356), or social anxiety disorder (p=0.977).

The effect size was not significant and the sample size exceeded the OIS (N=1,196 > 398), indicating no serious imprecision. The corrected effect size and its CI did not exceed ±0.25, indicating no clinically meaningful difference in efficacy compared to other active therapies.

Remission rates for anxiety disorders did not differ significantly between PDT and other active therapies (log odds ratio = 0.12, 95% CI: –0.76 to 0.99, p=0.761).

At follow-up of up to one year after termination, outcomes of PDT did not differ from other active therapies (g=0.08, 95% CI: –0.25 to 0.42, n=10, I^2=73%). Excluding one outlier[64] reduced heterogeneity to a moderate level (g=–0.03, 95% CI: –0.25 to 0.19; n=9, I^2=46%, N=914). At follow-up over more than one year after termination, PDT did not differ from other active therapies either (g=0.21, 95% CI: –0.45 to 0.87, n=5, I^2=85%). When removing one outlier[64], heterogeneity was considerably reduced (g=0.00, 95% CI: –0.20 to 0.20; n=4, I^2=17%, N=617). Both corrected effect sizes were not statistically significant and the sample sizes exceed the OIS, indicating no serious imprecision (914, 617 > 398).

Quality measures

The quality of the eligible meta-analysis was found to be good, with 10 out of the 11 relevant items fulfilled. The quality of primary studies, as assessed by the RCT-PQRS, was sufficient (total score ≥ 24) for most studies (65%). However, for the comparison with all control conditions, only 33% of studies scored ≥ 24, due to inclusion of several older studies. For comparisons with active therapies, the majority of RCTs (64%) were of sufficient quality.

As to treatment fidelity, most of the 17 included studies used a treatment manual (89%), included experienced/qualified therapists (89%), and monitored the treatment during the trial by supervision (72%). Treatment integrity was empirically studied in 33% of studies.

Adequate random sequence generation, allocation concealment, blinding of assessors (or use of only self-report measures) and intention-to-treat analysis were reported in 47%, 41%, 71% and 59% of the studies, respectively (see also supplementary information). The corresponding values for the comparison with all controls were 29%, 29%, 57% and 43%; those for the comparison with active therapies were 47%, 40%, 67% and 60%, respectively.

Secondary outcomes

Several studies covered in the recent included meta-analysis[38] reported data on tolerability, detecting no or only a few adverse events.

Improvement in functioning was not assessed in that meta-analysis. An earlier meta-analysis[41] found no differences between PDT and other psychotherapies in improving interpersonal functioning (g=–0.03, 95% CI: –1.19 to 1.14, n=3, N=512). The number of patients exceeded the OIS (512 > 398), but the CI was wide.

A recent meta-analysis on effectiveness of routinely delivered psychotherapies[45] found large pre-post effect sizes in anxiety outcomes (d=–0.80, 95% CI: 0.71-0.09), with no differences between PDT and CBT (d=0.00). One RCT found no differences in cost-effectiveness between PDT and solution-focused therapy in anxiety disorders[65].

Further results

A meta-analysis on Internet-delivered PDT[43] reported a small effect size compared to control conditions in anxiety outcomes (g=–0.32, 95% CI: –0.55 to –0.09; I^2=0%, n=5, N=359). Risk of bias was low for most studies, and publication bias was not found, although the number of studies was small. An RCT found no differences in outcome between PDT and CBT applied via the Internet in generalized anxiety disorder (0.14, 95% CI: –0.50 to 0.78)[66].

GRADE

According to the results presented above, PDT achieved a medium to large effect size (g=–0.72) compared to all control conditions in the reduction of anxiety symptoms, and a small effect size compared to other active therapies (g=0.06). There were no serious indications of inconsistency, indirectness, imprecision or publication bias in critical outcomes (see Table 2). Most studies (65%) showed acceptable quality as assessed by the RCT-PQRS, except for the comparison with (active and inactive) controls, due to the inclusion of several older studies. Treatment fidelity was sufficient for most studies, except for comparisons with active controls. The quality of the meta-analysis was rated as good. Furthermore, there was a relatively wide range of studies (n=17), with moderate heterogeneity and CIs indicating enough precision, except for comparisons with active controls. The benefits outweighed the costs and harms, as required by GRADE[6,60].

For comparisons with all controls, most studies showed an unclear or high risk of bias in critical outcomes for random sequence generation, allocation concealment and completeness, but not for blinding (see also supplementary information). For the comparison with active therapies, risk of bias was low in most studies for masking and completeness of data, while it was unclear or high for random sequence generation and allocation concealment. As noted above, the GRADE guidelines recommend to be conservative with regard to rating down the quality of evidence[61]. Thus, the review panel decided to downgrade the evidence for PDT in

anxiety disorders by one level, rating the evidence as moderate for critical outcomes. For the comparison with active controls, since the evidence was based on only three small old RCTs of low quality, the review panel rated the quality as low (see Table 2).

Personality disorders

The eligible recent meta-analysis assessing the efficacy of PDT for personality disorders, in comparison with control conditions or active therapies, included 16 RCTs, dealing with borderline or Cluster C personality disorders[38,39]. Although there was more heterogeneity between PDT methods used to treat these disorders compared to depressive and anxiety disorders, they are all based on psychodynamic theory and technique and have core dimensions in common[17,25,67].

Efficacy of PDT vs. control conditions

For core personality disorder symptoms, PDT achieved a medium effect size in comparison to all control conditions (g=–0.63, 95% CI: –0.87 to –0.41, n=5, I^2=11%, N=239) (see Table 3). Compared to active controls, PDT achieved a medium effect size (g=–0.65, 95% CI: –0.99 to –0.32, I^2=15%, n=4, N=200). The number of studies was too small to determine any effect of publication bias.

We used a standardized mean difference (SMD) = ±0.43 as a conservative estimate for a minimum clinically meaningful effect size[72], resulting in an OIS of 136[53]. The effect size of PDT in the reduction of personality disorder symptoms compared to all controls was statistically significant, and the sample size exceeded the OIS (N=239 > 136). Thus, there was no serious imprecision. The lower boundary of the CI represents a large effect size, while the upper boundary is close to –0.43, thus still representing a clinically meaningful effect size. For the comparison with active controls, the sample size (N=200) exceeds the OIS as well, but the lower boundary of the CI is below –0.43.

For suicidality, PDT was superior to active control groups, with a large effect size (g=–0.79, 95% CI: –1.38 to –0.20, n=5, I^2=72%). Removing one outlier[71] reduced heterogeneity to a moderate level and the effect size to medium (g=–0.67, 95% CI: –1.13 to –0.20, n=4, I^2=40%, N=239). We used an SMD of ±0.53 as a minimal clinically meaningful effect size, resulting in an OIS of 90[53]. The sample size exceeds the OIS (N=239 > 90). The effect size and the lower boundary of the CI can be regarded as clinically meaningful, but the upper boundary falls below the margin.

There were no indications of serious indirectness with regard to patients, treatments outcomes, or comparisons.

Efficacy of PDT vs. active therapies

No significant differences between PDT and other active therapies with regard to core personality disorder symptoms were found (g=0.05, 95% CI: –0.25 to 0.35, n=7, I^2=54%, N=473). Removing one possible outlier[69] reduced heterogeneity (g=–0.04, 95% CI: –0.31 to 0.22, n=6, I^2=38%, N=473). There was no evidence for publication bias, but the number of studies was small. There were no differences in effect sizes between trials for borderline and Cluster C personality disorders (p=0.953).

Differences to other active therapies with regard to core personality disorder symptoms were insignificant. The sample size exceeded the OIS (N=473 >136). Thus, precision was adequate. The corrected effect size is small and its CI does not exceed ±0.43, implying no clinically significant difference in efficacy compared to other active therapies.

There were no significant differences in follow-up studies comparing PDT with active therapies with regard to core personality disorder symptoms (g=0.00, 95% CI: –0.48 to 0.49, I^2=64%, N=370). Removing one outlier[70] reduced heterogeneity (g=–0.18, 95% CI: –0.38 to 0.03, n=4, I^2=5%). The corrected effect size was neither statistically nor clinically significant, and the sample size exceeds the OIS (N=370 > 136), indicating no serious imprecision.

Quality measures

The quality of the eligible meta-analysis was found to be very good, with 11/11 relevant items fulfilled. The quality of primary studies, as assessed by the RCT-PQRS, was sufficient (total score ≥ 24) for most studies (81%).

As to treatment fidelity, all the 16 included studies used a treatment manual (100%); most studies described adequate qualification of therapists (87.5%) and monitored the treatment during the trial by supervision (94.5%). A smaller percentage empirically assessed treatment integrity (50%).

Adequate random sequence generation, allocation concealment, blinding of assessors (or use of only self-report measures) and intention-to-treat analysis were reported in 50%, 44%, 69% and 50% of the studies, respectively (see also supplementary information). The corresponding values for the comparison with all controls were 60%, 40%, 80% and 40%; those for the comparison with active controls only were 75%, 50%, 75% and 50%; those for the comparison with active therapies were 43%, 43%, 71% and 47%, respectively.

Secondary outcomes

For improvement of functioning, PDT yielded a medium effect size compared to all controls (g=–0.66, 95% CI: –1.01 to –0.32, n=7, I^2=57%). When a potential outlier was removed[73], heterogeneity was reduced (g=–0.72, 95% CI: –1.04 to –0.41, n=6, I^2=42%, N=431). We used an SMD of ±0.45 as a minimal clinically meaningful effect size[72], resulting in an OIS of 124[53]. The sample size exceeds the OIS (N=431 > 124). The effect size and the lower boundary of the CI can be regarded as clinically meaningful, but the upper boundary falls below the margin.

For interpersonal problems (g=–0.05, 95% CI: –0.20 to 0.12; n=4, I^2=2%, N=394) and functioning (g=0.12, 95% CI: –0.12 to 0.36, n=4, I^2=2%, N=394), there were no significant differences between

Table 3 Synthesis of evidence profiles for psychodynamic therapy (PDT) in personality disorders

Outcomes	Quality assessment								Summary of findings		Quality of evidence
	Quality of systematic review (QSR)	Quality of studies (QS) (RCT-PQRS ≥ 24)	Treatment fidelity (FI)	GRADE				Publication bias	N patients (both arms)	Effect size (95% CI)	GRADE QSR, QS, FI Total rating
				Risk of bias	Inconsistency	Indirectness	Imprecision				
PDT vs. all controls											
Personality disorder symptoms (critical) 5 RCTs	11/11	81%	+	Yes	No	No	No	Undetected	239	–0.63 (–0.87, –0.41) I^2=11%	–+++ +++ Moderate
Functioning (important) 6 RCTs (1 outlier excluded)	11/11	67%	+	Yes	No	No	No	Not assessed	431	–0.72 (–1.04, –0.41) I^2=42%	–+++ +++ Moderate
Suicidality (important) 4 RCTs (1 outlier excluded)	11/11	75%	+	Yes	No	No	Yes	Not assessed	239	–0.67 (–1.13, –0.20) I^2=40%	–++– +++ Low
PDT vs. active controls											
Personality disorder symptoms (critical) 4 RCTs	11/11	75%	+	Yes	No	No	No	Not assessed	200	–0.65 (–0.99, –0.32) I^2=15%	–+++ +++ Moderate
PDT vs. active therapies											
Personality disorder symptoms (critical) 6 RCTs (1 outlier excluded)	11/11	83%	+	Yes	No	No	No	Not detected	473	–0.04 (–0.31, 0.22) I^2=38%	–+++ +++ Moderate
Personality disorder symptoms follow-up (important) 4 RCTs (1 outlier excluded)	11/11	100%	+	Yes	No	No	No	Not assessed	370	–0.18 (–0.38, 0.03) I^2=5%	–+++ +++ Moderate
Functioning (important) 4 RCTs	11/11	100%	+	No	No	No	No	Not assessed	394	0.12 (–0.12, 0.36) I^2=2%	++++ +++ High
Interpersonal problems (important) 4 RCTs	11/11	100%	+	No	No	No	No	Not assessed	394	–0.05 (–0.20, 0.12) I^2=2%	++++ +++ High
Follow-up interpersonal problems (important) 4 RCTs (1 outlier excluded)	11/11	100%	+	Yes	No	No	No	Not assessed	370	–0.23 (–0.28, 0.17) I^2=0%	–+++ +++ Moderate

RCTs – randomized controlled trials; RCT-PQRS – Randomized Controlled Trial Psychotherapy Quality Rating Scale; GRADE – Grading of Recommendations Assessment, Development, and Evaluation

PDT and other active therapies. The sample sizes exceeded the OIS (N=62) determined for functioning.

An RCT found PDT to be superior to dialectical behavior therapy and supportive therapy in improving reflective functioning and attachment in borderline personality disorder, thus showing an additional gain[74]. For improving reflective functioning, the effect size in favour of PDT was large (d=–0.84) compared to supportive therapy and medium (d=–0.55) compared to dialectical behavior therapy[74].

Two RCTs suggest that PDT is a cost-effective treatment in personality disorders and high utilizers of psychiatric services[75,76]. The efficacy of PDT for personality disorders has not been specifically tested in minorities. In one RCT of PDT, non-occurrence of any adverse events was explicitly reported[75].

GRADE

According to the results presented above, there is a relatively wide range of studies of PDT in personality disorders (n=16). PDT achieved a clinically meaningful medium effect size compared to all controls (g=–0.63) and active controls (g=–0.65) in the reduction of core personality disorder symptoms. No differences in efficacy compared to other active therapies were detected (g=–0.04). We did not find serious indications of inconsistency, indirectness, imprecision or publication bias (see Table 3). The CIs were relatively wide, but comparable to those of other active therapies[72]. Most studies showed a sufficient quality (81%) as assessed by the RCT-PQRS, and sufficient treatment fidelity. The quality of the meta-analysis was rated as very good.

For comparisons with all controls and active controls in personality disorders, most studies showed a low risk of bias for random sequence generation and blinding, and an unclear or high risk for allocation concealment and completeness (see also supplementary information). For comparisons with active therapies, most studies showed a low risk of bias for blinding, but an unclear or high risk for all other dimensions. As noted above, the GRADE guidelines recommend to be conservative in rating down the quality of evidence. Thus, taking all results into account, the review panel decided to downgrade the evidence for PDT in personality disorders by one level due to risk of bias, rating the evidence for critical outcomes as moderate (see Table 3).

Somatic symptom disorders

The eligible recent meta-analysis assessing the efficacy of PDT for somatic symptom disorders, in comparison with control conditions or active therapies, included 17 RCTs (N=2,106)[40]. There was some heterogeneity between the PDT methods, but they were all based on psychodynamic theory and technique[17].

Efficacy of PDT vs. control conditions and active therapies

PDT was significantly superior to control conditions in improving somatic symptoms, with a large effect size (SMD=–0.84, 95% CI: –1.35 to –0.33, n=11, N=895). There was evidence for possible publication bias (Egger's regression asymmetry test = –3.49, 95% CI: –5.65 to –1.33, p=0.047). Excluding one outlier[77] reduced heterogeneity to a moderate level, resulting in a medium effect size (SMD=–0.47, 95% CI: –0.70 to –0.23, n=10, I^2=55%, N=776).

Compared to active controls, PDT achieved a moderate effect size (SMD=–0.41; 95% CI: –0.74 to –0.09, n=7, N=644, I^2 = 70%).

PDT was significantly superior to control conditions in 3-6 month follow-ups (SMD=–0.45, 95% CI: –0.69 to –0.20, n=4, I^2=30%, N=479). At >6 month follow-up, the effect size was large (SMD= –1.17, 95% CI: –2.07 to 0.27, n=6, N=801), but I^2 was also large at 97%. When removing one outlier[77], the effect size was significant but small (SMD=–0.17, 95% CI: –0.32 to –0.02, n=5, I^2=26%, N=702). Compared to active controls 3-6 months after end of therapy, PDT achieved a medium effect size (SMD=–0.45, 95% CI: –0.69 to –0.20, n=4, I^2=30%, N=479).

We used d=±0.25 as a conservative estimate for a minimum clinically meaningful effect size, resulting in an OIS of 398[53]. The effect size in the reduction of somatic symptoms was significant and the sample size exceeded the OIS (N=776 > 398) for PDT vs. control conditions. The lower boundary of the CI represents a medium to large effect size; the upper boundary is slightly below –0.25 and may still represent a clinically meaningful effect size. The width of the CI is comparable to psychotherapy in somatic symptom disorders in general (see supplementary information). Five RCTs suggest that PDT is at least as efficacious as other therapies, including CBT[40].

There were no indications of serious indirectness with regard to patients, treatments outcomes, or comparisons in any of the analyses.

Quality measures

The quality of the eligible meta-analysis was found to be very good, with 11/11 relevant items fulfilled. The quality of primary studies was assessed according to the criteria defined by Guidi et al[78]: of the 17 studies, 94% described the longitudinal development of the somatic condition, 100% described treatment components, 76.4% reported past/current medication use, 64.7% described weakness of controls, 41.1% used observer and self-rated instruments, while only 17.6% described adverse effects beyond dropout rates, and 24% reported rates of deterioration after treatment beyond dropout rates.

As to treatment fidelity, all but one study used a treatment manual or manual-like guideline (94%), 53% of studies monitored treatments by video or audio recordings, and 53% checked treatment integrity by adherence ratings.

Adequate random sequence generation, allocation concealment, blinding of assessors (or use of only self-report measures) and report of complete outcome data were found in 59%, 53%, 59% and 76% of all studies; in 70%, 70%, 80% and 80% of the studies including all controls, and in 71%, 71%, 86% and 86% of studies including active controls only, respectively.

Secondary outcomes

PDT achieved a medium effect size compared to control conditions in improving functioning at short-term (SMD=–0.58, 95% CI: –1.16 to –0.01, n=5, I^2=88%, N=641). In the follow-up >6 months after end of therapy, a non-significant effect size was achieved compared to all controls (SMD=–0.05, 95% CI: –0.63 to 0.73, n=3, I^2=89%, N=641) (see Table 4).

One RCT suggests that PDT is a cost-effective treatment in somatic symptom disorders[79]. No studies have addressed the efficacy of PDT in minorities. No or only a few adverse events were reported in studies of PDT in somatic symptom disorders.

GRADE

There is a relatively wide range of RCTs of PDT in somatic symptom disorders (n=17). PDT was significantly superior to all controls with a medium effect size (SMD=–0.47). In addition, there is preliminary evidence from individual RCTs that PDT is at least as efficacious as other empirically-supported therapies. Treatment effects were found to be stable at follow-ups. There is evidence to suggest that the benefits outweigh the costs and harms, as required by GRADE[6,60]. We did not find serious inconsistency, indirectness or imprecision. There seems to be some publication bias.

Most studies showed a sufficient quality and treatment fidelity, and the quality of the meta-analysis was rated as good. For comparisons of PDT with all controls and active controls, most studies showed a low risk of bias in critical outcomes for random sequence generation, allocation concealment, blinding and completeness. Taking these results into account, the review panel decided to rate the evidence for PDT in somatic symptom disorders as high for critical outcomes (see Table 4).

Mechanisms of change in PDT

Our systematic search yielded one recent meta-analysis reporting a significant moderate correlation (r=0.31) between insight and outcome across a variety of psychotherapeutic approaches, including PDT[46]. Studies in depressive disorders, anxiety disorders and Cluster C personality disorders found that gains in insight preceded improvements in outcome of PDT[80-82]. These effects were found to be specific to PDT[80-82]. In personality disorders, the effect of transference work in patients with more severe interpersonal difficulties was found to be mediated by both improvements in insight and affect awareness[83].

A recent meta-analysis found a significant moderate correlation of 0.28 between alliance and outcome across different psychotherapies, with no significant differences between approaches[84]. For PDT, the correlation was 0.24[84]. With regard to diagnoses, associations were similar in anxiety, depressive and personality disorders[84]. There is preliminary evidence from studies examining within-patient effects that the alliance may have a causal role in improving outcomes[32], including studies of PDT (in depressive disorders)[85]. Specifically for PDT, it has been documented that the temporal precedence of alliance predicting symptom change becomes stronger with time over the course of long-term therapy[86].

Change in defense mechanisms was found to be related to outcome in studies of PDT in patients with depressive, anxiety and personality disorders, with correlations between 0.28 and 0.64[87,88]. The largest correlations were found for improvements in depression and functioning (0.64, 0.60)[87,88]. However, only a few studies of defense mechanisms examined temporal precedence[32].

A recent study found that PDT outcome in patients with borderline personality disorder was strongly related to improvements in reflective functioning (r=0.89)[89]. However, this study did not examine whether change in reflective functioning preceded change in outcome.

There is some evidence that emotion processing plays a role in PDT of somatic symptom disorders[90]. Furthermore, recent studies highlight the importance of both insight and emotional experiencing as mechanisms of change in PDT for depressive, anxiety and personality disorders[83,91].

Summary and recommendations

A synthesis of the most recent evidence for PDT in depressive, anxiety, personality and somatic symptom disorders as reviewed above is given in Tables 1-4, while a summary for quality of evidence and recommendations is provided in Table 5.

According to the revised EST criteria[3], there is evidence for the efficacy of PDT in these disorders based on recent systematic quantitative reviews, covering a relatively wide range of studies, showing a sufficient conceptual homogeneity between treatments, with sufficient quality of most individual studies (except for anxiety disorders comparing PDT with active controls), sufficient quality of meta-analyses, and sufficient treatment fidelity. No serious indirectness, imprecision, inconsistency or publication bias concerning critical outcomes was found, with the possible exception of publication bias in somatic symptom disorders. Clinically meaningful effects in target symptom improvement compared to (active) controls were found, with moderate heterogeneity after removing outliers, as well as stable effects in longer-term follow-ups, and low risk of adverse events. Clinically meaningful effect sizes in functioning were found in all disorders with the exception of anxiety disorders. Differences in comparison to other active therapies were small and not clinically significant, suggesting equivalence in efficacy. Furthermore, for PDT in the aforementioned disorders, there is some evidence for presumed mechanisms of change. There is also some preliminary evidence that PDT is cost-effective, effective under conditions of routine clinical practice, and efficacious in some sub-populations of the above disorders, which represent contextual factors as listed in the new EST model[3]. A positive balance between benefits, costs and harms exists.

In sum, the results for PDT in the examined disorders fulfill several criteria for high quality evidence according to the new EST model[3]. Some limitations exist as well. There is room for further

Table 4 Synthesis of evidence profiles for psychodynamic therapy (PDT) in somatic symptom disorders

Outcomes	Quality assessment								Summary of findings		Quality of evidence
	Quality of systematic review (QSR)	Quality of studies (QS)	Treatment fidelity (FI)	GRADE				Publication bias	N patients (both arms)	Effect size (95% CI)	GRADE QSR, QS, FI Total rating
				Risk of bias	Inconsistency	Indirectness	Imprecision				
PDT vs. all controls											
Somatic symptoms (critical) 10 RCTs (1 outlier excluded)	11/11	+	+	No	No	No	No	Possible	776	–0.47 (–0.70, –0.23) I^2=55%	++++ +++ High
3-6 month follow-up somatic symptoms (important) 4 RCTs	11/11	+	+	Yes	No	No	No	Not assessed	479	–0.45 (–0.69, –0.20) I^2=30%	–+++ +++ Moderate
>6 month follow-up somatic symptoms (important) 6 RCTs (1 outlier excluded)	11/11	+	+	Yes	No	No	No	Not assessed	702	–0.17 (–0.32, –0.02) I^2=26%	–+++ +++ Moderate
Functioning (important) 5 RCTs	11/11	+	+	No	Yes	No	Yes	Not assessed	641	–0.58 (–1.16, –0.01) I^2=88%	+–+– +++ Low
>6 month follow-up functioning (important) 3 RCTs	11/11	+	+	No	Yes	No	Yes	Not assessed	377	–0.05 (–0.63, 0.73) I^2=89%	+–+– +++ Low
PDT vs. active controls											
0-3 month follow-up somatic symptoms (critical) 7 RCTs (1 outlier excluded)	11/11	+	+	No	No	No	No	Not assessed	644	–0.41 (–0.74, –0.09) I^2=70%	++++ +++ High
3-6 month follow-up somatic symptoms (important) 4 RCTs	11/11	+	+	Yes	No	No	No	Not assessed	479	–0.45 (–0.69, –0.20) I^2=30%	–+++ +++ Moderate

Table 4 Synthesis of evidence profiles for psychodynamic therapy (PDT) in somatic symptom disorders *(continued)*

	Quality assessment								Summary of findings		Quality of evidence
				GRADE							
Outcomes	Quality of systematic review (QSR)	Quality of studies (QS)	Treatment fidelity (FI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	N patients (both arms)	Effect size (95% CI)	GRADE QSR, QS, FI Total rating
>6 month follow-up somatic symptoms (important) 6 RCTs (1 outlier excluded)	11/11	+	+	Yes	No	No	No	Not assessed	702	–0.17 (–0.32, –0.02) I^2=26%	–+++ +++ Moderate
0-3 month follow-up functioning (important) 4 RCTs	11/11	+	+	No	Yes	No	Yes	Not assessed	504	–0.57 (–1.22, 0.09) I^2=91%	+–+– +++ Low
>6 month follow-up functioning (important) 3 RCTs	11/11	+	+	No	Yes	No	Yes	Not assessed	378	–0.16 (–0.70, 0.38) I^2=82%	+–+– +++ Low

RCTs – randomized controlled trials; GRADE – Grading of Recommendations Assessment, Development, and Evaluation

Table 5 Summary of the status of psychodynamic therapy (PTD) as an empirically supported treatment for common mental disorders

	Comparison (critical outcome)	Effect size (95% CI)	GRADE	Quality of evidence	Efficacy demonstrated across several patient sub-populations	Evidence for mechanisms of change	Recommendation
Depressive disorders	PDT vs. all controls	–0.58 (–0.33, –0.83)	++++	High		Yes	Strong
	PDT vs. active controls	–0.51 (–0.68, –0.35)	++++	High	Yes		
	PDT vs. active therapies	–0.03 (–0.23, 0.17)	–+++	Moderate			
Anxiety disorders	PDT vs. all controls	–0.72 (–1.06, –0.37)	–+++	Moderate	Yes	Yes	Strong
	PDT vs. active controls	–0.64 (–1.14, –0.14)	–++–	Low			
	PDT vs. active therapies	0.06 (–0.11, 0.23)	–+++	Moderate			
Personality disorders	PDT vs. all controls	–0.63 (–0.87, –0.41)	–+++	Moderate	Yes	Yes	Strong
	PDT vs. active controls	–0.65 (–0.99, –0.32)	–+++	Moderate			
	PDT vs. active therapies	–0.04 (–0.31, 0.22)	–+++	Moderate			
Somatic symptom disorders	PDT vs. all controls	–0.47 (–0.70, –0.23)	++++	High	Yes	Yes	Strong
	PDT vs. active controls	–0.41 (–0.74, –0.09)	++++	High			

GRADE – Grading of Recommendations Assessment, Development, and Evaluation

research on mechanisms of change in PDT, controlling for temporal precedence. In addition, further studies in minorities and on cost-effectiveness of PDT are needed. With regard to improvements in functioning, the quality of evidence was low in anxiety and somatic symptom disorders.

The new EST model provides three options of recommendation: "very strong", "strong" or "weak"[3]. According to the results presented above, there is high quality evidence for depressive disorders and somatic symptom disorders, and moderate quality evidence for anxiety and personality disorders, that PDT achieves clinically meaningful effects in target symptoms and functioning compared to controls and is associated with low risk of harms and reasonable costs[3]. In addition, there is moderate quality evidence that there are no meaningful differences in efficacy between PDT and other active therapies. Thus, the criteria of the new EST model suggest that a "strong" recommendation for PDT in depressive, anxiety, personality and somatic symptom disorders is most appropriate (see Table 5).

DISCUSSION

This umbrella review suggests that PDT represents an evidence-based psychotherapy for depressive, anxiety, personality and somatic symptom disorders. Limitations of research on PDT were identified as well. For some analyses, our review relied on a limited number of RCTs. Some of these RCTs were old and of poor quality. The included meta-analyses aggregated different categorical diagnoses – such as different forms of depressive, anxiety, personality or somatic symptom disorders – due to the limited number of RCTs per condition. However, this is consistent with the transdiagnostic approach of the new EST model, demonstrating efficacy of PDT across several patient populations.

On the other hand, our review has several strengths. We applied several criteria specified by the new EST model not used in some other recent reviews using that model[92-94], including all the assessments required by GRADE guidelines (risk of bias of individual studies; rating of inconsistency, indirectness and imprecision via the OIS) as well as the assessment of the quality and treatment fidelity of individual studies, and of clinical significance of effect sizes. Furthermore, we primarily included only recent meta-analyses published in the past two years, reviewed cost-effectiveness and mechanisms of change, and systematically reported effect sizes for the different comparison conditions.

Across all evidence-based treatments, the rates of response and remission are limited[30,95]. Thus, a focus of future research on PDT should be on helping the considerable proportion of patients not responding to the available treatments. As a related issue, it is important to find out which patients benefit from which therapy, taking possible moderators into account such as disorder severity, comorbid disorders, personality features, staging of disorder and previous treatment failures/resistance, and family history of mental illness, aiming at a personalized treatment approach[96-99].

Another focus should be on treatment dose, addressing for which patients which number of sessions, session frequency or treatment duration is required. Further individual RCTs of PDT are required in those areas where only a few or old RCTs are available, as well as for specific mental disorders such as bipolar or psychotic disorders. A focus on unified transdiagnostic protocols addressing syndromes rather than categorical diagnoses represents another promising approach which is in accordance with both the transdiagnostic nature of PDT and the new model for EST[21-25]. Focusing on transdiagnostic features such as work-related problems, including perfectionism and procrastination, is another understudied area[100]. More studies are required in which PDT is tailored to minorities and underserved groups. Finally, available treatments may be improved by process-outcome research identifying empirically-supported mechanisms of change[101].

Psychotherapy is a field of rivalry between different approaches. However, patients should be offered a variety of research-supported treatments. The limited rates of remission and response for evidence-based treatments show that no single approach fits all patients. Further studies are needed to identify treatment moderators showing who benefits from which treatment.

ACKNOWLEDGEMENTS

The authors dedicate this paper to their late friend and colleague S. Rabung, who substantially contributed to the paper but did not live to see it published. They are grateful to E. Driessen for her contribution. Supplementary information on this study is available at https://osf.io/txkw5/.

REFERENCES

1. Chambless DL, Hollon SD. Defining empirically supported therapies. J Consult Clin Psychol 1998;66:7-18.
2. Task Force on Promotion and Dissemination of Psychological Procedures. Training and dissemination of empirically-validated psychological treatments. Report and recommendations. Clin Psychol 1995;48:3-23.
3. Tolin D, McKay D, Forman EM et al. Empirically supported treatments: recommendations of a new model. Clin Psychol Sci Pract 2015;22;317-38.
4. Sakaluk JK, Kilshaw RE, Williams AJ et al. Evaluating the evidential value of empirically supported psychological treatments (ESTs): a meta-scientific review. J Abnorm Psychol 2019;128:500-9.
5. Atkins D, Eccles M, Flottorp S et al. Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches. The GRADE Working Group. BMC Health Serv Res 2004;4:38.
6. Guyatt GH, Oxman AD, Vist GE et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336: 924-6.
7. Guyatt G, Gutterman D, Baumann MH et al. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians task force. Chest 2006;129:174-81.
8. Guyatt GH, Oxman AD, Schunemann HJ et al. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol 2011;64:380-2.
9. Guyatt G, Oxman AD, Akl EA et al. GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383-94.
10. Leichsenring F, Leweke F, Klein S et al. The empirical status of psychodynamic psychotherapy – an update: Bambi's alive and kicking. Psychother Psychosom 2015;84:129-48.
11. Leichsenring F, Luyten P, Hilsenroth MJ et al. Psychodynamic therapy meets evidence-based medicine: a systematic review using updated criteria. Lancet Psychiatry 2015;2:648-60.
12. Leichsenring F, Steinert C. The efficacy of psychodynamic psychotherapy: an up-to-date review. In: Kealy D, Ogrodniczuk JS (eds). Contemporary psychodynamic psychotherapy. London: Elsevier, 2019:49-74.
13. Leichsenring F, Steinert C. Applications of psychodynamic psychotherapy to specific disorders. In: Crisp H, Gabbard G (eds). Gabbard's textbook of psychotherapeutic treatments, 2nd ed. Washington: American Psychiatric Association Publishing, 2023:123-48.
14. Connolly Gibbons MB, Crits-Christoph P, Hearon B. The empirical status of psychodynamic therapies. Annu Rev Clin Psychol 2008;4:93-108.
15. Chambless DL. Task Force on Promotion and Dissemination of Psychological Procedures. A report adopted by the Division 12 Board of the American Psychological Association. Oklahoma: American Psychological Association, 1993.
16. Leichsenring F, Abbass A, Heim N et al. Empirically supported psychodynamic psychotherapy for common mental disorders – An update applying revised criteria: systematic review protocol. Front Psychiatry 2022;13:976885.
17. Gabbard GO. Long-term psychodynamic psychotherapy. Washington: American Psychiatric Publishing, 2004.
18. Gunderson JG, Gabbard G. Making the case for psychoanalytic therapies in the current psychiatric environment. J Am Psychoanal Assoc 1999;47:679-704.
19. Wallerstein RS. The psychotherapy research project of the Menninger Foundation: an overview. J Consult Clin Psychol 1989;57:195-205.
20. Ioannidis JP. Interpretation of tests of heterogeneity and bias in meta-analysis. J Eval Clin Pract 2008;14:951-7.
21. Leichsenring F, Steinert C. Towards an evidence-based unified psychodynamic protocol for emotional disorders. J Affect Disord 2018;232:400-16.
22. Leichsenring F, Schauenburg H. Empirically supported methods of short-term psychodynamic therapy in depression – Towards an evidence-based unified protocol. J Affect Disord 2014;169C:128-43.
23. Leichsenring F, Salzer S. A unified protocol for the transdiagnostic psychodynamic treatment of anxiety disorders: an evidence-based approach. Psychotherapy 2014;51:224-45.
24. Busch FN, Singer MB, Milrod BL et al. Manual of panic focused psychodynamic psychotherapy – extended range. New York: Taylor & Francis, 2012.
25. Caligor E, Kernberg O, Clarkin JF et al. Psychodynamic therapy for personality disorder pathology: treating self and interpersonal functioning. Washington: American Psychiatric Publishing, 2018.
26. Barlow DH, Farchione TJ, Bullis JR et al. The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders compared with diagnosis-specific protocols for anxiety disorders: a randomized clinical trial. JAMA Psychiatry 2017;74:875-84.
27. Fairburn CG, Cooper Z, Doll HA et al. Transdiagnostic cognitive-behavioral therapy for patients with eating disorders: a two-site trial with 60-week follow-up. Am J Psychiatry 2009;166:311-9.
28. Aromataris E, Fernandez R, Godfrey C et al. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015;13:132-40.
29. Shea BJ, Reeves BC, Wells G et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008.
30. Cuijpers P, Karyotaki E, Ciharova M et al. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: a meta-analysis. Acta Psychiatr Scand 2021;144:288-99.
31. Kocsis JH, Gerber AJ, Milrod B et al. A new scale for assessing the quality of randomized clinical trials of psychotherapy. Compr Psychiatry 2010;51:319-24.
32. Crits-Christoph P, Connolly Gibbons MB. Psychotherapy process outcome research – advances in understanding causal connections. In: Barkham M, Castonguay LG, Lutz W (eds). Bergin and Garfield's handbook of psychotherapy and behavior change, 7th ed. New York: Wiley, 2021:263-95.
33. Guyatt GH, Oxman AD, Kunz R et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395-400.
34. Higgins JP, Thompson SG, Deeks JJ et al. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
35. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. New York: Wiley, 2009.
36. Guyatt GH, Oxman AD, Kunz R et al. GRADE guidelines 6. Rating the quality of evidence – imprecision. J Clin Epidemiol 2011;64:1283-93.
37. Schünemann HJ, Higgins JPT, Vist GE et al. Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JPT, Chandler J, Cumpston M et al (eds). Cochrane handbook for systematic reviews of interventions, Version 6.3. London: Cochrane, 2022.
38. Barber JP, Muran JC, McCarthy KS et al. Research on psychodynamic therapies. In: Barkham M, Castonguay LG, Lutz W (eds). Bergin and Garfield's handbook of psychotherapy and behavior change, 7th ed. New York: Wiley, 2021:387-419.
39. Keefe JR, McMain SF, McCarthy KS et al. A meta-analysis of psychodynamic treatments for borderline and Cluster C personality disorders. Personal Disord 2020;11:157-69.
40. Abbass A, Town J, Holmes H et al. Short-term psychodynamic psychotherapy for functional somatic disorders: a meta-analysis of randomized controlled trials. Psychother Psychosom 2020;89:363-70.
41. Keefe JR, McCarthy KS, Dinger U et al. A meta-analytic review of psychodynamic therapies for anxiety disorders. Clin Psychol Rev 2014;34:309-23.
42. Driessen E, Hegelmaier LM, Abbass AA et al. The efficacy of short-term psychodynamic psychotherapy for depression: a meta-analysis update. Clin Psychol Rev 2015;42:1-15.
43. Lindegaard T, Berg M, Andersson G. Efficacy of Internet-delivered psychodynamic therapy: systematic review and meta-analysis. Psychodyn Psychiatry 2020;48:437-54.
44. Driessen E, Dekker JJM, Peen J et al. The efficacy of adding short-term psychodynamic psychotherapy to antidepressants in the treatment of depression: a systematic review and meta-analysis of individual participant data. Clin Psychol Rev 2020;80:101886.
45. Gaskell C, Simmonds-Buckley M, Kellett S et al. The effectiveness of psychological interventions delivered in routine practice: systematic review and

meta-analysis. Adm Policy Ment Health 2023;50:43-57.
46. Jennissen S, Huber J, Ehrenthal JC et al. Association between insight and outcome of psychotherapy: systematic review and meta-analysis. Am J Psychiatry 2018;175:961-9.
47. Thoma NC, McKay D, Gerber AJ et al. A quality-based review of randomized controlled trials of cognitive-behavioral therapy for depression: an assessment and metaregression. Am J Psychiatry 2012;169:22-30.
48. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. Biol Psychiatry 2006;59:990-6.
49. Cuijpers P, Turner EH, Koole SL et al. What is the threshold for a clinically relevant effect? The case of major depressive disorders. Depress Anxiety 2014;31:374-8.
50. Cuijpers P, Karyotaki E, Reijnders M et al. Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression. Epidemiol Psychiatr Sci 2019;28:21-30.
51. Cipriani A, Furukawa TA, Salanti G et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 2018;391:1357-66.
52. Leichsenring F, Steinert C, Rabung S et al. The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: an umbrella review and meta-analytic evaluation of recent meta-analyses. World Psychiatry 2022; 21:133-45.
53. Faul F, Erdfelder E, Lang AG et al. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39:175-91.
54. Bastos AG, Guimaraes LS, Trentini CM. The efficacy of long-term psychodynamic psychotherapy, fluoxetine and their combination in the outpatient treatment of depression. Psychother Res 2015;25:612-24.
55. Rosso G, Martini B, Maina G. Brief dynamic therapy and depression severity: a single-blind, randomized study. J Affect Disord 2013;147:101-6.
56. Eielsen M, Ulvenes PG, Rossberg JI et al. The effectiveness of an intensive inpatient psychotherapy program for chronic depression: a naturalistic comparison with wait list. BMC Psychiatry 2022;22:745.
57. Town JM, Abbass A, Stride C et al. Efficacy and cost-effectiveness of intensive short-term dynamic psychotherapy for treatment resistant depression: 18-month follow-up of the Halifax depression trial. J Affect Disord 2020;273:194-202.
58. Barber JP, Barrett MS, Gallop R et al. Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder: a randomized, placebo-controlled trial. J Clin Psychiatry 2012;73:66-73.
59. Connolly Gibbons MB, Gallop R, Thompson D et al. Comparative effectiveness of cognitive therapy and dynamic psychotherapy for major depressive disorder in a community mental health setting: a randomized clinical noninferiority trial. JAMA Psychiatry 2016;73:904-11.
60. Andrews JC, Schunemann HJ, Oxman AD et al. GRADE guidelines: 15. Going from evidence to recommendation – determinants of a recommendation's direction and strength. J Clin Epidemiol 2013;66:726-35.
61. Balshem H, Helfand M, Schunemann HJ et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401-6.
62. Steinert C, Bumke PJ, Hollekamp RL et al. Resource activation for treating post-traumatic stress disorder, co-morbid symptoms and impaired functioning: a randomized controlled trial in Cambodia. Psychol Med 2017;47:553-64.
63. Milrod B, Leon AC, Busch F et al. A randomized controlled clinical trial of psychoanalytic psychotherapy for panic disorder. Am J Psychiatry 2007;164: 265-72.
64. Durham RC, Murphy T, Allan T et al. Cognitive therapy, analytic psychotherapy and anxiety management training for generalised anxiety disorder. Br J Psychiatry 1994;165:315-23.
65. Maljanen T, Knekt P, Lindfors O et al. The cost-effectiveness of short-term and long-term psychotherapy in the treatment of depressive and anxiety disorders during a 5-year follow-up. J Affect Disord 2016;190:254-63.
66. Andersson G, Paxling B, Roch-Norlund P et al. Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: a randomized controlled trial. Psychother Psychosom 2012;81:344-55.
67. Fonagy P, Bateman A, Luyten P et al. Psychoanalytic/psychodynamic approaches to personality disorders. In: Lejuez CW, Gratz KL (eds). The Cambridge handbook of personality disorders. Cambridge: Cambridge University Press, 2020:427-39.
68. Zanarini MC, Vujanovic AA, Parachini EA et al. Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD): a continuous measure of DSM-IV borderline psychopathology. J Pers Disord 2003;17:233-42.
69. Giesen-Bloo J, van Dyck R, Spinhoven P et al. Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. Arch Gen Psychiatry 2006;63: 649-58.
70. Emmelkamp PMG, Benner A, Kuipers A et al. Comparison of brief dynamic and cognitive-behavioral therapies in avoidant personality disorder. Br J Psychiatry 2006;189:60-4.
71. Bateman A, Fonagy P. Effectiveness of partial hospitalization in the treatment of borderline personality disorder: a randomized controlled trial. Am J Psychiatry 1999;156:1563-9.
72. Storebo OJ, Stoffers-Winterling JM, Vollm BA et al. Psychological therapies for people with borderline personality disorder. Cochrane Database Syst Rev 2020;5:CD012955.
73. Amianto F, Ferrero A, Piero A et al. Supervised team management, with or without structured psychotherapy, in heavy users of a mental health service with borderline personality disorder: a two-year follow-up preliminary randomized study. BMC Psychiatry 2011;11:181.
74. Levy KN, Meehan KB, Kelly KM et al. Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. J Consult Clin Psychol 2006;74:1027-40.
75. Abbass AA, Sheldon A, Gyra J et al. Intensive short-term psychotherapy for DSM-IV personality disorders: a randomized controlled trial. J Nerv Ment Dis 2008;196:211-6.
76. Guthrie E, Moorey J, Margison F et al. Cost-effectiveness of brief psychodynamic-interpersonal therapy in high utilizers of psychiatric services. Arch Gen Psychiatry 1999;56:519-26.
77. Svedlund J, Sjodin I, Ottosson JO et al. Controlled study of psychotherapy in irritable bowel syndrome. Lancet 1983;2:589-92.
78. Guidi J, Brakemeier EL, Bockting CLH et al. Methodological recommendations for trials of psychological interventions. Psychother Psychosom 2018; 87:276-84.
79. Creed F, Fernandes L, Guthrie E et al. The cost-effectiveness of psychotherapy and paroxetine for severe irritable bowel syndrome. Gastroenterology 2003;124:303-17.
80. Gibbons MB, Crits-Christoph P, Barber JP et al. Unique and common mechanisms of change across cognitive and dynamic psychotherapies. J Consult Clin Psychol 2009;77:801-13.
81. Jennissen S, Connolly Gibbons MB, Crits-Christoph P et al. Insight as a mechanism of change in dynamic therapy for major depressive disorder. J Couns Psychol 2021;68:435-45.
82. Kallestad H, Valen J, McCullough L et al. The relationship between insight gained during therapy and long-term outcome in short-term dynamic psychotherapy and cognitive therapy for cluster C personality disorders. Psychother Res 2010;20:526-34.
83. Hoglend P, Hagtvet K. Change mechanisms in psychotherapy: both improved insight and improved affective awareness are necessary. J Consult Clin Psychol 2019;87:332-44.
84. Fluckiger C, Del Re AC, Wampold BE et al. The alliance in adult psychotherapy: a meta-analytic synthesis. Psychotherapy 2018;55:316-40.
85. Zilcha-Mano S, Solomonov N, Chui H et al. Therapist-reported alliance: is it really a predictor of outcome? J Couns Psychol 2015;62:568-78.
86. Volz M, Jennissen S, Schauenburg H et al. Intraindividual dynamics between alliance and symptom severity in long-term psychotherapy: why time matters. J Couns Psychol 2021;68:446-56.
87. Bond M, Perry JC. Long-term changes in defense styles with psychodynamic psychotherapy for depressive, anxiety, and personality disorders. Am J Psychiatry 2004;161:1665-71.
88. Perry JC, Bond M. Change in defense mechanisms during long-term dynamic psychotherapy and five-year outcome. Am J Psychiatry 2012;169:916-25.
89. De Meulemeester C, Vansteelandt K, Luyten P et al. Mentalizing as a mechanism of change in the treatment of patients with borderline personality disorder: a parallel process growth modeling approach. Personal Disord 2018;9:22-9.
90. Town JM, Lomax V, Abbass AA et al. The role of emotion in psychotherapeutic change for medically unexplained symptoms. Psychother Res 2019;29:86-98.
91. Fisher H, Rafaeli E, Bar-Kalifa E et al. Therapists' interventions as a predictor of clients' emotional experience, self-understanding, and treatment outcomes. J Couns Psychol 2020;67:66-78.
92. Boness CL, Hershenberg R, Kaye J et al. An evaluation of cognitive behavioral therapy for insomnia: a systematic review and application of Tolin's criteria for empirically supported treatments. Clin Psychol 2020;27:e12348.

93. Boness CL, Votaw V, Schwebel FJ et al. An evaluation of cognitive behavioral therapy for substance use: an application of Tolin's criteria for empirically supported treatments. PsyArXiv 2022; doi: 10.31234/osf.io/rpwh9.
94. Tolin D, Melnyk T, Marx B. Exposure and response prevention for obsessive-compulsive disorder. https://www.div12.org/wp-content/uploads/2019/10/Treatment-Review-ERP-for-OCD.pdf.
95. Springer KS, Levy HC, Tolin DF. Remission in CBT for adult anxiety disorders: a meta-analysis. Clin Psychol Rev 2018;61:1-8.
96. Maj M, Stein DJ, Parker G et al. The clinical characterization of the adult patient with depression aimed at personalization of management. World Psychiatry 2020;19:269-93.
97. Maj M, van Os J, De Hert M et al. The clinical characterization of the patient with primary psychosis aimed at personalization of management. World Psychiatry 2021;20:4-33.
98. Stein DJ, Craske MG, Rothbaum BO et al. The clinical characterization of the adult patient with an anxiety or related disorder aimed at personalization of management. World Psychiatry 2021;20:336-56.
99. McIntyre RS, Alda M, Baldessarini RJ et al. The clinical characterization of the adult patient with bipolar disorder aimed at personalization of management. World Psychiatry 2022;21:364-87.
100. Steinert C, Heim N, Leichsenring F. Procrastination, perfectionism, and other work-related mental problems: prevalence, types, assessment, and treatment – a scoping review. Front Psychiatry 2021;12:736776.
101. Hayes SC, Hofmann SG. "Third-wave" cognitive and behavioral therapies and the emergence of a process-based approach to intervention in psychiatry. World Psychiatry 2021;20:363-75.

DOI:10.1002/wps.21104