Psychodynamic therapy meets evidence-based medicine: a systematic review using updated criteria

Falk Leichsenring, Patrick Luyten, Mark J Hilsenroth, Allan Abbass, Jacques P Barber, John R Keefe, Frank Leweke, Sven Rabung, Christiane Steinert

Psychodynamic therapy (PDT) is an umbrella concept for treatments that operate on an interpretive-supportive continuum and is frequently used in clinical practice. The use of any form of psychotherapy should be supported by sufficient evidence. Efficacy research has been neglected in PDT for a long time. In this review, we describe methodological requirements for proofs of efficacy and summarise the evidence for use of PDT to treat mental health disorders. After specifying the requirements for superiority, non-inferiority, and equivalence trials, we did a systematic search using the following criteria: randomised controlled trial of PDT; use of treatment manuals or manual-like guidelines; use of reliable and valid measures for diagnosis and outcome; adults treated for specific mental problems. We identified 64 randomised controlled trials that provide evidence for the efficacy of PDT in common mental health disorders. Studies sufficiently powered to test for equivalence to established treatments did not find substantial differences in efficacy. These results were corroborated by several meta-analyses that suggest PDT is as efficacious as treatments established in efficacy. More randomised controlled trials are needed for some mental health disorders such as obsessive-compulsive disorder and post-traumatic stress disorder. Furthermore, more adequately powered equivalence trials are needed.

Introduction

Psychotherapy is effective for the treatment of a broad range of mental disorders, symptoms, and problems. The use of any form of psychotherapy should be supported by sufficient evidence. Psychodynamic therapy is an umbrella concept for treatments that operate on an interpretive-supportive continuum. By interpretive interventions insight into wishes, affects, object relations or defence mechanisms is enhanced. Supportive interventions include fostering a therapeutic alliance, setting goals, or strengthening psychosocial capacities such as reality testing or impulse control. The use of more supportive or more interpretive (insight-enhancing) interventions is tailored to the patient’s needs. There is a range of manualised psychotherapeutic interventions that vary in the extent to which they focus on supportive or expressive elements.

In this review, we address the methodological and statistical requirements to determine efficacy in psychotherapy. We differentiate between testing superiority, non-inferiority, and equivalence. We focus specifically on the latter, because testing equivalence has not yet been widely implemented in psychotherapy research. Although studies often claim to have shown equivalence in outcome they often do not meet the requirements to do so. We apply these considerations specifically to PDT, which is frequently used in clinical practice, to update and expand the evidence for PDT in specific mental disorders (panel 1). However, these considerations apply to any psychotherapeutic or pharmacological treatment.

Methodology to determine efficacy: Grades of evidence

Randomised controlled trials (RCTs) are viewed by most as the gold standard, but RCT methodology has both strengths and weaknesses. For example, a randomised controlled efficacy study maximises the internal validity of a study, ie, the observed effects can be causally related to the applied treatments, at the possible expense of external validity, ie, generalisability to real-world conditions in clinical practice. In contrast, effectiveness studies investigate the effects of an intervention in routine clinical care and therefore have high external validity, but at the possible expense of internal validity. Thus, efficacy and effectiveness studies address different research questions. For treatments that have been evaluated in RCTs, studies are needed to investigate their effectiveness in real-life conditions.

In an RCT, a treatment might be compared with different control conditions, eg, no treatment, a placebo, a treatment as usual, an alternative treatment, or a treatment with known efficacy. The strictest test of efficacy is to compare the novel treatment with a treatment of proven efficacy, because this study design controls for both specific and non-specific (or common) factors.

A treatment comparison with a waiting list condition (no treatment) controls for the natural course of the disorder only, whereas comparison with another psychotherapy, a placebo, or a treatment as usual controls for factors common to all types of psychotherapy (eg, therapeutic alliance, expectations, motivation, general support and attention). This implies that the different study designs and comparison conditions are associated with different research questions (ie, is a treatment efficacious when controlling for the natural course, for common factors or for common and specific factors?). Furthermore, a treatment might be expected to be superior, non-inferior, or equivalent to another treatment or condition. Thus, to distinguish between testing for superiority, equivalence, and non-inferiority is important.

Superiority

For a treatment to be considered superior to another treatment, the treatment group must show a statistically
significant better outcome than that of the comparison condition. However, statistical significance is only a necessary but not sufficient condition for showing superiority. In addition, the magnitude of difference must also be taken into account in the form of between-group effect sizes (eg, the different measures proposed by Cohen,9 or the number needed to treat [NNT], the area under the receiver operating characteristic [ROC] curve [AUC], or success rate difference [SRD]).11,12 For several reasons, odds ratios are not recommended as a measure of effect size.12 The magnitude of difference is important because a small difference in outcome might be statistically, but not clinically, significant.

To demonstrate superiority, the two-sided test for differences is typically used. To detect, for example, a medium difference in means of d=0.5 between treatments with a power of 0.80 by a two-sided test at α=0.05, 64 patients per group are required (table 1).9

**Equivalence**

Equivalence trials are used to show that a novel treatment is no better and no worse in outcome than an established treatment. If the traditional two-sided test for differences is used to test equivalence or non-inferiority in outcome, the conclusions are often incorrect because a two-sided test does not take into account a margin of equivalence. The margin of equivalence (−ΔE, ΔE) defines a range of values for which the efficacies are close enough to be considered equivalent. In practical terms, the margin is the maximum clinically acceptable difference that one is willing to accept.11,12 Furthermore, a non-significant result implies only that equality cannot be ruled out, which is not the same as proving equality. A more appropriate test for equivalence is the two one-sided test (TOST) procedure (panel 2).11,12 Outcomes are equivalent if the CI of the empirically found difference is within the equivalence margin (panel 1).11,12 The required sample size and the statistical power directly depend on the size of the equivalence margin.13 The traditional two-sided test and the equivalence test (TOST) often yield inconsistent results.12

**Non-inferiority**

A non-inferiority (NI) trial tests the hypothesis that the efficacy of a test treatment is no more than ΔNI lower than that of a treatment whose efficacy is established. The non-inferiority hypothesis will be accepted at a significance level of α if the lower limit of (1−2α)×100% CI for the difference is above −ΔNI.11 To test for non-inferiority, a one-sided test at α=0.025 is recommended.19 Non-inferiority trials are based on the assumption that the test treatment is superior to the standard treatment on an outcome that is unrelated to efficacy such as side-effects or costs. Guidelines for non-inferiority trials were recently published.13 With regard to the non-inferiority margin they suggest that the difference between the test intervention and the active comparator should be less than 50% of the difference between the active comparator and placebo. Furthermore, the difference in response rates between the test intervention and the active control should be no more than 5%.14 However, the guidelines for non-inferiority trials do not include data or suggestions to determine sample size or power analysis.14 Table 1 lists the sample size per group that is needed to demonstrate non-inferiority with a power of 0.80 for a variety of non-inferiority margins. For a difference of 5% or less, 1579 patients per group would be needed (p1=p2=0.5). Samples of this size are difficult to achieve in psychotherapy and

---

**Table 1**

<table>
<thead>
<tr>
<th>Equivalence means</th>
<th>Equivalence proportions</th>
<th>Superiority/ non-inferiority means</th>
<th>Superiority/ non-inferiority proportions</th>
<th>Non-inferiority proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10/0.056</td>
<td>1.00/0.475</td>
<td>0.50/0.475</td>
<td>0.00/0.475</td>
<td></td>
</tr>
<tr>
<td>0.20/0.112</td>
<td>1.35/0.880</td>
<td>0.90/0.880</td>
<td>0.10/0.880</td>
<td></td>
</tr>
<tr>
<td>0.30/0.168</td>
<td>1.48/0.940</td>
<td>0.60/0.940</td>
<td>0.10/0.940</td>
<td></td>
</tr>
<tr>
<td>0.40/0.223</td>
<td>2.00/1.000</td>
<td>0.50/0.950</td>
<td>0.00/0.950</td>
<td></td>
</tr>
<tr>
<td>0.50/0.276</td>
<td>1.00/0.975</td>
<td>0.40/0.975</td>
<td>0.00/0.975</td>
<td></td>
</tr>
<tr>
<td>0.60/0.329</td>
<td>1.00/0.975</td>
<td>0.30/0.975</td>
<td>0.00/0.975</td>
<td></td>
</tr>
<tr>
<td>0.70/0.379</td>
<td>1.00/0.975</td>
<td>0.20/0.975</td>
<td>0.00/0.975</td>
<td></td>
</tr>
<tr>
<td>0.80/0.428</td>
<td>1.00/0.975</td>
<td>0.10/0.975</td>
<td>0.00/0.975</td>
<td></td>
</tr>
<tr>
<td>0.90/0.475</td>
<td>1.00/0.975</td>
<td>0.00/0.975</td>
<td>0.00/0.975</td>
<td></td>
</tr>
</tbody>
</table>

All calculations were performed with nQuery software. If not otherwise specified, α was set at 0.05. TOST=two one-sided tests. Data are number of participants per group in relation to Cohen's d/success rate difference (according to Kraemer and Kupfer).10a For testing equivalence or non-inferiority in proportions, we calculated the required sample sizes for p1=p2=0.5; TOST 90% CI (1−2α)×100%. For other proportions, the sample size per group needs to be specifically calculated. For p1=p2=0.20, for example, 35 patients per group are needed to show equivalence if a margin of 0.267 is accepted. If the success' rate of the new treatment is expected to exceed the success rate of the standard treatment, relatively low sample sizes may be needed to demonstrate non-inferiority, eg, 55% vs 60% using a margin of −0.12 requires 2×145 patients (97.5% CI). If 60% success is expected for the standard and 55% for the new treatment, 2×590 are needed, all other parameters being constant.11 One two-sided test α=0.05, one one-sided test α=0.025 lower 97.5% CI limit. §α=0.05 one two-sided test: p1=0.5 p2=0.5 + SRD. ||p1=p2=0.5 lower 97.5% CI limit.

---

**Panel 1: What is new?**

- We review the criteria for evidence-based psychotherapy by differentiating between superiority, non-inferiority and equivalence trials
- We update the criteria for comparisons with efficacious treatments (ie, equivalence margin, two one-sided test procedure [TOST], sample size, and power)
- We provide guidelines to review superiority and equivalence trials
- We apply these updated criteria to present evidence of psychodynamic therapy
- We present a systematic review on the evidence for psychodynamic therapy
pharmacotherapy research. For non-inferiority trials, more pragmatic solutions or intermediate steps are needed. Equivalence and non-inferiority trials rely on the assumptions that the superiority of the active control compared with placebo or other treatment has been previously established and will be maintained during the trial and that the trial is sufficiently similar to previous trials that showed the active comparator to be superior to placebo or other treatments.\(^1\) In particular, the patient population and the treatment characteristics must be consistent with the previous trials.\(^1\)

**Guidelines for reviewing equivalence and non-inferiority trials**

**Reviews**

Recently, recommendations for reviewing equivalence and non-inferiority trials were presented.\(^1\) For comparisons of PDT with treatments of established efficacy, we followed these recommendations by (1) defining a margin of equivalence, (2) taking sample size and statistical power into account and (3) using direct rather than indirect language when concluding equivalence or non-inferiority (wording such as “there was no statistical difference” may be factually correct to describe the result of a study, but a conclusion regarding equivalence should be directly described as “not inferior to” or “as effective as”).\(^1\) Additionally, (4) we looked for possible biases that may lead to either underestimation or overestimation of differences between treatments (i.e., deviations in patients, measures, or treatments from those in the trials which established efficacy of the active comparator).\(^1\) To fulfil these criteria, we (1) preliminarily adopted the proposal by Chambless and Hollon of a “moderate” difference between treatments as an equivalence margin and (2) examined whether the statistical power of the included studies was sufficient to demonstrate equivalence, had the logic of the TOST procedure been applied. Thus, for the time being we regarded RCTs as sufficiently powered (≥0.80) to show equivalence (in means) of an active psychotherapy (e.g., PDT) to an established treatment if the data analysis was based on at least n=70 patients in each condition, using a moderate effect size of d=0.5 as an equivalence margin and a significance level of α=0.05 for each of the two one-sided tests (TOST, table 1).

**Meta-analyses**

Some individual studies might not fulfil the power criterion described above. By use of meta-analysis to aggregate a series of studies, a statistical power can be achieved that is higher than that of the individual studies.\(^1\) In a meta-analysis, statistical power depends on the number of studies included, the number of patients per study, and the degree of heterogeneity between studies.\(^1\) Hedges and Pigott\(^1\) proposed a convention for small, medium, and large levels of heterogeneity. This convention is to set the between-study variance equal to 0.33, 0.67, or 1.0 times of the within-study variance. For small, medium, and large levels of heterogeneity and with the assumption of 25 patients per group, only four, five, and six studies, respectively, are required to detect a medium effect size of d=0.5 with a power of 0.80 or above at α=0.05 by two-sided test using the random effects model (three, four, and five studies for a one-sided test). For this calculation, we used the formulas provided by Hedges and Pigott\(^1\) and Borenstein and colleagues.\(^1\) For a small difference of d=0.2, the number of studies required would be 21, 27, and 32 for a two-sided test and 17, 21, and 25 for a one-sided test, respectively.

**Empirical evidence for PDT**

We identified 17 studies that were not included in our published systematic review\(^1\) on PDT in specific mental disorders.\(^1\) We applied our updated criteria to the 64 RCTs that were identified in this search. Outcomes refer to either the target symptoms of the respective disorders (e.g., depression in depressive disorders), or the comorbid symptoms (e.g., depression in anxiety disorders), or social and personality functioning as assessed in the respective study. In most RCTs, short-term to medium-term (8–40 sessions) PDT was studied (table 2). Several studies also included long-term (12–36 months) PDT.

**Depressive disorders**

In several RCTs, PDT was superior to waiting list control conditions or alternative treatments for the improvement of depression (table 2).\(^2\) PDT was also superior to treatment as usual in patients with maternal depression\(^2\) and patients with breast cancer.\(^10\) Internet guided psycho-dynamic self-help was superior to internet-delivered structured support.\(^3\) Furthermore, PDT combined with pharmacotherapy was superior to pharmacotherapy alone or combined with supportive therapy in major depressive disorder.\(^1\) Results from a small pilot study\(^16\) showed large effect sizes in favour of PDT compared with treatment as usual, but the study was not sufficiently powered for a superiority trial and the differences did not achieve statistical significance (table 2).
## Major depressive disorder

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barber et al, 2012*</td>
<td>51</td>
<td>Pharmacotherapy: n=55; placebo: n=50</td>
<td>20 sessions; 16 weeks</td>
</tr>
<tr>
<td>Barkham et al, 1996*</td>
<td>18</td>
<td>CBT: n=18</td>
<td>8 vs 16 sessions</td>
</tr>
<tr>
<td>Connolly Gibbons et al, 2012*</td>
<td>21</td>
<td>TAU: n=19</td>
<td>12 sessions</td>
</tr>
<tr>
<td>Driessen et al, 2013*</td>
<td>177</td>
<td>CBT: n=164</td>
<td>16 sessions</td>
</tr>
<tr>
<td>Cooper et al, 2003*</td>
<td>50</td>
<td>CBT: n=43; counselling: n=48; treatment as usual: n=52</td>
<td>10 sessions</td>
</tr>
<tr>
<td>de Jonghe et al, 2004*</td>
<td>106</td>
<td>PDT plus pharmacotherapy: n=85</td>
<td>16 sessions</td>
</tr>
<tr>
<td>Johansson et al, 2012*</td>
<td>46</td>
<td>Structured support: n=46</td>
<td>10 weeks</td>
</tr>
<tr>
<td>Salminen et al, 2008*</td>
<td>26</td>
<td>Fluoxetine: n=25</td>
<td>16 sessions</td>
</tr>
<tr>
<td>Shapiro et al, 1994*</td>
<td>58</td>
<td>CBT: n=59</td>
<td>8 vs 16 sessions</td>
</tr>
</tbody>
</table>

### Depressive caregivers

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gallagher-Thompson &amp; Steffen, 1994*</td>
<td>30</td>
<td>CBT: n=36</td>
<td>16–20 sessions</td>
</tr>
</tbody>
</table>

### Dysthymic disorder

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maina et al, 2005*</td>
<td>10</td>
<td>Supportive therapy: n=10; waiting list: n=10</td>
<td>15–30 sessions; Mean=19.6</td>
</tr>
</tbody>
</table>

### Depressive disorders in patients with breast cancer

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beutel et al, 2014*</td>
<td>78</td>
<td>Treatment as usual: n=79</td>
<td>Up to 25 sessions</td>
</tr>
</tbody>
</table>

### PDT combined with pharmacotherapy in MDD

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burndal et al, 2002*</td>
<td>35</td>
<td>Clomipramine: n=39</td>
<td>10 weeks</td>
</tr>
<tr>
<td>de Jonghe et al, 2001*</td>
<td>72</td>
<td>Pharmacotherapy: n=57</td>
<td>16 sessions</td>
</tr>
<tr>
<td>Maina et al, 2007*</td>
<td>18</td>
<td>Brief supportive therapy combined with pharmacotherapy: n=17</td>
<td>15–30 sessions</td>
</tr>
</tbody>
</table>

### Mixed samples of patients with depressive and/or anxiety disorders

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bressi et al, 2010*</td>
<td>30</td>
<td>Treatment as usual: n=30</td>
<td>40 sessions; 1 year</td>
</tr>
<tr>
<td>Johansson et al, 2013*</td>
<td>50</td>
<td>Supportive interventions: n=50</td>
<td>10 weeks</td>
</tr>
<tr>
<td>Knekt et al, 2008*</td>
<td>STPP: 101</td>
<td>SFT: n=97</td>
<td>LTTP: 222 sessions; STPP: 18.5 sessions; SFT: 9.8 sessions</td>
</tr>
</tbody>
</table>

### Complicated grief

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCallum and Piper, 1990*</td>
<td>27</td>
<td>Waiting list: n=27</td>
<td>12 sessions</td>
</tr>
<tr>
<td>Piper et al, 2001*</td>
<td>53</td>
<td>Supportive therapy: n=54</td>
<td>12 sessions</td>
</tr>
</tbody>
</table>

### Social anxiety disorder

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bogels et al, 2014*</td>
<td>22</td>
<td>CBT: n=22; waiting list: n=27</td>
<td>35 sessions</td>
</tr>
<tr>
<td>Knijnik et al, 2004*</td>
<td>15</td>
<td>Credible placebo control group: n=15</td>
<td>12 sessions</td>
</tr>
<tr>
<td>Leichsenring et al, 2013*</td>
<td>207</td>
<td>Cognitive therapy: n=209; waiting list: n=79</td>
<td>30 sessions</td>
</tr>
</tbody>
</table>

### PDT combined with pharmacotherapy in social anxiety disorder

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knijnik et al, 2008*, 2009*</td>
<td>29</td>
<td>Pharmacotherapy: n=29</td>
<td>12 sessions</td>
</tr>
</tbody>
</table>

### Generalised anxiety disorder

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leichsenring, Salzer et al, 2009*</td>
<td>28</td>
<td>CBT: n=29</td>
<td>30 sessions</td>
</tr>
<tr>
<td>Andersson et al, 2012*</td>
<td>27</td>
<td>ICBT: n=27; waiting list: n=27</td>
<td>8 weeks</td>
</tr>
</tbody>
</table>

### Panic disorder

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milrod et al, 2007*</td>
<td>26</td>
<td>Applied relaxation: n=23</td>
<td>24 sessions</td>
</tr>
<tr>
<td>Beutel et al, 2013*</td>
<td>36</td>
<td>CBT: n=18</td>
<td>24 sessions</td>
</tr>
</tbody>
</table>

### PDT combined with pharmacotherapy in panic disorder

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiborg &amp; Dahl, 1996*</td>
<td>20</td>
<td>Pharmacotherapy alone: n=20</td>
<td>15 sessions</td>
</tr>
</tbody>
</table>

### PDT combined with pharmacotherapy in obsessive-compulsive disorder

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maina et al, 2010*</td>
<td>27</td>
<td>Pharmacotherapy: n=10</td>
<td>10–16 sessions</td>
</tr>
</tbody>
</table>

### Post-traumatic stress disorder

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brom et al, 1989*</td>
<td>29</td>
<td>Desensitisation: n=31; hypnotherapy: n=29; waiting list: 23</td>
<td>18.8 sessions</td>
</tr>
</tbody>
</table>

### Somatoform disorders

<table>
<thead>
<tr>
<th>Study</th>
<th>N (PDT)</th>
<th>Comparison group</th>
<th>Duration of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creed et al, 2003*</td>
<td>59</td>
<td>Paroxetine: n=43; treatment as usual: n=86</td>
<td>8 sessions</td>
</tr>
<tr>
<td>Faramarzi et al, 2013*</td>
<td>24</td>
<td>Medical treatment: n=25</td>
<td>16 sessions</td>
</tr>
<tr>
<td>Guthrie et al, 1991*</td>
<td>50</td>
<td>Supportive listening: n=46</td>
<td>8 sessions</td>
</tr>
</tbody>
</table>

(Table 2 continues on next page)
With regard to comparisons of established treatments such as CBT or pharmacotherapy, two studies were sufficiently powered to test for equivalence or non-inferiority. No significant differences in outcome were found in these RCTs. However, the first study applied the traditional two-sided test rather than the TOST.
procedure. The other RCT tested for non-inferiority. In this RCT, non-inferiority of PDT compared with CBT was not shown for remission rates but was shown for continuous measures of depression post treatment. However, the difference in remission rates (21% for PDT vs 24% for CBT) was minimal from a clinical perspective. Results from several RCTs showed no differences in outcome between PDT and treatments with known efficacy, but the studies were not sufficiently powered to demonstrate equivalence if the criterion of at least 70 patients per group is applied. Because meta-analyses of relatively few studies achieve a higher statistical power than individual studies, it is of note that results from several meta-analyses showed individual PDT to be efficacious in depressive disorders with no differences to other established treatments.

Complicated grief
The efficacy of PDT in complicated grief was demonstrated in two RCTs. In these studies, PDT was superior to a waiting list condition or a supportive treatment.

Anxiety disorders
Results from several RCTs have provided evidence for the efficacy of PDT in the treatment of anxiety disorders (table 2). For panic disorder, PDT was superior to applied relaxation. In an RCT of PDT versus CBT in panic disorder, no significant differences in remission rates were found, but the study was not sufficiently powered to show equivalence. For social anxiety disorder, PDT was superior to a credible placebo or as efficacious as CBT. Results from a recent RCT showed PDT to be superior to a waiting list control condition and to be as efficacious as CBT in all outcomes (social anxiety, general psychopathology, and defence mechanisms). Success rates were above 50% and were found to be stable at the 3-month and 12-month follow-up. For comparison with CBT, the authors reported the study to be sufficiently powered to detect medium differences. A large multicentre RCT, both PDT and CBT were superior to a waiting list control condition. No differences between PDT and CBT were found with regard to response rates and reduction of depression. CBT showed statistically significant better outcomes on remission rates, self-reported social anxiety and interpersonal problems, yet these differences were small and below the threshold defined a priori as clinically meaningful. Furthermore, there were no differences in long-term effects on any outcome measure 6, 12, and 24 months after the end of therapy. Although originally designed as a superiority study to detect small differences in outcome, this study was sufficiently powered to test for equivalence if the power criterion proposed above is applied.

For generalised anxiety disorder, results from one study showed no significant differences between PDT and CBT in the primary outcome measure (Hamilton Anxiety Rating Scale), but the study was not sufficiently powered for an equivalence trial. In secondary measures (eg, worry and depression), CBT achieved statistically significant better outcomes. Treatment effects were stable 12 months after the end of therapy. In another RCT, internet-guided psychodynamic self-help proved to be superior to a waiting list control condition in generalised anxiety disorder. No differences compared with internet-guided CBT were reported, but this study was not sufficiently powered to demonstrate equivalence (table 2).

For a mixed sample including various categories of anxiety disorders, short-term PDT was superior to long-term PDT (and as efficacious as solution-focused therapy) with regard to recovery at the 7-month follow-up. In a mixed sample of patients with either depressive disorder or anxiety disorders, or both, PDT was superior to treatment as usual (pharmacotherapy).

Combination of PDT plus pharmacotherapy was shown to be superior to pharmacotherapy alone in the treatment of social anxiety disorder and panic disorder. For panic disorder, rates of remission and relapse prevention in PDT combined with pharmacotherapy was superior to pharmacotherapy alone.

Post-traumatic stress disorder
Results from an RCT showed no significant differences in outcome between PDT, hypnotherapy, and CBT. However, this study was not sufficiently powered to show equivalence (table 2). PDT was superior to a waiting list control condition in two of three measures and achieved the largest within-group effect sizes at follow-up.

Somatoform disorders or somatic symptom disorder
There is a substantial body of evidence for the efficacy of PDT in somatoform disorders, now referred to as somatic symptom disorder in DSM-5 (table 2). Evidence from RCTs is available for irritable bowel syndrome, functional dyspepsia, and somatoform pain disorder. In each of these RCTs, PDT was superior to treatment as usual or supportive therapy. Furthermore, results from a meta-analysis showed PDT to be efficacious in patients with somatic disorders.

Eating disorders
Results from a bulimia nervosa study showed that PDT was superior to CBT and nutritional counselling. Results from two other studies showed no difference in primary outcome measure (bulimic episodes and vomiting) between PDT, and CBT but these studies were
not sufficiently powered to demonstrate equivalence (table 2). Differences in favour of CBT were found in secondary measures.\(^{43,57}\) In another RCT,\(^{58}\) CBT was superior to PDT, but the study was controversial because PDT was manualised but not symptom-focused.\(^{56,57}\)

Results from two studies provided evidence for PDT in anorexia nervosa.\(^{60-66}\) One RCT compared manual-guided PDT, enhanced CBT, and optimised treatment as usual in the treatment of anorexia nervosa.\(^{66}\) At the end of treatment, significant improvements were noted in all treatments, with no differences in the primary outcome measure (body-mass index, BMI). At 12-month follow-up, however, PDT was significantly superior to optimised treatment as usual on rates of recovery, whereas enhanced CBT was not significantly superior.\(^{66}\) Recovery rates were 35% for PDT, 19% for enhanced CBT and 13% for optimised treatment as usual. This study was sufficiently powered to show equivalence (table 2).\(^{66}\) Two eating-disorder studies were not sufficiently powered to demonstrate equivalence to active treatments (table 2).\(^{43,55}\) For the comparison with routine treatment\(^{45}\) and waiting list,\(^{45}\) these studies were sufficiently powered to show superiority.

### Personality disorders

Results from two meta-analyses\(^{59,60}\) showed PDT to be efficacious in the treatment of personality disorders.

#### Cluster C personality disorders

Evidence for PDT in cluster C personality disorders was reported by two RCTs\(^{68,69}\) with no differences in outcome compared with CBT. These studies were not sufficiently powered to demonstrate equivalence (table 2). Results from another RCT showed CBT to be more effective than waiting list control and PDT in the treatment of avoidant personality disorder.\(^{61}\) The study has attracted controversy because of possible biases,\(^{106,107}\) for example, whether a disorder-specific manual-guided bona-fide version of PDT was used.\(^{106,107}\)

#### Cluster B personality disorders

Results from several RCTs show that borderline personality disorder can be successfully treated with PDT.\(^{70-76}\) Bateman and Fonagy\(^{72}\) showed that PDT (mentalisation-based therapy, MBT) was superior to a day treatment. In a second RCT, Bateman and Fonagy\(^{74}\) showed that MBT was superior to a manual-guided structured clinical management. MBT was superior with regard to self-reported and clinically significant problems, such as suicide attempts and hospital admission. Clarkin and colleagues\(^{75}\) compared two types of PDT (transference-focused therapy [TFP] and supportive psychodynamic therapy [SPT]) to dialectical behaviour therapy (DBT). No differences were found between the three treatments in several outcome measures, but TFP was superior to DBT with regard to improvements in reflective functioning and attachment.\(^{76}\) Only TFP and SPT were associated with improvements in anger and impulsivity, and only TFP was associated with change in irritability and verbal and direct assault.\(^{77}\) TFP produced significant improvements in ten of 12 outcome measures, DBT in five of 12 measures and SPT in six of 12 measures.\(^{77}\) The efficacy of TFP was corroborated by another RCT showing that TFP was superior to a treatment by experienced community therapists.\(^{78}\) Gregory and colleagues\(^{79}\) reported PDT (deconstructive therapy) to be superior to a treatment as usual condition in borderline patients with co-occurring alcohol use disorder. Another RCT compared PDT (ie, transference-focused psychotherapy, TFP) with schema-focused therapy (SFT).\(^{80}\) The authors reported statistically and clinically significant improvements for both treatments. However, SFT was reported to be superior to TFP in several outcome measures. Furthermore, a significantly higher dropout was reported in TFP. Concerns on the methodology used in this study, in particular regarding treatment integrity of TFP have been published.\(^{81,82}\) The two studies that compare PDT to another active treatment were not sufficiently powered to show equivalence, but both studies reported superiority of PDT or SFT at least in some measures.\(^{76,78}\)

### Heterogeneous samples of patients with personality disorders

In two RCTs,\(^{113,114}\) PDT was superior to waiting list control condition or minimal contact conditions in samples with heterogeneous personality disorders. Results from another RCT\(^{76}\) showed no differences in outcome between PDT and brief supportive therapy. However, two of these studies\(^{43,55}\) were not sufficiently powered to detect possible differences between PDT and the active comparators.

In a sufficiently powered study, manual-guided PDT was as effective as community-delivered PDT.\(^{83}\)

### Substance-related disorders

For opiate dependence, results from two RCTs\(^{67,69}\) provided evidence for the efficacy of PDT in several outcomes (eg, days worked, drug use, illegal income, depression, and general psychiatric symptoms). In the earlier study by Woody and colleagues,\(^{67}\) both PDT and CBT were superior to drug counselling (standard treatment). No differences were found between PDT and CBT, but the studies were not sufficiently powered to test for equivalence. In the later study by Woody and colleagues,\(^{69}\) PDT was superior to drug counselling. Thus, PDT proved to be efficacious in opiate addiction. By contrast, both PDT and CBT were reported to be inferior to individual drug counselling for cocaine dependence.\(^{79,80}\)

### Obsessive-compulsive disorder

In the only published RCT\(^{70}\) of PDT in obsessive-compulsive disorder, PDT combined with pharmacotherapy was not superior to pharmacotherapy alone.
High users of psychiatric services
PDT was superior to a treatment as usual condition in high users of psychiatric services. The sample primarily included patients with depressive and anxiety disorders.

Relationship distress: marital therapy
No significant differences were found between PDT and a behavioural couple therapy with regard to individual and relationship functioning. Both treatments were superior in this regard to a waiting list control group. Effects were maintained at 6-month follow-up. At 4-year follow-up, significantly more couples in the behavioural condition than in PDT had divorced (38% vs 3%). The study was sufficiently powered to show superiority, but not equivalence.

Long-term psychodynamic psychotherapy in complex mental disorders
In several meta-analyses, long-term psychodynamic psychotherapy (LTPP), which was defined as involving at least 50 sessions or last for at least 1 year, was superior to shorter or less intensive forms of treatment in patients with complex mental disorders, defined as chronic mental disorders, personality disorders, or multiple comorbid disorders. Superiority was shown for improvements in target problems, general psychiatric symptoms, personality and social functioning. Results from these studies are consistent with data on dose-effect relations, which suggest that for many patients with chronic mental disorders or personality disorders, short-term psychotherapy is not sufficient. For these patients, long-term treatments seem to be more effective. The statistical analysis and comparison conditions used in this meta-analysis have been reviewed and critiqued. Concerns have been addressed in several publications. Meta-analysis results were corroborated if, for example, only active comparison conditions were included, studies previously not included were considered and between-group effect sizes were analysed. Thus, in complex mental disorders, LTPP proved to be superior to shorter or less intensive treatments. Analogously, we expect other forms of long-term psychotherapy to be superior to shorter forms of these treatments in complex mental disorders. DBT or SFT for borderline personality disorders, for example, are also long-term treatments.

Discussion
PDT is frequently used in clinical practice. Efficacy research, however, has been neglected in PDT for a long time. There remain concerns among some psychodynamic therapists and researchers about applying the methodology of RCTs to PDT. Some psychodynamic therapists and researchers remain uncertain about the clinical use of RCTs for PDT. For example, the study of unconscious conflicts or processes poses a unique challenge to research on PDT. However, the outcome of PDT in the form of observable manifestations of improvement can be studied. With regard to the problem of treatment manualisation, the available RCTs using treatment manuals show that the complex interpersonal process of psychodynamic therapy can be manualised (table 2)—treatment manuals should not be mistaken as cookbooks. Present manuals allow for a wide range of flexibility in therapist behaviour. Even long-term PDT can be manual-guided as shown by the RCTs by Bateman and Fonagy, Clarkin and colleagues, and Vinnars and colleagues. Furthermore, the methodological quality of PDT studies was shown to be comparable to that of CBT studies, which suggests that RCT methodology could be as adequately applied to PDT as to CBT.

In recent years, efficacy research for PDT has increased, and evidence for its efficacy is beginning to accumulate. Results from our systematic review suggest that there is substantial evidence for the efficacy of PDT in depressive, anxiety, somatoform, eating, substance-related, and personality disorders. This level of evidence is consistent with a recent Cochrane Report that found PDT to be efficacious in common mental disorders. Effects of PDT were found to be stable or increased in follow-up assessments.

Although there is a growing body of evidence for the use of PDT to treat mental health disorders, there are also some limitations. Application of the updated inclusion criteria proposed in this review results in only six of the 64 studies of PDT being sufficiently powered to show equivalence to an established treatment. However, studies that compare CBT to an established treatment are not more highly powered: only two of 26 studies that compared psychotherapy with pharmacotherapy included at least 70 patients per group. In the psychotherapy research field, RCTs that explicitly test equivalence as defined in this Review are still extremely rare. Of the 64 RCTs in this Review, none were conceptualised as equivalence trials. The only equivalence study we found in the general literature addressed the cost-effectiveness of CBT for health anxiety. In medical research however, there has been an increase in the number and quality of equivalence and non-inferiority studies.

PDT was inferior to an efficacious treatment in only one of the six studies sufficiently powered according to the criteria we applied here. In this study, both PDT and CBT were inferior to individual drug counselling in the treatment of cocaine dependence. No substantial differences in efficacy between PDT and CBT were found in those studies sufficiently powered to test equivalence. Furthermore, clinical meaningfulness of small differences is questionable. The between-group effect sizes for comparisons of PDT between bona-fide therapies were generally found to be small, both in large-scale individual studies and in meta-analyses. Results of these meta-analyses led to greater confidence in the assumption that there are no
Search strategy and selection criteria

Studies included in this Review met the following criteria: randomised control trials (RCTs) of psychodynamic therapy (PDT) as defined in the introduction section; use of reliable and valid measures for diagnosis and outcome such as the structured clinical interview for DSM (SCID) or Hamilton Depression Rating Scale or Hamilton Anxiety Rating Scale; an adult sample treated for specific problems; use of treatment manuals or manual-like guidelines (ie, clear descriptions of a treatment that include the theoretical background, a set of technical recommendations for this specific psychodynamic treatment such as indications, interventions or timing, and detailed case examples). We searched PubMed and PsycINFO for studies of PDT in adults published between Jan 1, 1970, and March 14, 2015. The following search terms were used: (psychodynamic or dynamic or psychoanalytic*) and (therapy or psychotherapy or treatment) and (study or studies or trial*) and (outcome* or result or results or effect* or change*) and (psychiat* or mental* or psychol*) and (RCT* or control* or compar*). Manual searches of previous reviews and meta-analyses, textbooks, and reference lists of the included studies were also done. After completing literature searches, all hits (n=3139) were saved in EndNote. After removal of duplicates (n=314), two authors (CS, FaL) independently screened the titles and abstracts of the resulting 2825 articles using the selection criteria. Disagreements were solved by consensus. All potentially relevant articles were then retrieved for full-text review, which resulted in the inclusion of 64 RCTs (table 2).

Differences between bona-fide treatments because meta-analyses have greater statistical power than individual studies. If future research confirms that there are no substantial differences in outcome between the different forms of bona-fide psychotherapy in common mental disorders, the next question is which patients benefit more from which kind of therapy.

Treatment integrity of the PDTs investigated in this review is important. Treatment integrity is defined as the extent to which a treatment is carried out as intended, and is a crucial factor in psychotherapy research. Treatment integrity is often inadequately addressed, and this has been shown for all forms of psychotherapy including CBT as for CBT more studies are available than for other approaches. Further, the relation between adherence to a treatment model and the competent delivery of interventions or outcome is heterogeneous. At present it is not really clear which interventions of a treatment package such as MBT or DBT are associated with improvements. In several studies, however, significant relations between psychodynamic interventions and outcome were demonstrated. These studies provide another type of evidence for PDT, ie, that outcome is related to psychodynamic interventions. Furthermore, there is evidence to suggest that changes in psychodynamically relevant mediators such as core conflicts and insight are significantly related to outcome.

Different PDT approaches were used in the 64 RCTs included in this review. However, from an empirical perspective, how different the approaches are is not clear. Approaches for the treatment of anxiety disorders and depressive disorders were shown to be consistent with each other with a large overlap. Thus, the development of unified or transdiagnostic protocols for the psychodynamic treatment of common mental disorders is possible and could be an important target for future research and practice.

In this review, we used rigorous criteria to assess PDT evidence. The need for such rigorous criteria is demonstrated by a recent flawed meta-analysis that claimed to test the so-called dodo bird hypothesis of no differences in efficacy between bona-fide treatments. Results from this meta-analysis showed PDT to be inferior to CBT. However, the comparison of PDT to CBT was based on a selected sample of studies because only three studies of PDT were included. Of these studies, none could be considered as representative of bona-fide PDT. In the first study, no treatment manual was used and therapists were not trained for the study. In the second study, only two plus one sessions were offered to individuals with subsyndromal depression. Results from the third study showed differences in treatment integrity between PDT and CBT that questioned the internal and construct validity of the study. Furthermore, several RCTs that compared bona-fide PDT with CBT were omitted by Marcus and colleagues. These include several studies of depressive disorders, anxiety disorders, post-traumatic stress disorder, eating disorders, substance abuse disorders, and personality disorders. In total, 22 RCTs that compare PDT and CBT were not included in this meta-analysis—that is, almost eight times as many relevant studies were missed than were included. New relevant studies have also been published in 2014. In summary, the results reported by Marcus and colleagues are not consistent with several recent reviews and meta-analyses. Marcus and colleagues reported a between-group effect size of 0.16 for the primary outcome when comparing CBT to other treatments. The between-group effect size of 0.16 is a minimal difference according to established conventions, the clinical significance of which is not clear. This small difference is essentially consistent with the dodo bird hypothesis. Another recent network meta-analysis, whereby different treatments are compared by statistical inference, reported that PDT was superior to no-treatment only, but was not different from both pill and psychological placebo and was inferior to CBT. Results from this meta-analysis also shows several severe limitations, such as a small number of PDT studies, which have been discussed elsewhere.
Several open questions remain that require further research on PDT: new RCTs are needed, particularly for disorders such as obsessive-compulsive disorder and post-traumatic stress disorder for which only one relatively old RCT exists. More adequately powered equivalence trials are needed. Future studies on PDT should measure not only symptoms or DSM criteria, but also measures that are more specific to PDT. Future studies should also examine whether there are specific gains achieved only by PDT. Added value of PDT was demonstrated, for example, by Levy and colleagues who compared improvements in reflective functioning and attachment between PDT and DBT. To further improve PDT, future research should address the mechanisms of change. The question of “what works for whom” should also be examined.

Contributors
Fal and CS conceptualised and designed the Review, did literature searches, assessed eligibility of the studies for inclusion, extracted the data, and double checked extracted information. Fal drafted the report, to which all authors contributed significantly. All authors critically revised the Review for important intellectual content. All authors had access to all data and the final manuscript has been approved by all authors.

Declaration of interests
We declare no competing interests.

Acknowledgments
We thank Emmanuel Lesaffre (L-Biostat, KU Leuven, Belgium) for helpful comments on statistical analysis.

References


50 Wilborg IM, Dahl AA. Does brief dynamic psychotherapy reduce the relapse rate of panic disorder? *Arch Gen Psychiatry* 1996; 53: 689–94.


68 Woody GE, Luborsky L, McIllean AT, O’ Brien CP. Corrections and revised analyses for psychotherapy in methadone maintenance patients. *Arch Gen Psychiatry* 1990; 47: 788–89.


