Analyzing Effectiveness of Long-term Psychodynamic Psychotherapy

Aaron T. Beck; Sunil S. Bhar

JAMA. 2009;301(9):931 (doi:10.1001/jama.2009.179)

http://jama.ama-assn.org/cgi/content/full/301/9/931

Analyzing Effectiveness of Long-term Psychodynamic Psychotherapy

To the Editor: Drs Leichsenring and Rabung[1] reported that long-term psychodynamic psychotherapy (LTPP) is more effective than shorter forms of psychotherapy for complex mental disorders based on a between-group effect size of 1.8 from 7 comparative trials that they meta-analyzed. The authors did not indicate that they were concerned about this and other surprisingly large effect sizes they reported.

Between-group effect sizes can be presented as group differences in terms of standard deviations or as point biserial correlations between group (eg, LTPP vs shorter-term therapies) and treatment effect. They are equivalent and convertible using a formula or tables.[2] The authors, however, apparently erroneously calculated within-group pre-post effect sizes and point biserial correlations between group and within-group effect sizes, which is altogether different. It seems that they converted these correlations between group and within-group pre-post effect sizes to produce deviation-based effect sizes that do not appear reasonable.

As a result, although none of the 7 studies had an overall effect size greater than 1.45, the authors reported a combined effect size of 1.8, which is statistically impossible. In a slightly larger set of 8 trials, the authors reported that LTPP had a larger overall effect than shorter-term therapies (0.96 vs 0.47) but a point biserial correlation (0.60) equivalent to a between-group effect size of 1.5. However, between-group effect sizes must be smaller than within-group effect sizes when both groups have positive effects. Similarly, these methods generated an implausible between-group effect size of 6.9 for personality functioning based on 3 trials,[3-5] none of which reported an effect size for personality functioning larger than approximately 2.

In addition, we believe that this collection of studies was not suitable for meta-analysis. Each reviewed study had 15 to 30 patients in the LTPP treatment group. It seems unlikely that investigators would attempt to publish a negative study with so few patients (or that such a study would be accepted for publication), which means that all published studies would have an effect size of at least 0.50 to 0.75, the minimum for statistical significance. This is an artificial floor that guarantees a large effect when these studies are combined.

Brett D. Thombs, PhD
brett.thombs@mcgill.ca
Marielle Bassel, BA
Lisa R. Jewett, BA
Department of Psychiatry
McGill University
Montreal, Quebec, Canada

Financial Disclosures: None reported.

1. Leichsenring F, Rabung S. Effectiveness of long-term psychodynamic psychotherapy: a meta-analysis. *JAMA*. 2008;300(13):1551-1565.
2. McCartney K, Rosenthal R. Effect size, practical importance, and social policy for children. *Child Dev*. 2000;71(1):173-180.
3. Bachar E, Latzer Y, Kreitler S, Berry EM. Empirical comparisons of two psychological therapies: self psychology and cognitive orientation in the treatment of anorexia and bulimia. *J Psychother Pract Res*. 1999;8(2):115-128.
4. Bateman A, Fonagy P. Effectiveness of partial hospitalization in the treatment of borderline personality disorder: a randomized controlled trial. *Am J Psychiatry*. 1999;156(10):1563-1569.
5. 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(6):1027-1040.

To the Editor: In their meta-analysis on the effectiveness of LTPP, Drs Leichsenring and Rabung[1] reported evidence that LTPP is an effective treatment for complex mental disorders. Although the article addressed several limitations of the findings, other limitations should be considered as well.

First, Leichsenring and Rabung formulated an unconventionally broad research question. Systematic review objectives typically define the patient population, intervention, control treatment, outcomes, and study designs of interest.[2] These authors defined the patient population broadly as a group of adults with mental disorders. The sole criterion that outcomes should have fulfilled is that they were reliable. Control treatments were not defined at all. Thus, analyzed diagnoses, outcomes, and control groups showed a very large clinical heterogeneity. Although the reviewers tried to account for this through subgroup analyses, their method of building clusters of heterogeneous disorders and outcomes still allowed for considerable variation. Thus, it is unclear for what disorder, for what outcome variables, and in comparison with which control groups the evidence was shown.

Second, the inclusion of randomized controlled trials (RCTs) alone does not warrant internal validity of the findings. The comparison of LTPP to other treatments was performed by calculating point biserial correlations between

Letters Section Editor: Robert M. Golub, MD, Senior Editor.

the within-group effect sizes and type of treatment. This procedure considers treatment groups rather than studies as the unit of analysis and thus nullifies the effect of randomization, reducing findings to an observational level. Accordingly, internal validity of the findings may be biased.

Third, although the reviewers did not find any sign of publication bias, the forest plot suggests a positive association between effect size and meta-analytical study weight. Depending on how this weight was calculated (which was not reported), the reasons for this association may have an important effect on the interpretation. Furthermore, for tests of possible publication bias, the application of the standard error of effect size rather than sample size is recommended.[3]

We greatly appreciate the high-quality work of the authors on this important issue but at the same time caution against a conclusion that the overall effectiveness of LTPP for treating mental disorders should now be considered as definitely proven.

Levente Kriston, PhD
l.kriston@uke.uni-hamburg.de
Department of Medical Psychology
University Medical Center Hamburg-Eppendorf
Hamburg, Germany
Lars Hölzel, MA
Department of Psychiatry and Psychotherapy
University Medical Center Freiburg
Freiburg, Germany
Martin Härter, MD, PhD
Department of Medical Psychology
University Medical Center Hamburg-Eppendorf
Hamburg

Financial Disclosures: None reported.

1. Leichsenring F, Rabung S. Effectiveness of long-term psychodynamic psychotherapy: a meta-analysis. *JAMA*. 2008;300(13):1551-1565.
2. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med*. 1997;127(5):380-387.
3. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol*. 2001;54(10):1046-1055.

To the Editor: In their meta-analysis of the effectiveness of LTPP, Drs Leichsenring and Rabung[1] concluded that LTPP was significantly superior to shorter-term methods of psychotherapy with regard to overall outcome, target problems, and personality functioning. There are multiple problems with their meta-analysis that mitigate the scientific credibility of this conclusion.

First, the designation of "shorter-term methods of psychotherapy" included 5 treatments that did not constitute formal psychotherapy as it is generally understood. These treatments consisted of a waitlist control condition, nutritional counseling, standard psychiatric care, low-contact routine treatment, and treatment as usual (TAU) in the community. Such "mixing apples and oranges" is problematic in meta-analyses.

Second, the selection of studies appears biased in favor of LTPP. One study[2] supportive of LTPP had previously been classified in a meta-analysis by Leichsenring and Rabung[3] as involving *short-term* psychodynamic psychotherapy (STPP) but was reclassified in the present meta-analysis as involving *long-term* psychodynamic psychotherapy. Also, the exclusion of the study by Giesen-Bloo et al[4] that favored schema-focused therapy over LTPP appears arbitrary. The original article met the necessary criteria for inclusion since it defined patients in treatment for 3 years as completers and presented effect sizes on that basis.[4]

Third, the assertion of the superiority of LTPP was based on 8 studies of poor methodological quality. Generally, they failed to control for multiple confounding factors between treatment groups, such as therapist experience and competence and frequency of sessions. At least one comparison-control study did not involve random assignment,[5] and except for one study,[2] none reported data on treatment integrity. Consequently, it remains unclear whether the outcomes reported in these studies were due to the type of treatment or to the broad range of other intergroup differences.

Fourth, in comparisons between LTPP and 2 control therapies (STPP and family therapy), the effect sizes were actually larger for the control therapies. In 3 other comparisons, the results were mixed. Established therapies (cognitive therapy, dialectical behavior therapy [DBT], and dynamic supportive therapy) yielded larger pre-post effect sizes for some outcomes, while LTPP yielded larger effect sizes for others. An inference of superiority of LTPP over the established shorter-term therapies is unsupported by the data.

Aaron T. Beck, MD
abeck@mail.med.upenn.edu
Sunil S. Bhar, PhD
Department of Psychiatry
University of Pennsylvania
Philadelphia

Financial Disclosures: Dr Beck reported being president of the Beck Institute for Cognitive Therapy and Research, but he does not receive any financial compensation for this. Dr Bhar reported no disclosures.

1. Leichsenring F, Rabung S. Effectiveness of long-term psychodynamic psychotherapy: a meta-analysis. *JAMA*. 2008;300(13):1551-1565.
2. Svartberg M, Stiles TC, Seltzer MH. Randomized, controlled trial of the effectiveness of short-term dynamic psychotherapy and cognitive therapy for cluster C personality disorders. *Am J Psychiatry*. 2004;161(5):810-817.
3. Leichsenring F, Rabung S, Leibing E. The efficacy of short-term psychodynamic psychotherapy in specific psychiatric disorders: a meta-analysis. *Arch Gen Psychiatry*. 2004;61(12):1208-1216.
4. 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(6):649-658.
5. Korner A, Gerull F, Meares R, Stevenson J. Borderline personality disorder treated with the conversational model: a replication study. *Compr Psychiatry*. 2006; 47(5):406-411.

To the Editor: The meta-analysis by Drs Leichsenring and Rabung[1] concluded that LTPP is an effective treatment for complex mental disorders. However, there are some critical aspects of this research that were not considered.

First, 2 studies categorized as RCTs in the meta-analysis,[2,3] representing more than half of the total patient count in the RCT group, had no control condition other

than LTPP. These 2 studies were not designed to answer the question of whether LTTP is more effective than other (shorter) psychotherapeutic treatments; thus the studies provide no additional information compared with observational studies.

Second, TAU was included in the group of other psychotherapy methods, weakening the effect of these alternative methods. Of these alternative methods, a large number were based on psychodynamic theories.

Third, the studies of treatment of borderline personality disorder indicate that manualized LTPP is superior to TAU. However, in the active treatment comparison, a study with unfavorable outcome for LTPP[4] was excluded. In that study,[4] all treatments lasted 3 times longer and had a larger sample size compared with the only other comparison study.[5]

Fourth, the numbers of patients reported for the primary study by Clarkin et al[5] differed from those in the original report (30 LTPP, 30 DBT, 30 dynamic supportive treatment in the original study vs 30, 17, and 22, respectively, in the meta-analysis).

Stefan Roepke, MD
stefan.roepke@charite.de
Department of Psychiatry and Psychotherapy
Charité-University Medicine Berlin
Berlin, Germany

Babette Renneberg, PhD
Free University Berlin
Berlin

Financial Disclosures: None reported.

1. Leichsenring F, Rabung S. Effectiveness of long-term psychodynamic psychotherapy: a meta-analysis. *JAMA*. 2008;300(13):1551-1565.
2. Høglend P, Amlo S, Marble A, et al. Analysis of the patient-therapist relationship in dynamic psychotherapy: an experimental study of transference interpretations. *Am J Psychiatry*. 2006;163(10):1739-1746.
3. Vinnars B, Barber JP, Noren K, Gallop R, Weinryb RM. Manualized supportive-expressive psychotherapy versus nonmanualized community-delivered psychodynamic therapy for patients with personality disorders: bridging efficacy and effectiveness. *Am J Psychiatry*. 2005;162(10):1933-1940.
4. 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(6):649-658.
5. Clarkin JF, Levy KN, Lenzenweger MF, Kernberg OF. Evaluating three treatments for borderline personality disorder: a multiwave study. *Am J Psychiatry*. 2007;164(6):922-928.

In Reply: In response to Dr Thombs and colleagues, because the number of controlled trials of LTPP was relatively small, we assessed within-group effect sizes in all studies. In the controlled studies, we found significantly larger within-group effect sizes in the LTPP conditions than in the control conditions. To assess the extent of this difference, we transformed the point biserial correlations used for tests of significance into between-group effect sizes in the form of a d statistic.[1]

These between-group effect sizes are not identical to between-group effect sizes as commonly assessed by calculating the difference between 2 treatments for each study. The between-group effect size we assessed may be larger than the underlying within-group effect sizes because they refer to different units. Our comparison of LTPP with other treatments considers treatment group rather than patient as the unit of analysis. A between-group effect size of 1.8 indicates that in the studies of LTPP the mean within-group effect size differed from the mean within-group effect size of the control groups by 1.8 standard deviations. Our article should have stated that patients in LTPP treatment groups with responses at or above the mean were better off (showed higher treatment effects) than the mean patient in 96% of the comparison groups (rather than "96% of the patients in the comparison groups").

To report the corresponding between-group effect size assessed in the conventional way, we calculated the difference between 2 competing treatments for each study. For overall outcome, this between-group effect size (Hedges d) equals 0.65 ($P = .03$), implying that the mean patient treated by LTPP was better off than 75% of the patients in the control groups.

In response to Dr Kriston and colleagues, considering treatment groups rather than studies as the unit of analysis can indeed reduce the effect of randomization. It may weaken internal validity but does not necessarily imply serious bias. Observational studies may not systematically overestimate the effects of psychotherapy.[2] The between-group effect size of 0.65 is not affected by this issue.

A number of writers criticize the heterogeneity of treatments, patients, measures, outcomes, and control conditions in the included studies. However, the results were robust across diagnostic groups and research design. Moreover, we studied 5 specific outcome domains to reduce heterogeneity. Among the controlled studies of LTPP, there was no significant heterogeneity for pre-post effects.

Regarding possible publication or study selection bias, we applied several methods to test for publication bias but did not find any indication. Kriston et al suspect a positive association between effect size and meta-analytical study weight based on the forest plot. However, the size of the squares represents the product of study weight and effect size. Study weight and effect sizes were not significantly correlated. Using sample sizes vs standard error to test for publication bias leads to different results only if 2 treatments are compared and if the sample sizes are substantially different.[3]

In response to Drs Beck and Bhar, we did not include waitlist groups in the control conditions. The control conditions did include several TAU conditions, reducing the mean effect size of the alternative treatments. However, they also included long-term psychotherapy (eg, DBT), which would increase the mean effect of the alternative treatments. Because the average duration of therapy was longer in the LTPP conditions, we used the alternative treatments as an unspecific (mixed) control group including TAU and different established treatments. Consequently, we did not claim that LTPP is superior to any specific forms of established psychotherapy (eg, DBT) in

complex mental disorders, but rather to predominantly shorter forms of psychotherapeutic interventions in general. Our comparison of LTPP with other treatments controlled for the common factors of psychotherapy (eg, attention, expectation for improvement, understanding). Thus, it is "specific" as defined by Chambless and Hollon,[4] allowing us to conclude that the superiority of LTPP is due to specific factors of LTPP.

The study by Giesen-Bloo et al did not meet inclusion criteria because the majority of patients were still in treatment (precluding pre-post or follow-up assessments). The psychodynamic treatment studied by Svartberg et al met our definition of LTPP because treatment duration was at least 1 year, although labeled as short-term. Beck and Bhar criticize the methodological quality of the controlled studies. However, all of the controlled studies of complex mental disorders used treatment manuals and ensured treatment integrity by supervision, video recordings of sessions, and ratings of adherence and competence.

Beck and Bhar list selective results of individual studies in which the effect sizes of the control groups were larger than those of LTPP or in which LTPP yielded larger effect sizes for some measures while the control treatments yielded larger effect sizes for others. However, the role of meta-analysis is to synthesize often disparate results across individual studies to arrive at more general conclusions.

In response to Drs Roepke and Renneberg, in the RCTs by Høglend et al and Vinnars et al different forms of LTPP were compared. We therefore did not include these studies in the comparisons of LTPP with alternative treatments but treated them as observational studies. For the study by Clarkin et al, we listed the smaller sample sizes for DBT and supportive therapy because these were the subsamples on which the authors based their results. For LTPP, this number of patients was 23, but we listed the intention-to-treat sample size. Wherever possible, calculations of effect sizes were based on intention-to-treat samples for all treatment conditions.

Falk Leichsenring, DSc
falk.leichsenring@psycho.med.uni-giessen.de
Department of Psychosomatics and Psychotherapy
University of Giessen
Giessen, Germany

Sven Rabung, PhD
Department of Medical Psychology
University Medical Center Hamburg-Eppendorf
Hamburg, Germany

Financial Disclosures: None reported.

1. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum; 1988.
2. Shadish WR, Matt G, Navarro A, Phillips G. The effects of psychological therapies under clinically representative conditions: a meta-analysis. *Psychol Bull*. 2000; 126(4):512-529.
3. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. 1994;50(4):1088-1101.
4. Chambless DL, Hollon SD. Defining empirically supported therapies. *J Consult Clin Psychol*. 1998;66(1):7-18.

Ultrasound-Based Diagnostic Strategies for Deep Vein Thrombosis

To the Editor: In their randomized controlled trial, Dr Bernardi and colleagues[1] found equivalent safety for 2 ultrasound-based diagnostic strategies for managing outpatients with suspected deep vein thrombosis (DVT). The results of the study deserve some comments.

First, for the 2-point strategy the prevalence of proximal DVT detection on the second ultrasound test was 5.5% (14/256). This is much higher than has been previously reported. For example, in the study by Cogo et al,[2] which included more than 1700 patients with suspected DVT, this rate was less than 1%.[2] It has been suggested that the second ultrasound test is of limited added value, and strategies have been proposed using a single proximal venous examination, which proved to have an acceptable 3-month thromboembolic rate.[3] However, the use of a single proximal vein ultrasonography might be criticized if the prevalence of proximal DVT detected at the second ultrasound test is as high as 5.5%. The high prevalence of cancer in this study (30%) may account for the high rate of proximal extension, indicating a need for particular care when choosing a diagnostic strategy for cancer patients.

Second, the study by Bernardi et al might help estimate the frequency of early proximal extension of distal DVT, which differs considerably among studies (from 0%-40%[4]). With the 2-point strategy, 14 early proximal extensions were identified on the 7-day follow-up. Assuming that the 2 groups were balanced by the randomization, the frequency of early extension of distal DVT might be as high as 22% (14/65). Although speculative, this estimate would support anticoagulant treatment for distal DVT, a highly contentious subject that should be addressed in future studies.

Isabelle Quere, MD
isabelle.queree@wanadoo.fr
Montpellier University Hospital
Montpellier, France

Marc Righini, MD
Geneva University Hospital
Geneva, Switzerland

Financial Disclosures: None reported.

1. Bernardi E, Camporese G, Buller HR, et al; Erasmus Study Group. Serial 2-point ultrasonography plus D-dimer vs whole-leg color-coded Doppler ultrasonography for diagnosing suspected symptomatic deep vein thrombosis: a randomized controlled trial. *JAMA*. 2008;300(14):1653-1659.
2. Cogo A, Lensing AW, Koopman MM, et al. Compression ultrasonography for diagnostic management of patients with clinically suspected deep vein thrombosis: prospective cohort study. *BMJ*. 1998;316(7124):17-20.
3. Perrier A, Desmarais S, Miron MJ, et al. Non-invasive diagnosis of venous thromboembolism in outpatients. *Lancet*. 1999;353(9148):190-195.
4. Righini M, Paris S, Le Gal G, Laroche JP, Perrier A, Bounameaux H. Clinical relevance of distal deep vein thrombosis: review of literature data. *Thromb Haemost*. 2006;95(1):56-64.

To the Editor: Dr Bernardi and colleagues[1] compared the safety of excluding DVT by serial 2-point ultrasonography to whole-leg color-coded Doppler ultrasonography. Of par-