

This article was downloaded by: [Society for Psychotherapy Research (SPR)]

On: 21 January 2011

Access details: Access Details: [subscription number 762317397]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Psychotherapy Research

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713663589>

### The effectiveness of long-term psychotherapy: Methodological research issues

Saskia de Maat<sup>a</sup>; Jack Dekker<sup>b</sup>; Robert Schoevers<sup>a</sup>; Frans de Jonghe<sup>c</sup>

<sup>a</sup> Mentrum Institute for Mental Health Amsterdam, The Netherlands <sup>b</sup> Mentrum Institute for Mental Health Amsterdam, and Department of Clinical Psychology, Free University of Amsterdam, The Netherlands <sup>c</sup> Mentrum Institute for Mental Health Amsterdam, and Department of Psychiatry, University of Amsterdam, The Netherlands

**To cite this Article** de Maat, Saskia , Dekker, Jack , Schoevers, Robert and de Jonghe, Frans(2007) 'The effectiveness of long-term psychotherapy: Methodological research issues', *Psychotherapy Research*, 17: 1, 59 — 65

**To link to this Article:** DOI: 10.1080/10503300600607605

**URL:** <http://dx.doi.org/10.1080/10503300600607605>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## The effectiveness of long-term psychotherapy: Methodological research issues

SASKIA DE MAAT<sup>1</sup>, JACK DEKKER<sup>2</sup>, ROBERT SCHOEVERS<sup>1</sup>, & FRANS DE JONGHE<sup>3</sup>

<sup>1</sup>*Mentrum Institute for Mental Health Amsterdam*, <sup>2</sup>*Mentrum Institute for Mental Health Amsterdam, and Department of Clinical Psychology, Free University of Amsterdam*, <sup>3</sup>*Mentrum Institute for Mental Health Amsterdam, and Department of Psychiatry, University of Amsterdam, The Netherlands*

(Received 2 May 2005; revised 12 October 2005; accepted 1 November 2005)

### Abstract

In evidence-based medicine (EBM) hierarchy, randomized controlled trials (RCTs) are ranked higher than cohort studies. However, cohort intervention studies are frequently, and RCTs rarely, used to investigate long-term psychotherapy (LTP). The authors compare the two methods and provide critical discussion of their acceptability, feasibility, and decisive power in LTP. The only essential and unchangeable difference between RCTs and cohort studies is that the former always include randomized control groups and the latter never do, giving RCTs a head start on internal validity that cohort studies cannot match. However, randomization nearly always has dramatic consequences for LTP research: The control conditions that are most informative (no treatment, wait list, placebo) are so unacceptable for the patients that decisive RCTs are, in most cases, unfeasible, but more feasible RCTs are less decisive. In contrast, the decisive power of cohort studies is determined by their methodological quality and knowledge of the natural course of the investigated disorders. Cohort studies are as capable as RCTs of meeting all quality criteria for intervention research, except for randomization. The knowledge of the natural course of the disorders suitable for LTP treatment is limited but not nonexistent. In most cases of LTP research, decisive RCTs present insurmountable, method-inherent feasibility problems and represent not the highest but rather an irrelevant level of evidence. The authors conclude that cohort studies provide the best available evidence.

Psychotherapy is not an applied science but rather a scientifically based art and skill. The general criteria of science, among others, apply to the field. The American Psychological Association (Chambless & Hollon, 1998; Task Force on Promotion and Dissemination of Psychological Procedures, 1995) has defined quality criteria to assess whether or not the effectiveness of a psychotherapeutic method has been sufficiently substantiated empirically to deserve the qualification. Empirically supported psychological therapy (ESPT; Kendall, 1998) meets the standards of evidence-based medicine (EBM): “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine requires the integration of individual, clinical expertise with the best available, external, clinical evidence from systematic research and our patients’ unique values and circumstances” (Centre for Evidence-Based Medicine, n.d.).

Researchers in the field of treatment effectiveness (outcome research, not process research) agree that

empirical evidence can and must be ordered in a hierarchical system. A widely accepted hierarchy (Centre for Evidence-Based Medicine, 2001) is as follows:

- 1a. A systematic review of RCTs with consistent results
- 1b. One high-quality RCT
- 1c. One all-or-none study
- 2a. A systematic review of cohort studies or of patient–control studies with consistent results
- 2b. One cohort- or patient-control study or one lower quality RCT
- 2c. Outcome research
- 3a. A systematic review of case–control studies
- 3b. A case–control study
4. Case series or a lower quality cohort study
5. Expert’s opinions or generally accepted therapeutic methods

RCTs, or their systematic reviews and meta-analyses, occupy first place in this hierarchy and cohort studies second. A cohort is a well-defined

group of people who are monitored over a specific period of time to observe whether a certain outcome occurs. The relation between a determinant (in our case long-term psychotherapy [LTP]) and outcome (the effect) is scrutinized. If the determinant is an intervention, this design is also called a quasi-experimental study: “experiments that lack random assignment...but that otherwise have similar purposes and structural attributes to randomized experiments” (Shadish, Cook, & Campbell, 2002). Quasi-experimental studies may or may not include a comparison group. In this study, we apply the term *cohort study* because we believe it is most frequently used in EBM literature.

Psychotherapy has been extensively investigated with RCTs (see, e.g., the systematic reviews of Crits-Christoph, 1992; Leichsenring & Leibling, 2003; Leichsenring, Rabung, & Leibling, 2004; Perry, Banon, & Ianni, 1999). Together, these four reviews refer to 28 different RCTs, 27 of them short-term psychotherapy (STP) and only 1 of them LTP. RCTs have frequently studied short-term therapies and rarely long-term therapies in general and especially in psychotherapy. This state of affairs has led to the common opinion that LTP is not an ESPT. Its effectiveness would not have been researched scientifically, let alone convincingly demonstrated. This conclusion is based on two premises: (a) that conclusive RCTs are feasible in LTP research and (b) that RCTs are superior to cohort studies in all important methodological aspects. We discuss these views and assess their validity. We apply an arbitrary distinction between STP and LTP: The latter consists of at least 50 sessions and lasts at least 1 year.

## RCTs and LTP

### Randomization of RCTs

RCTs derive their name from their most important characteristic: the randomization of patients. This study design is also called the confirmatory-deductive methodology and is applied in efficacy research. RCTs strive for a maximum internal validity, so that it can be assumed, beyond reasonable doubt, that differences found between treatment and control groups (the dependent variable) are explained by the therapeutic intervention (the independent variable). Within reasonable limits, this method ensures like no other that chance is the only alternative explanation for the differences found between the studied groups. Factors possibly (co)determining results (confounders) are neutralized as much as possible. To this end, the groups have to be, apart from the intervention concerned, comparable to each other

(e.g., regarding demographical and clinical characteristics). The tool to achieve this is random assignment. This means patients are allocated by chance to either the experimental treatment or the control condition. As Kunz and Oxman (1998) state, “It is a paradox that unpredictability is introduced into the design of clinical trials by using random allocation to protect against the unpredictability of the extent of bias in the results of non-randomized clinical trials.”

The advantages of randomization are obvious and justify the high ranking of RCTs within the hierarchy of empirical evidence. However, this does not diminish the fact that, especially over the past decade, RCTs have been severely criticized as well, not only in the field of psychotherapy and regardless of treatment length (Crits-Christoph, 1997; Healy, 2001; Hollon, 1996; Howard, Moras, Brill, Martynovich, & Lutz, 1996; Kaptchuck, 2001; Leichsenring, 2004; Marks, 1997; Seligman, 1995). The criticism focuses primarily on the external validity of RCTs (i.e., on the generalizability of their results to daily practice). Many patients refuse participation, are excluded, or drop out. The manuals and protocols applied, the many assessments, and so on connect poorly to daily patient care. Therefore, the question arises whether conclusions drawn from such research are also valid in real-world applications. In short, RCTs opt for internal validity at the expense of external validity. To offer a synthesis in this discussion, Leichsenring (2004) has argued that the RCT and the quasi-experimental study do not differ in principle concerning their external and internal validity, given their own specific context (the laboratory and the field, respectively).

In this article, we focus on the consequences of randomization for the feasibility and conclusive power of RCTs in LTP investigations because we are convinced that this is even more of a problem for long-term treatments than for short-term treatments.

### Acceptability, Feasibility, and Conclusive Power of RCTs in LTP Research

Based on chance, RCTs compare a treatment to be examined (the experimental one, in this case LTP) with a control condition. The following seven alternatives qualify as control conditions: no treatment, wait list, placebo treatment, treatment as usual (TAU), the experimental LTP in a low dose, another LTP with an assumed but unproven efficacy, and another LTP with established efficacy.

*No treatment, wait list, placebo treatment.* RCTs comprising one of the first three control conditions (no treatment, wait list, or placebo) provide the

strongest proof of efficacy. As far as LTP is concerned, however, they merely are theoretical options, because few well-informed patients will participate in a study lasting at least 1 year and offering a 50% chance on no treatment, wait list, or treatment only resembling psychotherapy. Should anyone accept the chance on such control conditions and indeed be allocated to one, it can be safely assumed that he or she will seek treatment elsewhere in the course of the year. If, by chance, that patient would be compliant with the control condition, he or she is most certainly not representative of the study population. In short, these RCTs are almost unfeasible when they are conclusive, and they are inconclusive when they are feasible. It is a small wonder they are not found in the literature.

*TAU.* RCTs can compare LTP with TAU. (See, e.g., Bateman & Fonagy, 1999; Linehan, Armstrong, Suarez, Allmon, & Heard, 1991; Linehan, Dimeff, Reynolds, Comtois, Welch, & Kivlahan, 2002; Linehan, Schmidt, Dimeff, Craft, Kanter, & Comtois, 1999; Verheul, van den Bosch, Koeter, de Ridder, Stijnen, & van der Brink, 2003; all studied patients with borderline personality disorder [BPD]). This design is feasible because there is an acceptable TAU for BPD patients, namely treatment at a psychiatric outpatient clinic. Bateman and Fonagy (2000) clearly favor this study type: "Waiting-list controls cannot be used to control for change over long periods of time, and so are of limited use. The most stringent control group without ethical problems is TAU which should be used in the future, even though heterogeneity of interventions and differential responsiveness within groups may obscure results." To our knowledge, no RCTs compare LTP and TAU in non-BPD-patients, not because there is no LTP for these patients but because there is no acceptable TAU for well-informed patients. Without exaggeration, it may be said that TAU mostly consists of minimal therapy, hardly rising above the first three options mentioned previously. It has been said that TAU comes down to cheap, low-frequency treatment provided by minimally trained paraprofessionals with an overwhelming caseload struggling to get by (Scheel, 2000). What patient would be prepared to chance such control condition in a long-term study? Borderline patients do; what else could they do? In short, regarding LTP, the feasibility of RCTs comprising TAU as control condition is clearly limited.

*Experimental LTP in a low dose.* In this design, LTP in a dose considered normal is compared with LTP in a lower dose. In pharmacological studies, this generally means a dose considered ineffective, in fact

a placebo in disguise, unless several doses are tested in search of a dose-response curve. In the context of psychotherapy, the term *dose* refers to session frequency. This control condition is possibly more acceptable to patients, but a problem arises with the possible outcome of this research type. Only an apparent difference in effectiveness is conclusive; finding no difference may as well signify that the two treatments are efficacious as that they are both inactive. To correctly interpret a negative result (finding no difference between the two dosages), it is necessary to use the natural course of the disorder (the course of the disorder untreated) as a reference. If the natural course is known, however, cohort studies would suffice.

There is yet another drawback to comparing two psychotherapy doses. A large dose difference (e.g., two sessions weekly vs. one every 2 weeks) increases the probability of finding a difference in effect, thus enhancing the conclusive power of the study. However, it decreases the acceptability of the control condition to patients, thus reducing the feasibility of the study. In contrast, a small dose difference (e.g., four vs. three sessions weekly) increases the acceptability, and thus the feasibility, of the study but decreases the probability of finding a difference in effect and thus the conclusive power of the study. This RCT type is either conclusive but hardly feasible or highly feasible but poorly conclusive. It seems that the conclusive power and the feasibility of RCTs are two enemies convicted to one another.

*Another LTP with an assumed but unproven efficacy.* This study type compares two LTPs, the efficacy of which is assumed although it has not been proven in conclusive RCTs. An RCT comparing schema therapy (Young, Klosko, & Weishaar, 2003) with transference-focused psychotherapy (Kernberg, Clarkin, & Yeomans, 2002) is an example. Which of the two treatments represents the control condition depends on the researchers' perspective. The acceptability and the feasibility of this design are fair. However, here too only finding a difference in effect is conclusive (unless the natural course of the disorder is known, which would render the RCT superfluous). When researchers sincerely believe that two treatment methods are roughly equivalent, the probability of finding a difference is low. In short, in most cases, this study design is fairly feasible but poorly conclusive.

*Another LTP with established efficacy.* This design compares LTP with another LTP with undisputed efficacy. Dialectic behavior therapy (Linehan et al., 1991, 1999, 2002; Verheul et al., 2003) and mentalization-based therapy (Bateman & Fonagy, 1999) in

the treatment of BPD patients qualify as such control conditions. Acceptability and thus feasibility of this RCT type are passable. However, two new problems arise. The first involves the conclusive power of this research type. Finding a difference in favor of the experimental treatment would naturally be instructive, but the chances are slim. This study design is primarily applied when the aim is to find no difference in order to conclude that the new method is equal in efficacy to the established one (the “me too” method). No difference may be found, but that does not necessarily mean that the new method differs as clearly as the established one from a control condition that implies no treatment, a wait list, or a placebo. The second and most prominent problem is that there are no LTPs with RCT-proven efficacy for non-BPD patients.

### Cohort Studies and LTP

#### Internal and External Validity of Cohort Studies

When RCTs are not feasible, what is the highest level of empirical evidence? According to the Centre for Evidence-Based Medicine (2001) hierarchy of evidence, the answer must be cohort studies. Compared with RCTs, cohort studies have a lower internal validity. However, the external validity is higher because they present less selection bias. In addition, in most designs cohort studies investigate treatment methods conforming to norms that hold in daily practice, whereas RCTs do not.

#### Acceptability, Feasibility, and Conclusive Power of Cohort Studies in LTP

In cohort studies, only the acceptability of the treatment condition counts, not that of a control condition as well. Therefore, the feasibility of cohort studies is, especially in LTP, better than that of RCTs.

The conclusive power of cohort studies is still a matter of debate. Many believe that cohort studies tend to overestimate the effects of treatments, because there is no correction for potential confounders, meaning that the comparability of the treated and the untreated groups may be doubted (Sacks, Chalmers, & Schmidt, 1982; Kunz & Oxman, 1998). Theoretically, this is certainly possible, but whether overestimation does actually occur in practice remains to be seen. An exhaustive and impressive study, published in the *New England Journal of Medicine* (Concato, Shah, & Horwitz, 2000), compared the results of RCTs with those of cohort studies using the same interventions in five different somatic disorders. Review of 72 meta-analyses of RCTs, 24 meta-analyses of cohort

studies or case-control studies, and 6 meta-analyses involving both designs demonstrated that the results of the observational studies remarkably resembled those of the RCTs. Concato et al. concluded that “the popular belief that only randomized, controlled trials produce trust-worthy results and that all observational studies are misleading does a disservice to patient care, clinical investigation, and the education of health care professionals.” It may be clear that these findings are important in light of the discussion on research into the effectiveness of LTP.

The conclusive power of cohort studies depends on the methodological quality of studies themselves and on the knowledge of the natural course of the disorders studied.

*Methodological quality of cohort studies.* The actual structure of cohort studies is often looser than that of RCTs, resulting in the general misconception that cohort studies are of lower quality than RCTs. Except for randomization, the assessment criteria for the methodological quality of intervention cohort studies do not differ from those for RCTs. Both research types can be of superior or inferior quality. It is in this light that Leichsenring (2004) has proposed a hierarchy of evidence for both RCTs and quasi-experimental studies based on criteria that do not represent a sine qua non condition for high-level research. As a rule, they are not categorical (yes-no) but dimensional (studies meet them from a slight to considerable extent) and include, but are not limited to, prospective study design; a control group, preferably by means of extensive matching; specific, homogeneous, and representative study groups; clearly defined, specific, and representative treatments of adequate duration; follow-ups of adequate duration; pre-, post-, and follow-up assessments; reliable and valid instruments applied by independent assessors; adequate statistical processing; acceptable dropout percentage during treatment and follow-up; intention-to-treat and per protocol approaches and dropout accountability.

Cohort studies are prospective or retrospective. They can, but do not necessarily have to, include a control group (of course, not randomized; Kazin, 1998; Shadish et al., 2002). Single-group cohort designs monitor one study group. There is no control group. Pre- and postmeasuring makes each patient functioning as his or her own control. Multigroup cohort designs monitor at least two groups: One is exposed to the intervention (e.g., psychoanalysis) and the other is often an otherwise treated control group (e.g., psychoanalytic psychotherapy). If the groups are matched on relevant variables, the results of both groups can be cautiously compared. Matching involves variables (e.g.,

demographical and diagnostic characteristics), which might contribute to the differences in effect. Matching is not an all-or-none phenomenon. The smaller the number of variables included in the matching, the less the control group deserves its name. Epidemiological data regarding the general population may also serve as a control for the results of cohort studies. Untreated groups in epidemiological research (which reflect the natural course of a condition) can be compared with treated cohorts. Again, the more the persons from epidemiological studies are matched to the cohort, the stronger is the evidence.

*Natural courses of the disorders treated with LTP.* Knowledge of the natural course of the disorder enhances the conclusive power of cohort studies considerably. The natural course is the spontaneous remission rate of the disorder under study among untreated patients. Although the efficacy of the surgical removal of an inflamed appendix has not been investigated in any RCT, nobody considers this a problem. People are convinced they know sufficiently well the natural course of appendicitis to consider RCTs in this case superfluous and on ethical grounds unacceptable.

Knowledge of the natural course of disorders can be based on clinical experience (as is the case with appendicitis), but results from epidemiological research are preferable. In practice, it proves difficult to determine the real (because untreated) natural course of disorders treated with LTP. A number of studies may serve as an example. Spijker et al. (2000) studied the course of major depression (MDD) in The Netherlands. It appeared that 50% of all new cases recover within the first 3 months. After that period the probability of remission diminishes and becomes nil after 12 months. After 2 years 26% are still not remitted. Two studies suggest they provide relevant information on the natural course of BPD. Perry et al. (1999) concluded, based on a survey of five studies, that 2 and 10 years after treatment termination, respectively, 19% and 48% of BPD patients have recovered, for a remission rate of 3.6% per year between the 2nd and 10th years. Zanarini, Frankenburg, Hennen, and Silk (2003) concluded, based on one study ( $N=290$ ), that 2 and 6 years after treatment termination, respectively, 35% and 66% have recovered, for a remission rate of 7.7% per year between the 2nd and 6th years. Based on these results, the remission percentage of BPD could roughly be estimated at 6% a year. In this context, however, the term *recovered* only means that the patient no longer meets the *Diagnostic and Statistical Manual of Mental Disorders* criteria for BPD. Unfortunately, the three

studies mentioned previously relate to the course, not to the natural course, of the disorder. Perry et al. and Zanarini et al. relate to the course of BPS after termination of intensive treatment. In addition, most patients in these three studies had sought and received help during the follow-up period. It may be assumed that the natural course of disorder is even less favorable than presented (unless one assumes that treatment is to no avail or even worsens the outcome).

In short, knowledge of the natural course of disorders is scarce, but epidemiological research of untreated disorders is necessary to enhance the conclusive power of any long-term treatment study.

## Discussion

This article discusses two premises that are more or less explicitly taken for granted in research regarding the efficacy of LTP: (a) Conclusive RCTs are feasible in LTP research and (b) RCTs are superior to cohort studies in all important methodological aspects.

### Feasibility of Conclusive RCTs

In RCTs randomization warrants unsurpassed internal validity. However, the treatment condition as well as the control condition must be acceptable. Otherwise, the study is impracticable. Acceptability of the control condition is often difficult enough in STP studies. In LTP studies the problem, as a rule, is almost insurmountable. The consequences are considerable. First, the more limited the acceptability of the control condition, the smaller is the percentage of the study population that will agree to the study conditions and the larger the selection bias. However, our main argument involves another aspect. Conclusive control conditions (no treatment, wait list, placebo) are, as far as LTP is concerned, unacceptable. More acceptable control conditions (other treatment modalities), if they exist at all, are less conclusive. All too often, researchers have to choose between either hardly feasible designs with high conclusive powers or passably feasible designs with low conclusive power.

Thus far, the literature has paid insufficient attention to the limited acceptability of control conditions to patients and the subsequent problems for RCTs. There is extensive literature on patient dropout after treatment has started. There is also a good deal of literature on patients' refusal to start the treatment to which they have been randomized (the difference between intention-to-treat design and per protocol design). There are, however, almost no publications regarding patients' refusal to participate

in a study because they do not accept the possibility of being allotted to the control condition.

The problems outlined previously are limitations not of the treatment method but of the research method. The factors determining the limitations relate to the duration, not type, of treatment. They are by no means specific to LTP research but apply in general to research of long-term treatments, whether they are of a somatic or a psychological nature.

### Methodological Qualities of RCTs and Cohort Studies

General quality criteria apply to intervention research, some of which have been mentioned before. Neither RCTs nor cohort studies can meet all criteria because of, among other factors, randomization (at the expense of external validity) and lack of randomization (at the expense of internal validity). Apart from these inherent limitations, RCTs and cohort studies, in actual practice, vary in quality and are both able to meet all other criteria for intervention studies. We discussed two specific problems of cohort studies.

The inclusion of a control group enhances the credibility of cohort studies. The greater the number of variables included in the matching, the more decisive power the study has. However, the relevance of matching must not be overestimated. After all, it is never certain whether all relevant factors have been taken into account in the matching.

Furthermore, the conclusive power of cohort research is codetermined by the control knowledge of the natural (i.e., untreated) course of the disorders studied. Progress in epidemiological research is badly needed, but there are unresolved, possibly insoluble, problems. People who suffer understandably seek help, and in our society they often get it. Consequently, the natural, untreated course of many disorders suitable for LTP treatment is hardly known. There is still another problem if data regarding the general population are gathered. This group differs from the group treated in a cohort study in one essential aspect (possibly an important confounder): help-seeking behavior.

Where scientific arguments make default, clinical experience retains its rights. In other words, for lack of epidemiological knowledge regarding the natural course of the disorders suitable for LTP treatment, one has to accept the guidance of clinical experience. LTP seems mostly indicated for long-existing problems. Leichsenring and Leibling (2003) conclude that "several studies reported more improvement in personality disorders after longer treatment durations." Kopta, Howard, Lowry, and Beutler (1994)

also found that improvements in character problems take longer than symptom changes. In the context of stepped care, LTP is also indicated in case STP appears insufficiently effective. The latter is by no means exceptional; on the contrary, there are strong indications that the effects of short-term therapies are short lived. In an extensive review, Gloaguen, Cottraux, Cucherat, and Blackburn (1998) showed that relapse rates after short-term treatment for depression with psychotherapy and pharmacotherapy varied, on average, from 30% to 60%. Two time-honored adages apply here: (a) The longer the duration of disorders and the more frequently they have been treated to no avail, the less likely it is that they will remit spontaneously. Time rarely heals old and ineffectively treated wounds.

### Conclusion

RCTs constitute an inadequate research methodology for studying LTP effectiveness, except with BPD. RCTs represent not the highest but rather an irrelevant level of empirical evidence. Where RCTs are an impracticable and, therefore, inadequate research method, cohort studies provide the best available evidence.

### References

- Bateman, A., & Fonagy, P. (1999). The effectiveness of partial hospitalisation in the treatment of borderline personality disorder—A randomised controlled trial. *American Journal of Psychiatry*, 156, 1563–1569.
- Bateman, A. W., & Fonagy, P. (2000). Effectiveness of psychotherapeutic treatment of personality disorder. *British Journal of Psychiatry*, 177, 138–143.
- Centre for Evidence-Based Medicine Toronto. (n.d.). *Glossary of EBM terms*. Retrieved March 15, 2005, from <http://www.cebm.utoronto.ca/glossary/>
- Centre for Evidence-Based Medicine Oxford. (2001). *Levels of evidence*. Retrieved March 15, 2005, from [http://www.cebm.net/levels\\_of\\_evidence.asp](http://www.cebm.net/levels_of_evidence.asp)
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7–18.
- Concato, J., Shah, N., & Horwitz, R. I. (2000). Randomised controlled trials, observational studies and the hierarchy of research designs. *New England Journal of Medicine*, 342, 1887–1892.
- Crits-Christoph, P. (1992). The efficacy of brief dynamic psychotherapy: A meta-analysis. *American Journal of Psychiatry*, 149, 151–158.
- Crits-Christoph, P. (1997). Limitations of the "dodo bird" verdict and the role of clinical trials in psychotherapy research: Comments on Wampold et al. (1997). *Psychological Bulletin*, 122, 216–220.
- Gloaguen, V., Cottraux, J., Cucherat, M., & Blackburn, I. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders*, 49, 59–72.
- Healy, D. (2001). The dilemmas posed by new and fashionable treatments. *Advances in Psychiatric Treatment*, 7, 322–327.

- Hollon, S. (1996). The efficacy and effectiveness of psychotherapy relative to medications. *American Psychologist*, 51, 1025–1030.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist*, 51, 1059–1064.
- Kaptchuck, T. J. (2001). The double-blind randomised, placebo-controlled trial: Gold standard or golden calf? *Journal of Clinical Epidemiology*, 54, 541–549.
- Kazin, A. E. (1998). *Research design in clinical psychology*. Boston: Allyn & Bacon.
- Kendall, P. C. (1998). Empirically supported psychological therapies. *Journal of Consulting and Clinical Psychology*, 66, 3–6.
- Kernberg, O. F., Clarkin, J. F., & Yeomans, F. E. (2002). *A primer of transference focused psychotherapy for the borderline patient*. New York: Jason Aronson.
- Kopta, S. M., Howard, K. I., Lowry, J. L., & Beutler, L. E. (1994). Patterns of symptomatic recovery in psychotherapy. *Journal of Consulting and Clinical Psychology*, 62, 1009–1016.
- Kunz, R., & Oxman, A. D. (1998). The unpredictability paradox: Review of empirical comparisons of randomised and non-randomised clinical trials. *British Medical Journal*, 317, 1185–1190.
- Leichsenring, F. (2004). Randomized controlled versus naturalistic studies: A new research agenda. *Bulletin of the Menninger Clinic*, 68, 137–151.
- Leichsenring, F., & Leibling, E. (2003). The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders: A meta-analysis. *American Journal of Psychiatry*, 160, 1223–1232.
- Leichsenring, F., Rabung, S., & Leibling, E. (2004). The efficacy of short-term psychodynamic psychotherapy for specific psychiatric disorders: A meta-analysis. *Archives of General Psychiatry*, 61, 1208–1216.
- Linehan, M. M., Armstrong, H. E., Suarez, A., Allmon, D., & Heard, H. L. (1991). Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Archives of General Psychiatry*, 48, 1060–1064.
- Linehan, M. M., Dimeff, L. A., Reynolds, S. K., Comtois, K. A., Welch, S. S., & Kivlahan, D. R. (2002). Dialectic behavior therapy versus comprehensive validation therapy plus 12-step for the treatment of opioid dependent women meeting criteria for borderline personality disorder. *Drug Alcohol Dependence*, 67, 13–26.
- Linehan, M. M., Schmidt, H., Dimeff, L. A., Craft, J. C., Kanter, J., & Comtois, K. A. (1999). Dialectic behavior therapy for patients with borderline personality disorder and drug-dependence. *American Journal of Addiction*, 8, 279–292.
- Marks, H. M. (1997). *The progress of experiment. Science and therapeutic reform in the United States 1900–1990*. New York: Cambridge University Press.
- Perry, J. C., Banon, E., & Ianni, F. (1999). Effectiveness of psychotherapy for personality disorders. *American Journal of Psychiatry*, 156, 1312–1321.
- Sachs, H., Chalmers, T. C., & Smith, H., Jr. (1982). Randomized versus historical controls for clinical trials. *American Journal of Medicine*, 72, 233–240.
- Scheel, K. R. (2000). The empirical basis of dialectic behavior therapy: Summary, critique, and implications. *Clinical Psychology: Science and Practice*, 7, 68–86.
- Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist*, 50, 965–974.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Spijker, J., de Graaf, R., Bijl, R. V., Beekman, A. T., Ormel, J., & Nolen, W. A. (2000). Duration of major depressive episodes in the general population: Results from The Netherlands Mental Health Survey and Incidence Study. *British Journal of Psychiatry*, 181, 208–213.
- Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training and dissemination of empirically-validated psychological treatments. *Behavioral Research & Therapy*, 39, 299–308.
- Verheul, R., van den Bosch, L. M. C., Koeter, M. W. J., de Ridder, M. A. J., Stijnen, T., & van der Brink, W. (2003). Dialectical behaviour therapy for women with borderline personality disorder. *British Journal of Psychiatry*, 182, 135–140.
- Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema therapy. A practitioner's guide*. New York: Guilford Press.
- Zanarini, M. C., Frankenburg, F. R., Hennen, J., & Silk, K. R. (2003). The longitudinal course of borderline psychopathology. *American Journal of Psychiatry*, 160, 274–283.