Psychotherapy Research, October 2006; 16(5): 604–616

Routledge
Taylor & Francis Group

Change norms: A complementary approach to the issue of control groups in psychotherapy outcome research

FALK LEICHSENRING[1] & SVEN RABUNG[2]

[1]*University of Göttingen and Clinic of Tiefenbruan, Germany and* [2]*University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

(Received 22 December 2004; revised 16 June 2005; accepted 13 September 2005)

Abstract
Control groups are widely regarded as a sine qua non of psychotherapy outcome research. However, they produce high costs and are not applicable to long-term studies. The authors address the issue of control groups in psychotherapy research in a novel way: They systematically study the changes occurring in psychiatric patients who did not receive any specific treatment. Such data may serve as reference data for further outcome studies. It would then no longer be necessary to collect reference data repeatedly in each and every new study. The authors evaluated the changes that occurred in the control groups of available randomized controlled studies (RCTs) of psychodynamic psychotherapy in a first and preliminary trial. Twenty-six RCTs of psychodynamic psychotherapy, in which reliable and valid outcome measures were applied and which provided the necessary data to calculate effect sizes, were included in a meta-analysis. Effect sizes were calculated for the mean changes occurring in the control groups. The mean effect size of 0.12 (95% confidence interval [CI]: 0.05–0.28) corresponds to a small effect. To demonstrate its possible use, the upper limit of the 95% CI ($d = 0.28$), considered a very conservative estimate of the average control group effect size, was compared with that of selected studies of psychodynamic psychotherapy. The average change in psychiatric patients included in RCTs of psychotherapy who did not receive specific treatment may serve as a preliminary control group estimate (change norm) for open trials of psychotherapy. Future research should assess the mean expected change in more sophisticated controls (e.g., for specific psychiatric disorders, outcome measures, or patient variables).

Control groups are widely regarded as a sine qua non of psychotherapy outcome research. They can refer to either a no-treatment condition (e.g., a wait list) or an alternative treatment.[1] Comparison with a no-treatment condition is regarded as indispensable for the comparison of the observed effects with the spontaneous course of a disorder. We address the issue of control groups in psychotherapy research in a new way and propose a complementary approach.

The assumption of control groups as a sine qua non is reflected, for example, in the criteria of the American Psychological Association's (APA) Task Force on Promotion and Dissemination of Psychological Procedures of Division 12 (Clinical Psychology). According to these criteria, efficacy of psychotherapeutic methods can be demonstrated only by randomized controlled studies (RCTs), in which a therapy group is compared with a control condition (wait list, placebo group, alternative treatment) or an already established treatment (Chambless & Hollon, 1998; Chambless & Ollendick, 2001; Task Force on Promotion and Dissemination of Psychological Procedures, 1995). Also, in evidence-based medicine, RCTs are regarded as the gold standard for the demonstration of efficacy (Canadian Task Force on the Periodic Health Examination, 1979; Cook, Guyatt, Laupacis, Sacket, & Goldberg, 1995; Guyatt et al., 1995; Nathan & Gorman, 2002).

However, the exclusive position of RCTs as methods for demonstrating that a treatment works has been questioned (Beutler, 1998; Fonagy, 1999; Persons & Silberschatz, 1998; Roth & Parry, 1997; Seligman, 1995). Seligman (1995, p. 966) summarized his criticism of RCTs as follows: "The efficacy study is the wrong method for empirically validating psychotherapy as it is actually done, because it omits too many crucial elements of what is done in the field." Furthermore, it can be shown that in many RCTs randomization is only a formal criterion, and the intended control of confounding variables is probably not achieved: Hsu has demonstrated statistically that a large number of patients ($N = 40$ per group) is necessary to control effectively for a

Correspondence: Falk Leichsenring, Clinic of Psychosomatics and Psychotherapy, University of Goettingen, von-Siebold-Str. 5, 37075 Goettingen, Germany. E-mail: fleichs@gwgd.de

ISSN 1050-3307 print/ISSN 1468-4381 online
DOI: 10.1080/10503300600805217

greater number of confounding variables and to make randomization work well (Hsu, 1989). When samples are of a moderate size (20–40 per group), randomization works well only if the number of confounding variables is small (Hsu, 1989). In many RCTs, the actual number of patients falls below the necessary sample size; therefore, it is questionable whether even in RCTs all confounding variables were actually controlled for. Hsu (1989) cited the meta-analysis of Shapiro and Shapiro (1982), who found an average sample size of 11.98. Hsu (1989, p. 134) concluded that, in more than half of the comparative outcome studies examined by Shapiro and Shapiro, the samples were small enough to suggest a better than even chance of nonequivalence of the contrasted groups constructed by random assignment. Also, in more recent studies, the sample size is often not large enough to make randomization work well: For example, in Gloaguen, Cottraux, Cucherat, and Blackburn's (1998) meta-analysis of the effects of cognitive therapy in depression, only 38% of the groups had a sample size of at least 20 patients and only 13% had a size of at least 40 patients per group, thus fulfilling the aforementioned criteria by Hsu (1989). In addition, patients frequently drop out of control groups, additionally restricting the representativeness of the control groups for the therapy group even in RCTs.

Contrary to RCTs (efficacy studies), effectiveness studies are carried out under the conditions of clinical practice (Seligman, 1995). Thus, RCTs and effectiveness studies refer to different situational contexts or intended applications (Leichsenring, 2004): RCTs refer to laboratory conditions, whereas effectiveness studies refer to field conditions. For this reason, empirical evidence achieved under laboratory conditions cannot directly be transferred to field conditions and vice versa (Leichsenring, 2004). RCTs and effectiveness studies refer to different questions of research, and one type of study cannot be substituted by the other.

Per definition, effectiveness studies cannot use randomized assignment of patients to treatments; patients who are randomly assigned do not decide for a treatment or a therapist as they would in the field. Randomization is generally regarded as indispensable to ensure that the observed effects can only be attributed to the treatment applied (internal validity). The improvement of the internal validity of effectiveness studies is an important issue (Fonagy, 1999). Shadish, Cook, and Campbell (2002) have described experimental and quasi-experimental designs for generalized causal inference. According to Shadish et al. (2002) a causal inference from a quasi-experimental study must meet three basic requirements: Causes precede effects, causes covary with effects, and alternative explanations of the effects are implausible. Because quasi-experimental studies cannot use random assignment, they use other principles to show that alternative explanations of the effect are implausible. These principles include

1. Identification and study of plausible threats to internal validity
2. Use of additional design elements (e.g., observation at more pretest time points, additional comparison groups) or statistical controls
3. Coherent pattern matching, that is, prediction of complex patterns of results (e.g., of nonequivalent dependent variables or of interactions).

We would like to make a proposal referring to the use of additional design elements (i.e., for example, the use of additional comparison groups or of statistical controls, as mentioned under principle 2). Furthermore, in a more general sense, we address the issue of control groups in psychotherapy research in a new way: We propose to systematically study the changes that occur in patients with psychiatric disorders who did not receive psychotherapeutic or psychiatric treatment. This kind of research could provide an empirically assessed standard against which the results of open studies of psychotherapy research can be compared. These mean expected changes could be used as change norms, as discussed next.

For studies of the efficacy of psychotherapy, which are at present carried out in the form of RCTs, data about the mean expected changes in untreated patients could serve as substitutes for control groups. With reference data like this, it would probably no longer be necessary to collect such data repeatedly in each new study. Like norms of psychological tests, these data should be based on large and representative samples. Like data from normative samples, these mean expected changes should be specified for the relevant characteristics of the sample under study (e.g., age, gender, psychiatric disorders, or the acute–chronic dimension). Even cross-national and cross-ethnical differences between patient populations should be considered. If, for example, the mean expected changes in untreated patients vary across different psychiatric disorders, disorder-specific norms are required (e.g., separately for anxiety disorders, depressive disorders, or personality disorders). Like norms of psychological tests, these data have to be updated from time to time. An agreed-on standard battery of diagnostic instruments could be used to assess norms of change

(e.g., Strupp, Horowitz, & Lambert, 1997). Thus, our proposal implies a new program of research. Providing norms of change would save considerable costs as the expenses associated with establishing an untreated control group decrease. The necessary sample sizes are reduced, and the expense of screening procedures and diagnostic assessments can decrease considerably. Thus, they could serve as one of the just-mentioned additional design elements (comparison values, statistical controls) that can be applied to improve the internal validity of effectiveness studies. Admittedly, the degree of control of known and unknown factors influencing outcome that is achieved by a randomization, if correctly carried out, cannot be reached.[2]

Epidemiological studies are required to gain data of both short-term and long-term changes in untreated patients with psychiatric disorders. At present, data like this are not yet available. Therefore, in a first and preliminary attempt, we chose another approach to obtain such data of reference: to date in psychotherapy research, many RCTs have included control groups that provide data about the changes occurring in patients who did not receive any specific treatment. Thus, we decided to evaluate the changes that occurred in these control groups of available RCTs. Of course, assessing the changes in RCT control groups bears some important limitations, discussed later. We merely regard this approach as a first step and want to encourage more elaborate research to gain representative data of reference of changes in untreated patients.

Lambert and Bergin (1994) have demonstrated that the effects of psychotherapy exceed those of untreated groups and placebo treatments. Although the authors reported the difference in effect sizes between psychotherapy and untreated control groups or placebo groups, they regrettably did not report the pure effects of untreated control groups or of placebo groups in the form of an effect size.

In a meta-analytical approach, Grawe, Donati, and Bernauer (1994, p. 708) assessed the effects of control groups for a sample of 111 studies of psychotherapy. They found a global effect size for untreated control groups of 0.10. The authors did not report the standard deviation for those effect sizes; however, it can be assessed from a corresponding graph published by Grawe et al. (1994, p. 709). According to our calculation, the standard deviation was 0.33. Because this value is quite large compared with the mean, it indicates a wide variability of effect sizes in control groups across studies. One possible explanation for this result is that Grawe et al. (1994) did not differentiate between different disorders or methods of psychotherapy when evaluating the effect sizes of control groups. Although it may appear initially unusual, it may also be necessary to take into account the type of psychotherapeutic method against which the control groups were tested within an RCT: It cannot be excluded a priori that in the available RCTs of therapy A (e.g., psychodynamic psychotherapy) the patients systematically differ from those treated with therapy B (e.g., cognitive–behavioral therapy [CBT]) concerning relevant variables. For example, this could be the case if, before randomization, patients were selected with regard to criteria that are generally regarded as relevant for the indication of the method of therapy in question (e.g., patients sufficiently motivated and introspective to be successfully treated with short-term psychodynamic therapy). If these patient characteristics affect the rates of spontaneous remission, the changes in the control groups undergoing this form of therapy may be different from those in control groups undergoing other forms of therapy that do not use these selection criteria. For this reason, it seems to be necessary in a first step of research to assess the effects occurring in control groups separately for specific forms of psychotherapy. Of course, this hypothesis may be refuted by future empirical data. However, evaluating the changes occurring in control groups by type of therapy does not introduce a systematic bias. If future data show that type of therapy does not affect the effects of control groups, the data may be lumped together across type of therapy.

Most of the arguments put forward against the representativeness of RCTs for clinical practice (e.g., the use of specially trained therapists and of specific therapy manuals, an a priori fixed duration of therapy, random assignment of patients instead of patients making their own decision for a therapy and a therapist) refer to the therapy condition but not to the control condition of an RCT. For the control groups, the aforementioned factors are not relevant. Control groups offer the chance to study the changes that occur in patients who did not receive psychotherapy. However, there is a tendency to include only less severely disturbed patients with isolated disorders in RCTs (Roth & Parry, 1997). This tendency reduces the representativeness of the results for patients in clinical practice. Again, this problem primary refers to the therapy condition. The control condition should be less affected by this factor: It can be assumed that more severely disturbed patients, if anything, show less change without receiving psychotherapeutic treatment than less severely disturbed patients (see also the Discussion section). For this reason, the changes occurring in control groups of RCTs can be regarded to overestimate the changes occurring in clinical practice.

This should particularly be true in investigations of chronic disorders.

Using RCTs to investigate which changes occur in control groups has the great advantage in that randomization allows for the following conclusion to be made: If the effects of therapy A (e.g., psychodynamic psychotherapy) are studied in a randomized controlled group design, the effect sizes found in the control group are valid for exactly the same population of patients treated with therapy A (e.g., psychodynamic psychotherapy). As a result of randomization, the patients in the therapy group and those in the control group should be comparable with regard to known and unknown variables (e.g., psychiatric disorder, age, sex, personality variables). This is important with regard to the representativeness of the effect sizes occurring in control groups for the patients who were treated with therapy A.

In a preliminary move to gain first data of reference, this study assessed the effects of untreated control groups and of treatment as usual (TAU) groups from RCTs of psychodynamic psychotherapy. Thus, data of reference for a specific form of psychotherapy were assessed. For TAU groups, larger effect sizes than in wait list groups are expected. In this context, it is also possible to examine whether wait list groups and TAU groups differ concerning the occurring changes. Later in this article, we demonstrate how these expected values of change can be used. For this purpose, we compared the average control group effect we found in our meta-analysis with the effects of selected exemplary studies of psychoanalytic and psychodynamic psychotherapy.

Method

Selection of Studies

In addition to the usual search for studies in reviews, meta-analyses, and textbooks, a computerized search was carried out using Medline and PsycLit with the following key words: psychodynamic psychotherapy, psychoanalytic therapy, randomized controlled study.

Inclusion Criteria

We included those RCTs in which a form of psychoanalytic or psychodynamic psychotherapy was compared with a control condition, which could be a wait list group, a TAU group, a placebo therapy, or a group of untreated patients who were not offered psychotherapy after a waiting period. Only studies that used outcome measures for which reliability and validity had been demonstrated were included in the meta-analytic evaluation. Furthermore, studies had to provide data—means and standard deviations (pre and post)—that allowed for the calculation of effect sizes in the form of Cohen's *d* (Cohen, 1988). Studies that did not fulfill the aforementioned criteria were excluded. For several studies, it was not possible to calculate effect sizes because means or standard deviations were not reported (e.g., Guthrie, Creed, Dawson, & Tomenson, 1993; Sifneos, 1990).

Assessment of Effect Sizes

Effect sizes were assessed separately for each outcome measure used in a study. Effect sizes were calculated by dividing the difference between means before and after the waiting period (or TAU) by the standard deviation of the control group before the waiting period (or before TAU; Cohen, 1988). If necessary, signs were reversed so that a positive effect size always indicated improvement. Whenever multiple measures were applied in a study, we assessed the effect size for each measure separately and calculated the mean effect size to assess the overall effect of the study. For each of the 26 control groups, we assessed the magnitude of change in the form of Cohen's *d.* The effect size index gives the magnitude of change in units of standard deviations (e.g., $d = 1.5$ corresponds to a difference of 1.5 *SD*s).

Falk Leichsenring extracted the following information from the articles: (a) name of authors, (b) year of publication, (c) psychiatric disorder treated with psychodynamic or psychoanalytic therapy, (d) kind of control group, (e) sample size of control group, and (f) means and standard deviations for each outcome measure. This information was checked by Sven Rabung.

Areas of Functioning

We had planned to evaluate different areas of functioning (e.g., symptoms, interpersonal relations, well-being) or different psychiatric disorders separately, but it was not possible to do so because the number of studies providing the necessary data was not sufficiently large. Future studies should take the different areas of functioning and patient characteristics into account. Furthermore, because different outcome measures are differentially sensitive to change, future studies should assess change norms for specific outcome measures (e.g., Symptom Checklist-90-Revised [SCL-90-R]).

Results

Studies Included

Twenty-six studies fulfilled the criteria of inclusion (Table I).[3] For the studies included, the mean duration of the waiting period was 18.29 weeks

Table I. Studies of Psychoanalytic and Psychodynamic Therapy Including Control Groups.

Study	Control condition	Disorder	*N*
Dührssen & Jorswieck (1965)	TAU	Mixed	100
Levis & Carrera (1967)	Wait group	Mixed	10
Gillan & Rachman (1974)	TAU	Phobia	8
Newton & Stein (1974)	TAU	Alcoholism	29
Sloane, Staples, Cristol, & Yorkston (1975)	Wait group	Anxiety/personality disorders	33
Siegel, Rootes, & Traub (1977)	Wait group	Mixed	53
Meyer & Bolz (1981)	Wait group	Mixed	25
Rosser et al. (1983)	Wait group	Chronic bronchitis	17
Manos & Vasilopoulou (1984)	No treatment	Mixed	16
Deter (1986)	TAU	Asthma bronchiale	10
Thompson (Gallagher & Breckenridge, 1987)	Wait group	Depression	19
Marmar, Horowitz, Weiss, Wilner, & Kaltreider (1988)	Mutual help group	Conjugal bereavement	30
Brom, Kleber, & Defares (1989)	Wait group	Posttraumatic stress disorder	23
Snyder & Wills (1989)	Wait group + minimal treatment	Marital problems	20
Pilowsky & Barrow (1990)	Placebo tablet + supportive contact	Chronic pain	20
Woody, Luborsky, McLellan, & O'Brien (1990)	TAU	Opiate addiction	35
Winston et al. (1991)	Wait group	Personality disorders	26
Baldoni, Baldaro, & Trombini (1995)	TAU	Urethral syndrome (urinary symptoms and pain w/o organic lesions)	21
Shefler, Dasberg, & Ben-Shakhar (1995)	Wait group	Mixed	16
Woody, McLellan, Luborsky, & O'Brien (1995)	TAU	Opiate addiction	27
Bateman & Fonagy (1999)	TAU	Borderline personality disorder	19
Guthrie et al. (1999)	TAU	High users of psychiatric services	55
Hamilton et al. (2000)	TAU	Chronic dyspepsia	36
Monsen & Monsen (2000)	TAU+ no treatment (mixed)	Chronic pain (somatoform disorder)	20
Sandell et al. (1999, 2001)	Untreated	Mixed	12
Guthrie et al. (2001)	TAU	Deliberate self-poisoning	60

Note. TAU = treatment as usual.

($SD = 9.69$) and the mean duration of TAU was 28.07 weeks ($SD = 21.27$). Thus, control groups associated with a short- to medium-term waiting period or TAU were included in our meta-analysis. This result reflects the fact that, for both ethical and practical reasons, it is only possible to test short-term or medium-term therapies in RCTs, which include control conditions of equal duration, but not long-term psychotherapy or psychoanalysis (Sandell, Blomberg, & Lazar, 1999; Seligman, 1996).

Psychiatric Disorders Included

According to the diagnoses reported by the authors of the 26 studies, the patients included in these studies suffered from the following psychiatric disorders (see Table I): affective disorders (depression), anxiety disorders, personality disorders, borderline personality disorder, posttraumatic stress disorder, adjustment disorders, somatoform disorders (chronic pain, chronic dyspepsia), substance abuse (alcohol abuse, opiate addiction), high users of psychiatric services, patients who deliberately poisoned themselves, marital problems, and heterogeneous psychiatric disorders (seven studies; see Table I).

Thus, the patients included in the 26 studies cover most of the psychiatric disorders that are relevant for psychotherapy. The 26 studies make up a total sample of 740 patients. Thus, our calculation of effect sizes is based on a relatively large sample of patients who suffered from most of the psychiatric disorders relevant to psychoanalytic and psychodynamic psychotherapy, both of which speak in favor of the external validity of the data.

The mean sample size of the control groups was 28.46 ($SD = 19.85$). The size of a sample is important to ensure that randomization works well (Hsu, 1989). The 26 studies included both untreated control and TAU groups (see Table I).

Heterogeneity

Changes in control groups can be expected to be different for different disorders. Furthermore, they should be different for acute and chronic disorders. Therefore, we tested for heterogeneity of effect sizes. The value of one study (Sloane, Staples, Cristol, & Yorkston, 1975) was identified as an outlier ($d = 1.18$), which deviated considerably from the values of other studies (Norusis, 2000). Thus, the effect

sizes of the remaining studies can be regarded as not statistically significantly heterogeneous. To prevent a distortion of the estimate of the effect size in control groups by the outlier of this study, we substituted it by the mean effect size of the remaining 25 studies.

As a result of the relatively small number of controlled studies of psychodynamic psychotherapy, it was not possible to assess the effects of control groups both for a specific type of therapy and for specific disorders. Thus, our attempt can only be regarded as a first preliminary approach to draw attention to control group data. Future studies should take into account the type of disorder as well as other relevant patient characteristics (e.g., age, gender, or social variables).

Wait List Control Groups Versus TAU Groups

We initially tested whether the untreated wait list control groups differed from the TAU groups concerning the observed changes. Surprisingly, this was not the case: For the untreated wait list control groups, the effect size was 0.10 ($SD = 0.14$, $N = 9$); for the TAU groups, the effect size was 0.14 ($SD = 0.21$, $N = 15$). This difference is not significant, $t(21) = 0.42$, $p = .68$. In the remaining 2 of the 26 studies, the patients of the control groups were offered neither TAU nor any form of psychotherapy after a waiting period (Manos & Vasilopoulou, 1984; Sandell et al., 1999; Sandell, Blomberg, & Lazar, 2002). Because this is a third control condition, we did not include these two studies in the comparison of wait list to TAU because we wanted to evaluate the effect of waiting for psychotherapy in contrast to TAU. However, the average effect size for the entirely untreated control groups of these two studies was 0.08 and thus did not differ from the changes in wait list or TAU control groups. Because there were no significant differences between the different control conditions, we combined the data of the untreated groups and the TAU groups and calculated the mean effect size across all 26 studies. By aggregating all studies, the database is enlarged, and a higher representativeness can be achieved for the estimate of expected change in control groups of psychodynamic psychotherapy. The following results refer to the combined data of the wait list, TAU, and entirely untreated control groups. Aggregating data of TAU and untreated patients may be regarded as conceptually questionable. However, it depends on the respective question of research whether it makes sense to combine the data of TAU and untreated patients: If a researcher wants to know whether a given treatment is superior to nonpsychotherapy (or nonspecific psychotherapy), it is useful for him or her to compare the effects of the treatment with a non-psychotherapy-specific value that includes data of both TAU and untreated patients. For example, of the studies included in our meta-analysis, Monsen and Monsen (2000) did the comparison using a control group combining TAU and untreated patients. This is consistent with the criteria of APA Division 12 (Clinical Psychology) for empirically supported treatments. According to these criteria, a treatment can be regarded as empirically supported if it is statistically significantly superior to no treatment, placebo, or alternative treatment (Chambless & Hollon, 1998). For more specific questions of research, data of TAU and untreated patients should be treated separately. This should also be done if future studies confirm the assumption that the data of TAU and untreated patients differ significantly. We have reported the means and standard deviations for these conditions separately to allow for this (see prior discussion).

Estimate of Effect Size in Control Groups

For the 26 control groups of psychoanalytic and psychodynamic psychotherapy, we found a mean effect size of 0.12 ($SD = 0.19$). This is a small effect according to Cohen (1988): No treatment, wait list, or TAU led to only minimal improvements (i.e., 0.12 SD; $Mdn = 0.09$). Because the standard deviation is relatively large compared with the mean, the effects that occurred in the control groups varied considerably across studies. For this reason, we assessed a confidence interval for the average control group effect. The limits of a 95% confidence interval are 0.05 and 0.28 (Norusis, 2000). This means that the probability that the sample studied comes from a population whose average effect is between 0.05 and 0.28 is 95%. Even an effect size of 0.28 is small according to Cohen (1988).

Correlations With Duration of the Control Period

The changes in patients who do not receive a specific form of psychotherapy may vary with the duration of the waiting or TAU period. To test whether the effects occurring in control groups are small for short-term periods but larger for longer term periods, we assessed Spearman correlations between the effect sizes of the studies included and the duration of the waiting or TAU period. According to the results, the Spearman correlation was not significant ($r_s = -.16$). Thus, the effects occurring in the control conditions were independent of the duration of the wait or TAU period. They did not increase with time. This result favors the ecological validity of the results.

Furthermore, we compared the mean effect size of the 26 control groups of psychodynamic psychotherapy with the data that Grawe et al. (1994) reported for 111 studies of psychotherapy. The mean effect size we found (0.12, $SD = 0.19$) corresponds quite well with the value of 0.10 ($SD = 0.33$) reported by Grawe et al. (1994) for untreated control groups. The difference is not significant ($t = 0.41$). The larger standard deviation in the Grawe et al. data may be due to the fact that they did not control for the type of therapy under study.

Possible Applications

To demonstrate its possible use, we compared the average control group effect we acquired with those found in some selected studies of psychoanalytic or psychodynamic psychotherapy. As examples of short-term psychodynamic psychotherapy, we chose the studies of Shapiro et al. (1994), Milrod et al. (2000), and Crits-Christoph, Connolly, Azarian, Crits-Christoph, and Shappell (1996). The first study involved treatment of patients with major depression; the second, panic disorder; and the third, generalized anxiety disorder. As an example for moderate-length term psychodynamic psychotherapy, we chose the study by Stevenson and Meares (1992), in which patients with borderline personality disorder were treated. Finally, as examples of psychoanalytic therapy, we selected the studies of Rudolf, Manz, and Ori (1994); Manz, Henningsen, & Rudolf, (1995), Sandell et al. (1999, 2001), Luborsky et al. (2001), and Leichsenring, Biskup, Kreische, and Staats (2005). In these studies, patients with complex heterogeneous disorders were treated. For each of the studies, we assessed the mean effect size of psychoanalytic and psychodynamic psychotherapy applying the same procedure as described previously: Effect sizes were calculated by dividing the difference between means before and after therapy by the standard deviation before therapy (Cohen, 1988). Again, we assessed effect sizes separately for each outcome measure used in a study. If necessary, signs were reversed so that a positive effect size always indicates improvement. Whenever multiple measures were applied in a study, we assessed the effect size for each measure separately and calculated the mean effect size to assess the overall effect of the study.

For each study, the effect size achieved by psychotherapy was compared with the average effect size of control groups by t tests according to Hays (1981, p. 274, formula 8.2.3). Taking into account the large variability in control group effect sizes, we did not use the mean effect size occurring in control groups (0.12) but the upper limit of its 95% confidence interval (0.28). Using the upper limit implies a stricter empirical test. In a second step, the between-group effect size was calculated for each comparison according to Cohen (1988, p. 67, formula 2.5.3).

As can be seen in Table II, all effect sizes of psychoanalytic or psychodynamic psychotherapy reported in the studies mentioned previously significantly exceeded the average effect of the 26 control groups. The corresponding t values can be used to calculate between-group effect sizes (Cohen, 1988, p. 67). These between-group effect sizes give the difference between therapy and control group in units of standard deviations (Cohen, 1988). According to the results, the differences between the average control group effect and the effects yielded by psychoanalytic or psychodynamic psychotherapy correspond to large effects according to Cohen (1988): The applied forms of psychoanalytic or psychodynamic psychotherapy yielded effect sizes that substantially exceeded the effects of control groups (see Table II). This is true for both the effects in outcome measures that are specific or nonspecific to the disorder. In the study of Luborsky et al. (2001), for example, the effect of psychoanalytic therapy exceeded the average control group effect by at least 4.74 *SD*s (see Table II). These results are consistent with a meta-analysis that reported psychodynamic therapy to be significantly superior to wait list and TAU (Leichsenring, Rabung, & Leibing, 2004).

Discussion

In psychotherapy research, there are several attempts to assess expected treatment responses (Howard, Kopta, Krause, & Orlinsky, 1986; Kopta, Howard, Lowry, & Beutler, 1994; Lambert, 2005; Lueger et al., 2001). In this study, we have illustrated a complementary approach to the issue of control patients. We proposed to systematically assess the magnitude of change that occurs in patients with psychiatric disorders who did not receive any specific psychotherapeutic or psychiatric treatment. These data can be collected and treated like normative data of psychological tests and may serve as reference data against which the effects found in studies of psychotherapy can be compared. Epidemiological studies are required to gain data of both short-term and long-term changes in untreated patients with psychiatric disorders. At present, data like these are not yet available.

In a preliminary move to gain first data of reference, we meta-analytically evaluated 26 studies of psychodynamic psychotherapy. We found a mean control group effect size of 0.12, which can be

Table II. Comparison of Selected Trials of Psychoanalytic and Psychodynamic Therapy With the Upper Limit of the 95% Confidence Interval for the Average Effect Size of 26 Control Groups of Psychoanalytic and Psychodynamic Therapy[a].

Study	Disorder	Outcome measure *d*	*t*[b]	Between-group *d*[c]
Milrod et al. (2000)[d]	PD	Anxiety		
		Post: 1.32 ($N=14$)	20.48*	6.79
		6MFU: 1.17 ($N=10$)[e]	14.81*	5.51
		All measures		
		Post: 1.10 ($N=14$)	16.15*	5.35
		6MFU: 0.98 ($N=10$)	11.65*	4.34
Crits-Christoph et al. (1996)[d]	GAD	Anxiety: post: 1.70 ($N=26$)	38.11*	10.57
		All measures: post: 1.14 ($N=26$)	23.08*	6.40
Shapiro et al. (1994)[d]	Depression	Depression		
		Post: 2.41 ($N=28$) [f]	59.32*	16.16
		3MFU: 2.39 ($N=25$)[f]	55.53*	15.55
		All measures		
		Post: 1.91 ($N=28$)[f]	45.40*	12.36
		3MFU: 1.85 ($N=25$)[f]	41.32*	11.57
Stevenson & Meares (1992)	BPD	BPD criteria[g]: post: 2.40 ($N=30$)	61.11*	16.37
		All measures: post: 0.95 ($N=30$)	19.31*	5.17
Rudolf et al. (1994), Manz et al. (1995)[e]	Mixed	PSKB: post: 0.90 ($N=44$)	21.64*	5.35
Sandell et al. (2001)[e]	Mixed	SCL: post: 1.55 ($N=24$)	32.75*	9.27
Luborsky et al. (2001)[e]	Mixed	GAF, HSRS: post: 0.98 ($N=17$)	15.19*	4.74
Leichsenring et al. (2005)	Mixed	SCL, IIP		
		Post: 1.31 ($N=36$)	41.37*	10.65
		12MFU: 1.62 ($N=23$)	40.89*	11.70

Note. PD = panic disorder; MFU = months follow-up; GAD = generalized anxiety disorder; BPD = borderline personality disorder; PSKB = Psychischer und Sozialkommunikativer Befund (Rating of the Psychosocial State); SCL = Symptom Checklist-90-R Global Severity Index; GAF = Global Assessment of Functioning; HSRS = Health Sickness Rating Scale; IIP = Inventory of Interpersonal Problems-64.

[a]Control group effect estimate: $d=0.28$, $SD=0.19$. [b]Comparison with the control group effect estimate; $t=(M-\mu)/(s/N^{1/2})$ (Hays, 1981, p. 274). [c]Therapy vs. control group effect estimate; $d=t(1/n_1+1/n_2)^{1/2}$ (Cohen, 1988, p. 67). [d]Short-term psychodynamic therapy. [e]Psychoanalytic therapy. [f]Effects for 16 sessions (from Shapiro et al., 1995, pp. 533–534). [g]Based on *Diagnostic and Statistical Manual of Mental Disorders* (third edition).

*$p<.01$.

regarded as a small effect (Cohen, 1988). Even if one uses the upper limit of its 95% confidence interval (i.e., 0.28) as a very conservative estimate of the average effect occurring in control groups, this effect is still small according to Cohen (1988). Accordingly, TAU, wait list, or no treatment yielded little improvement in these patients.

Generally, the changes occurring in control groups consist of several components (i.e., the natural course of the disorder, error variance, and the therapeutic effect of waiting and TAU). Surprisingly, the untreated wait list control groups did not differ significantly from the TAU groups concerning the observed effect sizes. Waiting for therapy produced changes comparable to those of routine treatments. However, preliminary data of another meta-analysis have provided some evidence that TAU is superior to no treatment but inferior to psychotherapy (Leichsenring et al., 2004). These results are consistent with the overall findings reported by Grissom (1996, p. 979) that "the ranking for therapeutic success is generally therapy, placebo, and control (do-nothing or wait)." TAU generally can be expected to be superior to placebo and inferior to psychotherapy. However, further studies including a larger sample of studies are necessary to confirm this assumption.

It is important to determine whether the average control group effect that we found in the 26 studies can be regarded as representative for patients treated with psychoanalytic and psychodynamic psychotherapy. Otherwise, it cannot be used for comparisons. Several aspects of our meta-analytic evaluation suggest some representativeness: First, the sample on which the evaluation was based included most of the disorders relevant for psychotherapy. Second, the results are based on a relatively large sample of more than 700 patients. Third, through randomization, comparability with the patients who were treated in the therapy group with psychodynamic psychotherapy should be ensured. Fourth, the mean effect size we found does not differ significantly from that reported by Grawe et al. (1994) for 111 untreated control groups across all forms of psychotherapy ($d=0.10$). Low rates of response to placebo or wait list, which are consistent with our finding, were reported for chronic depression (Kocsis et al., 1998) and for panic disorder and phobia (Mattick,

Andrews, Hadzi-Pavlovic, & Christensen, 1990). Fifth, because patients in the field usually suffer from more complex disorders than those included in RCTs (Guthrie, 2000; Seligman, 1995), the changes occurring in clinical practice in untreated patients or those treated with TAU can be expected to be even lower. Thus, the estimate of the expected change we have assessed in this meta-analysis can be assumed to overestimate the changes occurring in patients in the field of clinical practice. Sixth, the effect size did not show correlations with the duration of waiting or TAU. This suggests that the mean change we found may be appropriate for comparison with the effects of both short-term and long-term psychotherapy. However, further studies are necessary to confirm this assumption.

For these reasons, the mean control group effect that occurred in the 26 studies we evaluated can be regarded to have some representativeness of psychodynamic psychotherapy. This estimate should be especially appropriate for studies in which patients with heterogeneous disorders were treated because it is based on data from different disorders that were lumped together by the meta-analytic evaluation. Shadish, Cook, and Campbell (2002) have described quasi-experimental designs for generalized causal inference. They suggested including additional comparison groups or statistical controls. Change norms of the kind that we assessed in our meta-analysis may serve this function. They may contribute to the improvement of the inferences that can be made from quasi-experimental studies. However, they are only one of many design elements on which the inferences may be based.

To demonstrate how the mean expected change can be used, we applied it to selected studies of psychodynamic and psychoanalytic therapy. The effects of psychoanalytic or psychodynamic psychotherapy reported in the selected studies were shown to significantly and substantially exceed the effects occurring in control groups of psychodynamic psychotherapy. These results are consistent with meta-analyses that showed that psychodynamic psychotherapy yielded large effect sizes in the treatment of depression, personality disorders, and other specific psychiatric disorders (Leichsenring, 2001; Leichsenring & Leibing, 2003; Leichsenring et al., 2004). Furthermore, the effect sizes yielded in psychodynamic therapy exceeded those of wait list or TAU groups and were equivalent to those of other therapies (e.g., CBT; Leichsenring et al., 2004).

Certainly, our preliminary meta-analytical evaluation includes some important limitations: Only the type of therapy was taken into account, not the type of disorder. However, changes in control groups can be expected to be different for different disorders. Furthermore, they should be different for acute and chronic disorders. In this meta-analysis, the greatest improvement was found in the phobic patients treated by TAU who were included in the investigation by Gillan and Rachman (1974). The greatest deteriorations were found in patients with borderline personality disorder in the study by Bateman and Fonagy (1999), also treated by TAU.[4] However, the effect sizes were not statistically significantly heterogeneous. Because of the relatively small number of controlled studies of psychodynamic psychotherapy, it was not possible to assess the effects of control groups both for a specific type of therapy and for specific disorders. Further research should assess changes of control groups for specific disorders. On the one hand, this can be done by further meta-analytic evaluations of existing studies across different forms of therapy: There are as many control groups of specific disorders as randomized controlled studies that demonstrate efficacy of a treatment for specific disorders. The effect sizes calculated in our study could also be included in an evaluation across different forms of therapy. Another strategy of research would be to study the spontaneous course of specific disorders (e.g., in epidemiological studies) and to gain expected values for the changes of untreated disorders. Additionally, the original data of the available studies could be evaluated and a data bank established. Furthermore, it would be interesting to apply the approach described here to other forms of treatments and to their control groups (e.g., to assess the mean changes occurring in control groups of psychopharmacological treatments). We suppose that psychiatric patients who do not receive psychopharmacological treatments also show little improvements compared with those who did receive these treatments (e.g., Kocsis et al., 1998). Because the presently available outcome measures are differentially sensitive to change, future studies should also assess the effects of control groups in specific outcome measures (e.g., in the SCL-90-R, the Inventory of Interpersonal problems, or disorder-specific measures). Another limitation is that we could only use data from studies of short-term or moderate-length psychodynamic psychotherapy. Thus, the question arises as to whether the changes we found in the control groups are valid for long-term therapies. Although the changes we found in the control groups did not show correlations with the duration of wait or TAU, studies of the long-term course in untreated patients with psychiatric disorders are necessary.

We regard our proposal as complementary to the methodology of RCTs and effectiveness studies. RCTs may be applied, for example, if a newly developed method of therapy is to be tested under

controlled conditions. Effectiveness studies are required to show whether a treatment works in the field. Reference data of untreated patients can be used to substitute for control groups, for example, to save the costs implied by a control group or as reference data for studies of long-term treatments.

Notes

[1] We refer to the latter type of condition (comparison to an other treatment) as comparison group and to no-treatment conditions as control groups.

[2] However, in many randomized controlled studies, the actual number of patients falls below the sample size necessary for randomization to work well (Hsu, 1989).

[3] We thank Rolf Sandell and his coworkers, who provided us with the data to assess the effects of their untreated control group (personal communication, May 6, 2002).

[4] This result is consistent with our assumption made early in this article that more severely disturbed patients show less positive changes without receiving psychotherapeutic treatment than less severely disturbed patients.

References

Baldoni, F., Baldaro, B., & Trombini, G. (1995). Psychotherapeutic perspectives in urethral syndrome. *Stress Medicine*, *11*, 79–84.

Bateman, A., & Fonagy, P. (1999). Effectiveness of partial hospitalization in the treatment of borderline personality disorder: A randomized controlled trial. *American Journal of Psychiatry*, *156*, 1563–1569.

Beutler, L. E. (1998). Identifying empirically supported treatments: What if we didn't? *Journal of Consulting and Clinical Psychology*, *66*, 113–120.

Brom, D., Kleber, R. J., & Defares, P. B. (1989). Brief psychotherapy for posttraumatic stress disorders. *Journal of Consulting and Clinical Psychology*, *57*, 607–612.

Canadian Task Force on the Periodic Health Examination. (1979). The periodic health examination. *Journal of the Canadian Medical Association, 121, 1193–1254*.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*, 7–18.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685–716.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cook, D. J., Guyatt, G. H., Laupacis, A., Sacket, D. L., & Goldberg, R. J. (1995). Clinical recommendations using levels of evidence for antithrombotic agents. *Chest*, *108*, 227S–230S.

Crits-Christoph, P., Connolly, M. B., Azarian, K., Crits-Christoph, K., & Shappell, S. (1996). An open trial of brief supportive-expressive psychotherapy in the treatment of generalized anxiety disorder. *Psychotherapy*, *33*, 418–430.

Deter, H. C. (1986). Cost-benefit analysis of psychosomatic therapy in asthma. *Journal of Psychosomatic Research*, *30*, 173–182.

Dührssen, A., & Jorswieck, E. (1965). An empirical-statistical study of the efficacy of psychoanalytic treatment. *Nervenarzt*, *36*, 166–169.

Fonagy, P. (1999). Process and outcome in mental health care delivery: A model approach to treatment evaluation. *Bulletin of the Menninger Clinic*, *63*, 288–304.

Gillan, P., & Rachman, S. (1974). An experimental investigation of desensitization in phobic patients. *British Journal of Psychiatry*, *124*, 392–401.

Gloaguen, V., Cottraux, J., Cucherat, M., & Blackburn, I. M. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders*, *49*, 59–72.

Grawe, K., Donati, R., & Bernauer, F. (1994). *Psychotherapie im Wandel. Von der Konfession zur Profession* [Psychotherapy in change. From confession to profession]. Göttingen: Hogrefe.

Grissom, R. J. (1996). The magical number .7 +/− .2: Meta-meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo, and control. *Journal of Consulting and Clinical Psychology*, *64*, 973–982.

Guthrie, E. (2000). Psychotherapy for patients with complex disorders and chronic symptoms. The need for a new research paradigm. *British Journal of Psychiatry*, *177*, 131–137.

Guthrie, E., Creed, F., Dawson, D., & Tomenson, B. (1993). A randomised controlled trial of psychotherapy in patients with refractory irritable bowel syndrome. *British Journal of Psychiatry*, *163*, 315–321.

Guthrie, E., Kapur, N., Mackway-Jones, K., Chew-Graham, C., Moorey, J., Mendel, E., et al. (2001). Randomised controlled trial of brief psychological intervention after deliberate self poisoning. *BMJ*, *323*, 135–138.

Guthrie, E., Moorey, J., Margison, F., Barker, H., Palmer, S., McGrath, G., et al. (1999). Cost-effectiveness of brief psychodynamic-interpersonal therapy in high utilizers of psychiatric services. *Archives of General Psychiatry*, *56*, 519–526.

Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., & Cook, R. J. (1995). Users' guides to the medical literature: IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA*, *274*, 1800–1804.

Hamilton, J., Guthrie, E., Creed, F., Thompson, D., Tomenson, B., Bennett, R., et al. (2000). A randomized controlled trial of psychotherapy in patients with chronic functional dyspepsia. *Gastroenterology*, *119*, 661–669.

Hays, W. L. (1981). *Statistics*. New York: Holt.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist*, *41*, 159–164.

Hsu, L. M. (1989). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, *57*, 131–137.

Kocsis, J. H., Frances, A. J., Voss, C., Mason, B. J., Mann, J. J., & Sweeney, J. (1998). Imipramine and social-vocational adjustment in chronic depression. *American Journal of Psychiatry*, *145*, 997–999.

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.

Lambert, M. J. (2005). Early response in psychotherapy: Further evidence for the importance of common factors rather than "placebo effects.". *Journal of Clinical Psychology*, *61*, 855–869.

Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds), *Handbook of psychotherapy and behavior change* (4th ed,. pp. 143–189). New York: Wiley.

Leichsenring, F. (2001). Comparative effects of short-term psychodynamic psychotherapy and cognitive-behavioral therapy in depression: A meta-analytic approach. *Clinical Psychology Review*, *21*, 401–419.

Leichsenring, F. (2004). Randomized controlled vs. naturalistic studies. A new research agenda. *Bulletin of the Menninger Clinic*, *68*, 115–129.

Leichsenring, F., Biskup, J., Kreische, R., & Staats, H. (2005). The effectiveness of psychoanalytic therapy. First results of the "Göttingen study of psychoanalytic and psychodynamic therapy. *International Journal of Psychoanalysis*, *86*, 433–455.

Leichsenring, F., & Leibing, 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., & Leibing, E. (2004). The efficacy of short-term psychodynamic therapy in specific psychiatric disorders: A meta-analysis. *Archives of General Psychiatry*, *61*, 1208–1216.

Levis, D. J., & Carrera, R. (1967). Effects of ten hours of implosive therapy in the treatment of outpatients: A preliminary report. *Journal of Abnormal Psychology*, *72*, 504–508.

Luborsky, L., Stuart, J., Friedman, S., Diguer, L., Seligman, D. A., Bucci, W., et al. (2001). The Penn Psychoanalytic Treatment Collection: A set of complete and recorded psychoanalyses as a research resource. *Journal of the American Psychoanalytical Association*, *49*, 217–234.

Lueger, R. J., Howard, K. I., Martinovich, Z., Lutz, W., Anderson, E. E., & Grissom, G. (2001). Assessing treatment progress of individual patients using expected treatment response models. *Journal of Consulting and Clinical Psychology*, *69*, 150–158.

Manos, N., & Vasilopoulou, E. (1984). Evaluation of psychoanalytic psychotherapy outcome. *Acta Psychiatrica Scandinavica*, *70*, 28–35.

Manz, R., Henningsen, C., & Rudolf, G. (1995). [Methodological and statistical aspects of therapy evaluation exemplified by the Berlin psychotherapy study]. *Psychotherapie Psychosomatik Medizinische Psychologie*, *45*, 52–59.

Marmar, C. R., Horowitz, M. J., Weiss, D. S., Wilner, N. R., & Kaltreider, N. B. (1988). A controlled trial of brief psychotherapy and mutual-help group treatment of conjugal bereavement. *American Journal of Psychiatry*, *145*, 203–209.

Mattick, R. P., Andrews, G., Hadzi-Pavlovic, D., & Christensen, H. (1990). Treatment of panic and agoraphobia. An integrative review. *Journal of Nervous and Mental Disease*, *178*, 567–576.

Meyer, A. E., & Bolz, W. (1981). The Hamburg short psychotherapy comparison experiment: Foreword and introduction. *Psychotherapy Psychosomatics*, *35*, 81–84.

Milrod, B., Busch, F., Leon, A. C., Shapiro, T., Aronson, A., Roiphe, J., et al. (2000). Open trial of psychodynamic psychotherapy for panic disorder: A pilot study. *American Journal of Psychiatry*, *157*, 1878–1880.

Monsen, K., & Monsen, T. J. (2000). Chronic pain and psychodynamic body therapy: A controlled outcome study. *Psychotherapy*, *37*, 257–269.

Nathan, P. E., & Gorman, J. M. (Eds.). (2002). *A guide to treatments that work* (2nd ed.). New York: Oxford University Press.

Newton, J. R., & Stein, L. I. (1974). Implosive therapy in alcoholism. Comparison with brief psychotherapy. *Quarterly Journal on the Study of Alcohol*, *35*, 1256–1265.

Norusis, M. J. (2000). *SPSS 10.0 guide to data analysis*. Upper Saddle River, NJ: Prentice Hall.

Persons, J. B., & Silberschatz, G. (1998). Are results of randomized controlled trials useful to psychotherapists? *Journal of Consulting and Clinical Psychology*, *66*, 126–135.

Pilowsky, I., & Barrow, C. G. (1990). A controlled study of psychotherapy and amitriptyline used individually and in combination in the treatment of chronic intractable, "psychogenic" pain. *Pain*, *40*, 3–19.

Rosser, R., Denford, J., Heslop, A., Kinston, W., Macklin, D., Minty, K., et al. (1983). Breathlessness and psychiatric morbidity in chronic bronchitis and emphysema: A study of psychotherapeutic management. *Psychological Medicine*, *13*, 93–110.

Roth, A. D., & Parry, G. (1997). The implications of psychotherapy research for clinical practice and service development: Lessons and limitations. *Journal of Mental Health*, *6*, 367–380.

Rudolf, G., Manz, R., & Ori, C. (1994). [Results of psychoanalytic therapy]. *Zeitschrift für Psychosomatische Medizin und Psychoanalyse*, *40*, 25–40.

Sandell, R., Blomberg, J., & Lazar, A. (1999). [Repeated long-term follow-up of long-term psychotherapy and psychoanalysis. First findings of the Stockholm Outcome of Psychotherapy (STOP) Project]. *Zeitschrift für Psychosomatische Medizin und Psychoanalyse*, *45*, 43–56.

Sandell, R., Blomberg, J., & Lazar, A. (2002). Time matters: On temporal interactions in long-term follow-up of long-term psychotherapies. *Psychotherapy Research*, *12*, 39–58.

Sandell, R., Blomberg, J., Lazar, A., Carlsson, J., Broberg, J., & Schubert, J. (2001). [Differences in long-term effects of psychoanalysis and extended psychotherapy: Findings of the Stockholm Outcome of Psychoanalysis and Psychotherapy Project]. *Psyche*, *55*, 277–310.

Seligman, M. E. (1995). The effectiveness of psychotherapy. The Consumer Reports study. *American Psychologist*, *50*, 965–974.

Seligman, M. E. (1996). Science as an ally of practice. *American Psychologist*, *51*, 1072–1079.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology*, *62*, 522–534.

Shapiro, D. A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin*, *92*, 581–604.

Shefler, G., Dasberg, H., & Ben-Shakhar, G. (1995). A randomized controlled outcome and follow-up study of Mann's time-limited psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*, 585–593.

Siegel, S. M., Rootes, M. D., & Traub, A. (1977). Symptom change and prognosis in clinic psychotherapy. *Archives of General Psychiatry*, *34*, 321–329.

Sifneos, P. E. (1990). *Short-term dynamic psychotherapy. Evaluation and technique*. New York: Plenum Press.

Sloane, R. B., Staples, F. R., Cristol, A. H., & Yorkston, N. J. (1975). Short-term analytically oriented psychotherapy versus behavior therapy. *American Journal of Psychiatry*, *132*, 373–377.

Snyder, D. K., & Wills, R. M. (1989). Behavioral versus insight-oriented marital therapy: Effects on individual and interspousal functioning. *Journal of Consulting and Clinical Psychology*, *57*, 39–46.

Stevenson, J., & Meares, R. (1992). An outcome study of psychotherapy for patients with borderline personality disorder. *American Journal of Psychiatry*, *149*, 358–362.

Strupp, H. H., Horowitz, L. M., & Lambert, M. J. (Eds.). (1997). *Measuring patient changes in mood, anxiety, and personality disorders: Toward a core battery*. Washington, DC: American Psychological Association.

Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training and dissemination of empirically-validated psychological treatments: Report and recommendations. *Clinical Psychology*, *48*, 3–23.

Thompson, L. W., Gallagher, D., & Breckenridge, J. S. (1987). Comparative effectiveness of psychotherapies for depressed

elders. *Journal of Consulting and Clinical Psychology*, *55*, 385–390.

Winston, A., Pollack, J., McCullough, L., Flegenheimer, W., Kestenbaum, R., & Trujillo, M. (1991). Brief psychotherapy of personality disorders. *Journal of Nervous and Mental Disease*, *179*, 188–193.

Woody, G., Luborsky, L., McLellan, A. T., & O'Brien, C. P. (1990). Corrections and revised analyses for psychotherapy in methadone maintenance patients. *Archives of General Psychiatry*, *47*, 788–789.

Woody, G. E., McLellan, A. T., Luborsky, L., & O'Brien, C. P. (1995). Psychotherapy in community methadone programs: A validation study. *American Journal of Psychiatry*, *152*, 1302–1308.

Zusammenfassung

Veränderungsnormen: Ein komplementärer Ansatz zum Kontrollgruppenproblem in der Psychotherapieergebnisforschung

Ziel: Kontrollgruppen werden oft als unabdingbare Bedingung in der Psychotherapie-Ergebnisforschung angesehen. Sie sind jedoch mit hohen Kosten verbunden, sie sind außerdem kaum realisierbar bei der Untersuchung von Langzeit-Psychotherapie. In dieser Arbeit gehen wir die Frage der Kontrollgruppe in der Psychotherapieforschung auf eine neue Art an. Wir schlagen vor, die Veränderungen systematisch zu untersuchen, die bei Patienten mit psychischen Störungen auftreten, die keine spezifische Behandlung bekommen. Solche Daten können als Referenz-Werte für zukünftige Ergebnisstudien dienen. Es wäre dann nicht länger notwendig, solche Daten in jeder Studie immer wieder neu zu erheben.

Methode: Um diesen Zugang zu exemplifizieren, untersuchten wir in einem ersten Ansatz die Veränderungen, die in den vorliegenden Kontrollgruppen von randomisierten kontrollierten Studien (RCTs) psychodynamischer Therapie auftraten. Wir schlossen 26 RCTs psychodynamischer Therapie, in denen reliable und valide Ergebnismasse verwendet worden sind und die die notwendigen Daten zur Bestimmung von Effektgrößen lieferten, in eine Meta-Analyse ein. Effektgrößen wurden für die mittlere Veränderung berechnet, die in den Kontrollgruppen dieser Studien auftraten.

Ergebnisse: Der mittlere Effekt, der in diesen 26 Kontrollgruppen auftrat, war d = 0.12 (95% Konfidenzintervall: 0.05–0.28). Dies entspricht einem kleinen Effekt. Um die mögliche Anwendung dieses Wertes zu demonstrieren, verglichen wir den oberen Grenzwert des 95% Konfidenzintervalls (d = 0.28), also einen sehr konservativen Schätzer der durchschnittlichen Kontrollgruppen-Effektgröße, mit ausgewählten Studien psychodynamischer Psychotherapie.

Schlussfolgerungen: Die durchschnittliche Veränderung, die bei Patienten mit psychischen Störungen in RCTs psychodynamischer Psychotherapie aufgetreten ist, kann als ein vorläufiger Schätzer für Kontrollgruppen (Veränderungsnorm) für unkontrollierte Psychotherapiestudien dienen. Zukünftige Studien sollten die durchschnittlich zu erwartenden Veränderungen in Kontrollgruppen spezifischer untersuchen (z.B. für spezifische psychische Störungen, Ergebnismasse oder Patientenvariablen.)

Résumé

Changer les normes : une approche complémentaire au sujet des groupes de contrôle dans la recherche de l'efficacité des psychothérapies

Le groupe de contrôle est généralement considéré comme indispensable pour la recherche de l'efficacité des psychothérapies. Cependant, il cause des coûts importants et ne s'applique pas à l'étude des traitements à long terme. Les auteurs approchent le sujet du groupe de contrôle d'une nouvelle manière : ils étudient systématiquement les changements chez des patients psychiatriques n'ayant reçu aucun traitement. Ces données serviront, par la suite, de données de référence pour d'autres études d'efficacité. Il ne serait ainsi plus nécessaire d'obtenir à répétition des données de référence pour chaque nouvelle étude. Les auteurs ont évalué les changements survenus dans des groupes de contrôles d'études randomisées et contrôlées (RCTs) disponibles de psychothérapies psychodynamiques, dans un premier test préliminaire. 26 RCTs de psychothérapies psychodynamiques, avec des mesures fidèles et valides et des nombres suffisants pour calculer la grandeur de l'effet, faisaient partie de cette méta-analyse. La grandeur de l'effet était calculée pour les changements moyens survenus dans les groupes de contrôle. La grandeur de l'effet moyenne de 0.12 (intervalle de confiance à 95% : 0.05–0.28) correspond à un petit effet. Pour démontrer son utilisation possible, la limite supérieure de l'IC à 95% (d = 0.28), considérée comme une estimation très conservatrice de la grandeur de l'effet moyen des groupes de contrôle, a été comparée avec celle d'études de psychothérapies psychodynamiques sélectionnées. Le changement moyen de patients psychiatriques inclus dans des RCTs de psychothérapies et qui n'avaient pas reçu de traitement particulier peut servir d'estimation préliminaire de groupe contrôle (norme de changement) pour des études ouvertes de psychothérapie. La recherche devrait, dans le futur, évaluer le changement moyen attendu pour des contrôles plus sophistiqués (p.ex., des troubles psychiatriques spécifiques, des mesures d'efficacité ou des variables des patients).

Resumen

Cambio de Norma: un enfoque complementario al tema de los grupos de control en la investigación de resultados en psicoterapia

Los grupos de control son ampliamente considerados como un *sine qua non* en la investigación de los resultados en psicoterapia. Sin embargo, provocan altos costos y no son aplicables a estudios prolongados. Los autores tratan este problema en una forma nueva: sistemáticamente estudian los cambios ocurridos en pacientes psiquiátricos que no recibieron ningún tratamiento específico. Tales datos pueden servir como referencia para posteriores estudios de resultados. En ese caso no sería ya necesario reunir datos de referencia en cada nuevo estudio. Los autores evaluaron los cambios ocurridos en los grupos de control de estudios controlados randomizados disponibles (RCTs) de psicoterapia psicodinámica en un ensayo primero y preliminar. En un metaanálisis se incluyeron veintiséis RCTs de psicoterapia psicodinámica, en los que se aplicaron medidas confiables y válidas de resultados y

que proveyeron los datos necesarios para calcular la amplitud de los efectos. Estos se calcularon para los cambios en la media que ocurrieron en los grupos de control. (*Effect sizes were calculated for the mean changes occurring in the control groups)*. La amplitud del efecto medio de 0.12 (95% del intervalo de confianza (CI): 0.05–0.28) corresponde a un efecto pequeño. Para demostrar su uso posible, el límite superior del 95% CI (d = 0.28), considerado un estimado muy conservador de la amplitud del efecto promedio del grupo control, se comparó con el de estudios seleccionados de psicoterapia psicodinámica. El cambio promedio en pacientes psiquiátricos incluido en los RCT de la psicoterapia que no recibieron tratamiento específico pueden servir como estimado preliminar del grupo control (cambio de norma) para ensayos abiertos de psicoterapia. Futuras investigaciones deberían estimar el cambio medio esperado en controles más sofisticados (v.g., para desórdenes psiquiátricos específicos, medidas de resultado o variables de los pacientes).

Resumo

Mudar as normas: Uma abordagem complementar ao tema dos grupos de controlo na investigação de resultados psicoterapêuticos

Os grupos de controlo são considerados, de um modo geral, como uma condição *sine qua non* na investigação de resultados psicoterapêuticos. No entanto, estes envolvem custos elevados e não são aplicáveis para os estudos de longo prazo. Os autores abordam o tema dos grupos de controlo na investigação psicoterapêutica de um novo modo: Estudam sistematicamente as mudanças que ocorrem em pacientes psiquiátricos que não receberam qualquer tratamento específico. Tais dados poderão servir como referência para estudos futuros de resultados terapêuticos. Deste modo, não seria necessário recolher dados de referência, repetidamente, em cada novo estudo. Os autores avaliaram as mudanças que ocorreram nos grupos de controlo dos estudos clínicos randomizados (ECR) de psicoterapia psicodinâmica disponíveis num primeiro ensaio preliminar. Foram incluídas numa meta-análise, vinte e seis ECR de psicoterapia psicodinâmica, nas quais se usaram medidas de resultados terapêuticos fidedignas e validas e que forneceram dados necessários para calcular a magnitude do efeito. A magnitude do efeito foi calculada para a média das mudanças que ocorreram nos grupos de controlo. A média da magnitude do efeito foi de 0.12 (intervalo de confiança 95% [IC]: 0.05–0.28) correspondendo a um efeito pequeno. Para demonstrar o seu possível uso, o limite superior do intervalo de confiança de 95% (d = 0.28), considerado uma estimativa conservadora da média do tamanho do efeito do grupo de controlo, foi comparado com o de estudos seleccionados de psicoterapia psicodinâmica. A mudança média dos pacientes psiquiátricos incluída nos ECR's de psicoterapia que não receberam tratamento específico poderá servir como uma estimativa preliminar do grupo de controlo (norma de mudança) para os ensaios abertos de psicoterapia. Estudos futuros deverão avaliar a mudança média esperada em controlos mais específicos (e.g., para perturbações psiquiátricas específicas, medidas de resultados ou variáveis do paciente).

Sommario

Cambiamento delle norme: un approccio complementare alla questione dei gruppi di controllo nella ricerca di esito in psicoterapia

I gruppi di controllo sono largamente considerati un sine qua non della ricerca di esito in psicoterapia. Tuttavia essi hanno alti costi e non sono applicabili agli studi a lungo termine.

Gli autori indirizzano la questione dei gruppi di controllo nella ricerca in psicoterapia in un modo innovativo: essi sistematicamente studiano i cambiamenti che avvengono nei pazienti psichiatrici che non hanno ricevuto nessun trattamento specifico. Tali dati potrebbero servire come dati di riferimento per ulteriori studi di esito. Presto, allora, sarebbe necessario prendere i dati di riferimento ripetibili in ogni e ciascuno studio.

Gli autori hanno stimato i cambiamenti che avvengono nei gruppi di controllo degli studi controllati randomizzati (RCTs) disponibili di psicoterapia psicodinamica in un primo e preliminare esperimento. Ventisei RCTs di psicoterapia psicodinamica, nei quali erano applicate misure di esito valide ed affidabili e che fornivano i dati necessari per calcolare l'effect sizes, sono stati inclusi in una meta-analisi. Gli effect sizes sono stati calcolati per i cambiamenti medi avvenuti nei gruppi di controllo. L'effect size medio di 0.12 (95% intervallo di confidenza (CI): 0.05–0.28) corrisponde ad un piccolo effetto. Per dimostrare il suo possibile uso, il limite superiore del 95% CI (d = 0.28), considerato molto cauto nella media dell'effect size del gruppo di controllo, è stato confrontato con quello degli studi selezionati di psicoterapia psicodinamica. Il cambiamento medio nei pazienti psichiatrici inclusi nei RCTs di psicoterapia che non avevano ricevuto trattamenti specifici potrebbe servire come valutazione preliminare del gruppo di controllo (cambiamento di norme) per esperimenti aperti di psicoterapia.

Ulteriori ricerche valuterebbero il peggior cambiamento atteso in controlli più sofisticati (per es. per disturbi psichiatrici specifici, misure di esito o variabili del paziente).

常模的改變：關於控制組在心理治療結果研究的一個補充式觀點

摘要

控制組廣泛地被視為心理治療結果研究必要的程序，然而他們也造成高度成本以及不適用於長期研究的問題。作者嘗試提出控制組如何在心理治療研究的新穎作法，也就是系統性地研究一些沒有接受任何治療的精神病患者的改變歷程。這樣的資料就可以做為未來心理治療結果研究的參照資料，如此將不需要在每一個新的研究重複蒐集參照資料。作者評估那些可進行隨機控制研究（RCTs）的心理動力治療的第一次以及預備試驗中，所發生在控制組裡的改變。研究對 26 個具良好信效度結果評量、且能提供統計考驗力的心理動力治療的 RCTs 進行整合分析。統計考驗力是根據改變發生在控制組的平均值所計算，結果是 0.12（95%的信心區間〔CI〕：0.05—0.28）相對屬於小的統計考驗力。為示範其使用的可能性，當做其他心理動力研究的參照資料，作者建議 95%的上限信心區間 CI（$d = 0.28$）可被視為非常保守的控制組統計考驗力的平均估計。在 RCTs 的心理治療中沒有接受任何治療的精神病患者的平均改變可視為預備控制組的估計資料。未來研究應評估在更精緻的控制研究中的平均預期改變，例如對特定的精神疾病、結果評量或患者變項。