Researchers have found it difficult to comply entirely with modern methodological standards when studying the outcomes of the very long processes in psychoanalysis. The major shortcomings have been in specifying the treatment given, in specifying its patients, and in controlling for confounding factors, such as selection bias for instance, through randomization. Where short-term psychodynamic psychotherapies are concerned, researchers are becoming more compliant with the standards and have been able to publish empirical support for their positive effects, especially with depressed patients. Moderate-length and long-term psychodynamic psychotherapies are less well supported, but research suggests they may be uniquely effective for patients with complex and long-standing personality-based problems or disorders. Psychoanalysis proper has
been tested in various designs, which were typically not well controlled. Studies suggest that psychoanalysis may generate results superior to those of psychodynamic psychotherapy, but to what extent this is due to mere treatment intensity or duration is not clear. Symptoms and focused problems tend to change more than relations and personality-based problems in response to treatment, and the treatment results tend to improve during follow-up. Studies of the effects of psychoanalytically oriented treatments on health care utilization have shown mixed results.

A Methodological Approach

How do we know that an intervention has had an effect? First of all, we would have to establish that the intervention has indeed taken place the way it was supposed to be delivered. Second, we would have to make sure that those who have been subjected to it have indeed improved—and stay improved, or continue to improve, after treatment. The general idea is to compare pre-treatment and posttreatment observations. Third, we must ascertain that the improvement is in response to the intervention and not to some other factor[s]. One such factor is chance. Our observations may have been influenced by irrelevant coincidences in ourselves, in the person being observed, in the observational instrumentation, or in the environment. This results in random errors, which are essentially controlled by increasing the size of the sample of observations and improving the reliability of the observations by controlling and standardizing the conditions under which they are made. Another type of “other factors” results in systematic errors, leading to unsatisfactory internal validity. This occurs when patients change systematically in response to factors other than the intervention itself, referred to as confounding factors or confounders. One such factor is almost always time—that is, the passing of time—because very often people change across time, and often for the better if they were bad off to begin with. Another factor has to do with maturation, which of course is especially relevant in the case of interventions with children and adolescents. Still another is events and trends occurring along with the intervention but merely coincidental to it, such as employment rates, crises, judicial changes, and other changes in society. We would also like to know whether improvement is due to specific technical ingredients of the intervention or to nonspecific factors inherent in any intervention, such as hope, attention, and caring. Also, where a formal study is concerned, the mere fact of being subjected to the extra attention and more or less intense formal observations is one more such factor that we would like to exclude as a cause, as would the vested interest of the person or persons who are doing the study. Finally, selection bias is a frequent confounder when patients are assigned to treatments on the basis of clinical assessments or preferences.

Ensuring internal validity calls for various controls, often in the form of one or more groups assigned to contrasting non-intervention conditions. In principle, we have to make the intervention/non-intervention contrast independent of possible confounders by keeping them constant, by matching the contrasting groups on them, or by randomizing the contrast (i.e., assigning patients to intervention or non-intervention on a chance basis). Obviously, establishing change and assigning it to its proper cause are subject to numerous complicating factors when a formal research study is being conducted, and even more so under natural conditions in the clinic. These complications and the measures taken to counteract them have to be taken into account when one is reviewing research on the outcomes of psychoanalytically informed treatments. The extent to which researchers have been able to control for these and similar complicating factors in outcome studies—or have been unable to control them or are uninterested in controlling them—critically determines whether the findings are convincing in building a case for psychoanalytically informed treatments in the eyes of third-party payers, and many second-party ones, too.

Short-Term Psychoanalytically Informed Therapies

Psychological therapies based on psychoanalytic principles come in many forms—for individuals or groups, couples or families: short-term or long-term; time-limited or not; low-frequency or high-frequency. There is now ample evidence that individual psychoanalytically oriented or psychodynamic short-term therapy will achieve treatment results that are respectable and on a
par with comparable nonpsychoanalytic or nonpsychodynamic therapies. Thus, after an initial, largely negative meta-analysis (Svartberg and Stiles 1993), an increasing number of meta-analyses have yielded considerably more positive results. Crits-Christoph (1992), Anderson and Lambert (1995), Leichsenring et al. (2004), Fonagy et al. (2005), Abbass et al. (2006), Connolly Gibbons et al. (2008), Lewis et al. (2008), and Gerber et al. (2011), among others, have compiled and reviewed primary studies of [mostly] short-term psychodynamic psychotherapies (STPP) with [mostly] adult patients. Abbass et al. (2006) concluded that “STPP shows promise, with modest to moderate, often sustained gains for a variety of patients” (p. 1). Yet they added, “However, given the limited data and the heterogeneity between the studies, these findings should be interpreted with caution.” In their review, Lewis et al. (2008) concluded that STPP is not superior to other forms of psychotherapy but is certainly not inferior, and perhaps more importantly, they cautioned that “further high-quality studies are required of STPP focused on specific clinical problems” (p. 454).

Besides the scarcity of studies, one problem that has been noted by several reviewers (e.g., Abbass et al. 2006; Connolly Gibbons et al. 2008; Fonagy et al. 2005) is that the psychodynamic therapy being delivered in many studies has been poorly specified and that the therapies delivered in different studies may have been quite different in their methods in terms of supportiveness, focus on transference processes, and so forth. Therefore, positive findings for one version of STPP cannot automatically be taken as support for other versions. At most, the various studies may possibly be interpreted to support some common psychodynamic or psychoanalytic factor or factors, but it is not yet clear which these are.

From a psychiatric perspective, it is also considered a problem that the meta-analyses and most of the primary studies were conducted on patient samples that were diagnostically mixed. That psychiatric diagnoses may be irrelevant in a psychoanalytic perspective, where intrapsychic dynamics are the primary problem, and that comorbidity seems to be the rule in psychiatric samples are weak arguments in any contemporary discussion of what constitutes evidence for effective treatment of psychological disturbances. Thus, many primary studies of psychoanalytically informed therapies are dismissed precisely on the ground that the patient samples are heterogeneous.

### Short-Term Psychodynamic Therapy With Diagnostically Specific Samples

Nevertheless, there are some meta-analyses that have focused on specific diagnostic groups of patients, especially depressed patients, in psychoanalytic, mostly short-term, therapies (Cuypers et al. 2008a; Driessen et al. 2010; Leichsenring 2001). Cuypers et al. (2008a) reported that “we found very few indications that several important types of psychological treatment for depression differ significantly from each other [in their efficacy]. No significant difference was found for cognitive-behavior therapy, psychodynamic therapy, behavioural activation treatment, problem-solving therapy, and social skills training” (p. 917). Comparing STPP and pharmacotherapy for depressed patients, de Maat et al. (2006) summarized findings from 10 studies and concluded that “psychotherapy and pharmacotherapy appear equally efficacious…. Both treatments have larger effects in mild than in moderate depression, but similar effects in chronic and nonchronic depression and at follow-up psychotherapy outperforms pharmacotherapy” (p. 566). Although a recent Finnish study confirmed the comparability between STPP and pharmacotherapy for depressed patients (Salminen et al. 2008), a recent meta-analysis claimed that pharmacological treatments may be superior with dysthymic patients and that SSRIs are slightly more effective with major depression (Cuypers et al. 2008b). There should be no doubt by now, however, that STPP has shown evidence of being efficacious in dealing with mild or moderate depressive states.

Given the central adaptive role of anxiety as a signal or message rather than an affliction, according to psychoanalytic theory, there are still relatively few studies of STPP with patients diagnosed with anxiety disorders. In their meta-analysis, Lewis et al. (2008) listed only two controlled studies [and three naturalistic ones], and Leichsenring (2005; Leichsenring and Leibing 2007) found none [but one involving moderate-length psychodynamic therapy with social phobics]. During the last few years, however, a manualized psychoanalytically informed therapy for panic disorder has been introduced by Milrod and tested with positive results (Milrod et al. 2007). Also, positive results have been reported for STPP with generalized anxiety disor-
Substance abuse and dependence constitute another group of diagnoses with a body of controlled research to suggest that the psychodynamic approach is a viable one (Crits-Christoph et al. 1999, 2008; Woody et al. 1987, 1995). In the clinical spectrum, one is then entering the band of personality disorders. Certainly, in view of the frequency of comorbidity with personality disorders in samples with Axis I syndromes, we may question which is the more fundamental condition and whether the personality disorder should not be prioritized in treatment. Leichsenring and Leibing (2003) performed a meta-analysis of published studies of psychodynamic and cognitive-behavioral therapies in the treatment of personality disorders. Most of the studies were naturalistic and nonrandomized, and in only 3 of 22 of the studies were direct comparisons made. Leichsenring and Leibing’s conclusion corroborates that of Perry and colleagues’ (1999) earlier meta-analysis of a partially overlapping sample of studies: psychodynamic therapies, like cognitive-behavioral ones, are quite effective for patients with these complex clinical states. An interesting detail in the Perry et al. analysis was that the effects in general were larger when based on judgments of independent observers. Furthermore, Leichsenring and Leibing’s (2003) analysis showed that this difference applied to the psychodynamic studies only and was reversed among the cognitive-behavioral studies. On the basis of these results, the indications for psychodynamic therapy with patients with personality disorders are quite compelling.

Although not included in these meta-analyses, more studies on samples of inpatients or day hospital patients with personality disorders have been reported (Bateman and Fonagy 1999, 2001, 2003; Chiesa and Fonagy 2000, 2003; Chiesa et al. 2004; Dolan et al. 1997; Gabbard et al. 2000). In view of the severity of illness in these patient populations, the treatment results are impressive, but it is not possible to sort out the contributions of the psychotherapy component in these treatment packages.

Particular attention, at least in Europe, has been paid to the development of mentalization-based treatment (MBT; Bateman and Fonagy 2004, 2006). Although derived in part from attachment and cognitive theories, MBT is basically psychodynamic in its approach. There is evidence accumulating for the efficacy of MBT protocols with quite disturbed borderline patients. Especially interesting, Bateman and Fonagy (2008) followed up on a patient sample 5 years after treatment discharge. The outcome parameters were socially obtrusive yet nonreactive criteria like suicide attempts, hospitalization, emergency visits, medication, and employment. On all parameters, the MBT group was still significantly and clearly superior to the comparison group whose members had undergone so-called treatment as usual (TAU). Bateman and Fonagy (2009) recently reported findings from a randomized controlled trial of MBT in comparison with structured clinical management with outpatients with a diagnosis of borderline personality disorder. Whereas there was significant improvement in both the MBT and the clinical management groups, improvement was greater in the MBT group based on both self-reported criteria and more “objective” ones like suicide attempts and hospitalization.

Another explicitly specified, psychoanalytically informed therapy is transference-focused psychotherapy (TFP; Clarkin et al. 1999). TFP has been evaluated in a strictly controlled comparison with dialectical behavior therapy (DBT; Linehan 1993) and psychoanalytically informed supportive therapy (Clarkin et al. 2007; Levy et al. 2006). From a psychoanalytic point of view, it is interesting that, besides matching DBT in terms of level of functioning and reduction of suicidality, TFP was the only condition to increase attachment security and reflective functioning.

After a study by Giesen-Bloo et al. (2006) in the Netherlands, with less positive results, Doering et al. (2010) reported on a German-Austrian study of borderline patients randomly assigned to 1 year of either manualized TFP or nonmanualized psychotherapy delivered by experienced private practitioners in the communities. TFP brought about significantly more positive outcomes in terms of treatment dropout, suicide attempts, and borderline symptoms.
The pioneer brand among manualized psychoanalytically informed psychotherapies is supportive-expressive psychotherapy (SEP; Luborsky 1984). Luborsky et al. (1988; Mintz et al. 1979) summarized the findings of a large stringently designed naturalistic study (the Penn Study) of SEP manualized psychotherapy with mostly anxious or depressed patients. Ratings by independent judges, as well as by the patients themselves, indicated moderate to large effects on different outcome measures. Follow-up in a subsample after 5 years showed that the gains had not only been maintained but had even tended to increase after termination. Recently, Vinnars et al. (2005, 2007) reported from a randomized study of SEP in a design similar to that of Doering et al.’s (2010) study, comparing manualized SEP with nonmanualized psychodynamic psychotherapy in a sample of patients with personality disorders. In both conditions, patients improved in terms of level of functioning, psychiatric symptoms, and diagnosis.

**Moderate-Length Psychotherapies**

These positive findings notwithstanding, it is an important observation that treatment duration in the studies summarized so far generally was rather low, from a psychoanalytical point of view. In fact, there are only a few studies of truly long-term therapies. The relatively brief duration of treatments in published studies became evident in a meta-analysis by Leichsenring and Rabung (2008). After a review of more than 4,000 papers, Leichsenring and Rabung were able to identify not more than 23 studies on psychoanalytically informed treatments longer than 1 year or 50 sessions. Although the authors chose to call the therapies described in those studies “long-term,” “moderate-length” therapies seems like a more appropriate categorization from a psychoanalytical perspective. In general, the patients had personality disorders, with multiple and long-standing mental disorders. The effects were impressively strong, in general, and significantly more so than those of shorter-term forms of psychotherapy when such comparisons had been done, and outcome was again correlated with treatment duration. Especially important, again, the effects were consistently larger at follow-up than at treatment termination.

**Psychoanalysis and Long-Term Psychoanalytically Oriented Psychotherapy**

**General Issues in Outcome Research on Psychoanalysis**

Three important questions about outcome research are especially relevant from a psychoanalytic perspective:

- Is research on psychodynamic psychotherapies generalizable to psychoanalysis?
- Is outcome research harmful to the treatment?
- How should we define psychoanalysis?

In view of the fact that the positive findings from research on psychodynamic psychotherapies were mainly based on studies of less-than-intense, moderate-length treatments, we may well wonder whether these findings are generalizable to intense, “really long-term” treatments or processes of the kind that are called “psychoanalysis” and are recognized as such in the psychoanalytic community. Considering the positive association between duration and outcome typically found in psychotherapy research (Orlinsky et al. 2004), a reasonable generalization would be that psychoanalysis proper would have even more positive effects. On the other hand, some authorities entrusted with the privilege of formulating treatment guidelines have concluded that the evidence for short-term treatment cannot automatically be extended to long-term treatment. So, an important issue is whether short- and moderate-term psychotherapies are merely lesser forms of psychoanalysis or whether they are qualitatively different modes of treatment. This question has been the subject of much discussion—and not much empirical research (Grant and Sandell 2004; Kächele 2010).

Another concern is based on the widespread belief among psychoanalysts that the research process itself, with its controls and its measurements, will influence the psychoanalytic or psychotherapeutic process in ways that paradoxically will rob its findings of any external validity and also jeopardize the treatment. Although little solid evidence exists either for or against these beliefs, the evidence there is suggests that patients and therapists are indeed affected by participating in research—but the influence seems to be positive more often than negative (Busch et al. 2000; Marshall et al. 2001).
The most direct way to test whether the relatively positive findings on psychoanalytically informed psychotherapies apply to psychoanalysis as well is of course to review comparable research on “psychoanalysis proper.” But what is psychoanalysis? Most conveniently and least controversially, it may be defined on the basis of frame factors [e.g., duration longer than 2 years, session frequency higher than twice a week, use of couch]. Because the border between psychoanalysis and psychotherapy is hard to distinguish, the following review will include, besides studies of psychoanalysis, studies of psychoanalytically oriented psychotherapy in which treatment duration exceeded 2 years and 200 sessions.

Meta-analyses

Using somewhat different inclusion criteria than Leichsenring and Rabung (2008) did in their meta-analysis, de Maat and colleagues (2009) published a meta-analysis of studies of what they considered long-term psychoanalytic therapies and psychoanalyses. Besides 9 studies that were included in Leichsenring and Rabung’s sample, de Maat et al. found 18 more studies satisfying their inclusion criteria, with a treatment duration of 1 year or longer and involving at least 50 sessions, and also satisfying certain quality criteria. Again, such treatment duration is not long from a psychoanalytic perspective, so it may be of interest to have a closer look at the sample based on a longer minimum duration for considering a treatment to at least approximate psychoanalysis. Nineteen of the 27 studies had treatments with a specified duration of 2 years or longer (and 2 more probably had), and 12 of those included a treatment labeled as “psychoanalysis,” implying 3 sessions or more per week. There were 14 studies with specified duration of at least 3 years, and 2 more possibly had a comparable duration; of these 16, 11 had a condition called “psychoanalysis.” Six of these were published after the year 2000, three from Europe and three from North America. The studies were mostly naturalistic cohort studies, and quite a few were retrospective. One, involving TFP of longer than 3 years’ duration, was a randomized study. De Maat and her colleagues (2009) found that the effect sizes were moderate to large, larger for psychoanalysis than for psychotherapy, larger at follow-up than at termination, and substantially larger for symptom reduction than for personality change. These conclusions are supported in two very recent meta-analyses [S. de Maat, J. Dekker, R. de Jonghe, et al., submitted for publication a; S. de Maat, F. de Jonghe, R. de Kraker, submitted for publication b].

Chart Reviews of Psychoanalytic Caseloads

Although some analysts have demonstrated their courage to follow up their own patients and publish their findings (Coriat 1917; Schjelderup 1955), such statistical summaries have been more convincing when the sample of patients has been selected from some independent caseload. In a careful and detailed review, Bachrach and colleagues (1991) summarized studies of such collective caseloads from psychoanalytic institutes in Berlin; London; Chicago; Topeka, Kansas; New York; and Boston. They concluded that patients in psychoanalysis derive therapeutic benefit—if they are suitable for psychoanalysis and their pretreatment level of functioning is high—although this may sound like circular reasoning.

Retrospective review of clinical case records for the purposes of treatment evaluation requires extremely stringent procedures—from the design of the initial records, to the coding and rating of the records, and finally to the analysis of these data—as well as an awareness of the methodological weaknesses that nevertheless remain. Bachrach et al. (1991) found many such weaknesses in the caseload studies. An exemplary chart review study that was more compliant with the stringent requirements outlined above is a review of almost 800 cases in child psychoanalysis or psychotherapy at the Anna Freud Centre in London (Fonagy and Target 1994, 1996; Target and Fonagy 1994a, 1994b). This review was as close to a total nonsampling study as is probably possible. Health-sickness ratings and formal diagnoses showed statistically and clinically significant improvements in about 75% of the cases, provided the treatments had been longer than 6 months. Outcome varied greatly with diagnosis. Children with emotional problems had generally better outcomes than children with disruptive acting-out problems. In general, more success was seen in the psychoanalysis cases than in the psychotherapy ones, and this was especially so in cases of more severe disturbance.

Retrospective Follow-Up Studies

True, a chart review study is retrospective in the sense that the reviewers revisit the charts. Yet, they were indeed there as the treatments took place. In contrast, a number of follow-up studies, in the United States [Erle and Goldberg 2003; Freedman et al. 1999]
al. 2005] and Germany [Dossmann et al. 1997; Keller et al. 2002], have adopted a purely retrospective design by approaching patients—or therapists or analysts—after termination to assess, in retrospect, the patients' posttreatment status as well as their status pretreatment. The model design used in these studies was the Consumer Reports Study survey [Seligman 1995]. The validity of this procedure depends on whether the retrospective assessment of the former patient is correct—that is, agrees with the assessment he or she would have made at the beginning of treatment—a dubious assumption. Also, the studies generally had serious selection problems.

The most sophisticated retrospective follow-up of psychoanalysis patients published so far was supported by the German Psychoanalytical Association [DPV] and directed by Leuzinger-Bohleber [2002a, 2002b; Beutel and Rasting 2002; Leuzinger-Bohleber et al. 2003]. The response from the DPV membership was strong and helped produce a sample of around 400 patients who had terminated a psychoanalysis or psychoanalytic psychotherapy with a DPV member during the years 1990 to 1993. Careful analysis established that the sample was representative. Probably the most significant and valid study of the DPV group was based on qualitative analyses of tape-recorded psychoanalytic follow-up interviews with almost 200 patients throughout Germany in an extremely well-organized logistic design. A particularly interesting approach was to identify a number of "ideal types" of cases on the basis of patterns of inferred change across three dimensions: self-reflection, object relations, and creativity and working ability. Eight such ideal types were defined and labeled as follows: Type I: "sound and positive"—the successful follow-up; Type II: "successful, but why?"—the unreflective success; Type III: "successful but without self-reflection or satisfying object relations"; Type IV: "creative and able to work, but still alone"; Type V: "the tragic ones—accepting their fates"; Type VI: "without success, but socially well integrated unreflective people"; Type VII: "the unsuccessful ones"; Type VIII: "the extremely traumatized" [Leuzinger-Bohleber 2002b]. As indicated by the multitude of types, the variation in outcomes was considerable and appeared to depend on the quality of the match between the personalities of the analyst/therapist and the patient, according to the researchers. No significant quantitative differences were established between the psychoanalysis and the psychotherapy cases. However, qualitative differences were suggested: among the psychoanalysis cases, the internalization of the self-analytic function had resulted in more elaborate and differentiated reflective functioning.

Real-Time Pre-Post Studies

There are a number of prospective studies adopting a so-called pre-post design, in which patients are followed from the beginning of treatment onward, with assessments made in real time before and after treatment. An ambitious research program at the Boston Psychoanalytic Institute was initiated by Kantrowitz for a prospective study of 22 cases [Kantrowitz et al. 1986, 1987a, 1987b, 1990a, 1990b, 1990c]. Using interviews, psychological tests, and carefully prescribed rating procedures before and after the analyses, and following up after 5 years as well, Kantrowitz and colleagues found that the treating analyst, at the time of termination, made a more positive evaluation of the outcome than the patient did 1 year later, whereas the psychological test battery indicated the least positive treatment results [Kantrowitz 1993]. At follow-up, slightly less than half of the patients had maintained their initial gains, and neither the patient nor the analyst or the tests were able to predict which patients would or would not maintain their gains [Kantrowitz et al. 1990a]. As in the DPV study, the match between the personalities of patient and analyst was felt to have been critical.

In a study in Göttingen, Germany, long-term psychoanalytic therapy during 3 years was offered to a diagnostically mixed sample of mostly depressed patients [Leichsenring et al. 2005]. A comprehensive questionnaire battery showed large improvements. At termination, more than 75% of the patients showed significant improvements, and 80% of the sample had significant improvements at follow-up, indicating further improvements after termination, in contrast to the finding by Kantrowitz et al. [1990a].

A collection of 17 complete tape-recorded psychoanalytic therapy during 3 years was offered to a diagnostically mixed sample of mostly depressed patients [Luborsky et al. 2001], to the research community for further studies. The cases were assessed by two independent raters early and late in the treatments, using standard outcome measures. The ratings showed significant and large mean effects, but also large variation around the mean, which offers ideal conditions for the comparative study of good and not-so-good outcome cases.

Several pre-post studies on psychoanalysis have included a comparison group with psychotherapy patients. The patients were assigned to the groups on clinical grounds or based on self-selection and thus not randomly assigned.

One of the most ambitious and extensive projects ever in the area of psychological treatments in general was launched at the Menninger Clinic in Topeka, Kan-
sas. Among the many publications of the project, there are two extensive summaries (Kernberg et al. 1972; Wallerstein 1986, 1989). Most of the 42 patients included were severely disturbed, with repeated previous treatment failures. The original plan was to compare a group in psychoanalysis and one in psychotherapy. However, the contrast gradually broke down, as several patients were later switched between the two forms. Typically, the psychoanalyses also had to be modified, in the sense that supportive measures became more salient than is assumed with “pure” psychoanalysis. The findings were based on case records, repeated interviews with patients, an extensive test battery, interviews with family members, health-sickness ratings, and a complex procedure of paired comparisons among cases on a host of variables. Global ratings suggested that almost 60% of the patients had moderate or very good improvement, and the psychoanalysis and the psychotherapy cases had about equally successful outcomes. Indeed, Wallerstein (1989) found the supportive ingredients more conducive to change than had been expected. Also unexpectedly, change was not necessarily dependent on the resolution of internal core conflicts.

A most interesting result was brought to light when Blatt (1992) assessed treatment outcomes in the Menninger Project in relation to his distinction between introjective and anaclitic patients. His categorization was based on the patients’ Rorschach protocols and showed that the anaclitic patients, preoccupied with interpersonal issues, tended to respond more positively to psychotherapy than to psychoanalysis, whereas the introjective patients, concerned with autonomy and self-definition, did better in psychoanalysis.

A series of studies in Berlin and Heidelberg, Germany, has compared groups of psychoanalytic cases with outpatients and inpatients in psychodynamic psychotherapy (Grande et al. 2006; Rudolf et al. 1994; von Rad et al. 1998). In general, the results have been quite positive both at termination and at follow-up, and especially so in the psychoanalysis cases. However, in contrast to several other studies (de Maat et al. 2009), and in contrast to the psychotherapy group, von Rad and colleagues (1998) found that symptom changes in the psychoanalysis group were not maintained at follow-up. Also, a patient satisfaction question at follow-up indicated less positive opinions among the analysands. Possibly, the demands of the psychoanalytic regime fostered higher expectations that could not be fulfilled with many patients. These demands may not be justified by the somewhat better effects on symptoms, unless the treatment will lead to more profound changes beyond symptomatology. Grande and his colleagues therefore developed a scale to assess so-called structural change; the Heidelberg Structural Change Scale (HSCS; Grande et al. 2004) was developed on the assumption that structural change will be reflected in increasing awareness and readiness to cope with, and work through, intrapsychic conflicts. Indeed, Grande et al. (2009) found that the patients’ evaluation of their treatment outcome at follow-up interviews even 3 years after termination was significantly predicted by changes in the HSCS at termination but not by corresponding measures of distress or interpersonal problems.

In another ambitious prospective, naturalistic German study, Brockmann and colleagues (2002, 2006) compared long-term (>3 years) psychoanalytic therapies with long-term (>2 years) behavior therapies, delivered by private practitioners. Although diagnostically equal, the groups differed in several respects: besides being less symptomatic initially, the psychoanalytic group tended to have more education, used less psychotropic medication, and had more often sought treatment on their own initiative. While both groups improved significantly until follow-up (at 3.5 years) in terms of symptom distress, the psychoanalytic group showed superior and continuous change until the final follow-up, after 7 years.

Longitudinal Studies

Whereas the traditional assessment design has been used to compare pretreatment, posttreatment, and, sometimes, follow-up measurements, a few studies have adopted designs to estimate continuous change trajectories. In the German TRANS-OP project, Puschner and colleagues (2007) used a clever design to construct average change trajectories in realtime during the first 2 years of treatment. This naturalistic study covered almost 500 cases of psychoanalysis or psychodynamic therapies that had been reimbursed by a private health insurance company. The sample was mixed with respect to diagnoses, with a majority of patients having affective and neurotic disorders. Of note, symptom distress showed an especially sharp decline before the first formally scheduled session. After 2 years, when about two-thirds of the patients had terminated their treatments, the psychoanalysis patients, from a more impaired level of distress, had improved somewhat faster than patients in psychodynamic therapy. Neither of these differences was statistically significant, however.

One design to study very long treatments without taking the time to follow each through from beginning to end in real-time is called the “accelerated longitudinal
design” (Bell 1953; Raudenbusch and Chan 1992). In the Stockholm Outcome of Psychoanalysis and Psychotherapy Project [STOPPP; Blomberg et al. 2001; Sandell et al. 2000], a panel of patients in psychoanalysis or psychotherapy, mostly psychodynamic, responded to an extensive questionnaire for 3 consecutive years. The sample was divided into subgroups depending on the position of each patient in terms of time in treatment, thus creating a time scale from 1 year before treatment to 3 years after termination. The psychoanalysis group had a significantly higher rate of positive change on a measure of psychological distress, from an initial level almost identical to that of the psychotherapy group. Especially interesting was that the really significant divergence appeared only after treatment termination. By 3 years after termination, the mean trajectory in the psychoanalysis group had reached close to the mean in a “normal,” nonpatient group. On the other hand, the Social Adjustment Scale, which measures the quality and “quantity” of one’s social relations, showed only modest, equal change in the two groups.

A similar design was used by a group at the Netherlands Psychoanalytic Institute in Amsterdam (Berghout and Zevalkink 2009, Berghout et al., in press, Zevalkink and Berghout 2006). On almost all clinical scales, there was significant improvement from before or early in treatment to termination, but nothing further during follow-up. The Minnesota Multiphasic Personality Inventory, assumed to reflect personality structures, showed similar changes on some, but not all, of its subscales, although these changes were consistently smaller. There were no systematic differences between the two forms of treatment on either type of measure, and rather elaborate attempts to predict individual treatment outcome had failed throughout.

**Randomized and “Quasi-Randomized” Studies**

Randomized studies of long-term psychotherapies and psychoanalyses are extremely rare. Nevertheless, a classic truly randomized comparison between high- and low-frequency long-term psychoanalytic psychotherapy was reported by Heinicke (1965; Heinicke and Ramsey-Klee 1986). The patient population was fairly specific: children with reading difficulties, 7–10 years of age. Interestingly, in the low-frequency group, the children improved at a faster rate than the children in the high-frequency group during the first year, but during the second year the four-times-a-week group caught up and surpassed the low-frequency group.

This “sleeper effect” was also found in a recent Finnish study (Knekt et al. 2008a, 2008b). In this study, 326 outpatients with predominantly mood or anxiety disorders were randomly assigned to one of three treatment groups: long-term psychodynamic psychotherapy, STPP, and solution-focused therapy, a cognitive-behavioral form. Significant reductions on depression and anxiety symptoms, as well as significantly raised levels of functioning and working ability, were noted during the 3-year follow-up. During the first year of treatment, STPP had a higher change rate than long-term psychodynamic psychotherapy, whereas during the second year of follow-up, no significant differences were found between the short-term and long-term therapies. After 3 years of follow-up, long-term psychodynamic psychotherapy was more effective, with 14%–37% lower scores on the outcome variables. No significant differences were found between the short-term therapies. A fourth, nonrandomized group of patients were offered psychoanalysis based on suitability considerations. Here, too, psychoanalysis patients were slower starters, but after 5 years of follow-up, psychoanalysis was the most effective, after a gradual and steady increase in recovery rates. According to preliminary analyses, about 80% of the patients receiving psychoanalysis recovered from their depressive symptoms, whereas the corresponding proportion for the other groups varied from 48% to 67% (Knekt et al. 2011).

In Munich, Germany, Huber and Klug (2004; Huber et al. 2007) compared psychoanalysis with psychodynamic therapy and behavior therapy in a sample of more than 100 patients diagnosed with unipolar depression. Balancing the requirements for internal and external validity, the authors used a design that included a “randomization board,” which decided, on the basis of an audiotaped intake interview, whether a patient should be randomly assigned or not. As it turned out, the board decided that all patients could be randomly assigned. Although selected on a random basis, the cognitive-behavioral group was included only later, unfortunately, and this reduced the credibility of the design. Follow-up with an extensive outcome assessment battery showed that the psychoanalysis group was superior to both psychotherapy groups on measures of relapse rate and interpersonal problems. Also, a structural type of measure, the Scales of Psychological Capacities (DeWitt et al. 1991), showed superior improvement for the psychoanalysis group (Huber et al. 2005).

Another interesting design, a so-called preference design (Brewin and Bradley 1989), is currently being used in a treatment study involving patients with chronic depression sponsored by the German Association for
Psychoanalytic Therapies [DGPT]. In this design, patients are first allowed to choose whether to be randomly assigned or not to a treatment and then, if not assigned, to choose the preferred treatment, of which psychoanalysis is one and CBT another. No results are yet published. Likewise, results are still forthcoming from a truly randomized study on moderate-length once-weekly psychoanalytic therapy with primary care patients with treatment-resistant depression in London, the Tavistock Adult Depression Study [TADS]. Also, a truly randomized trial of successful psychoanalysis with mother-infant couples is now forthcoming (Salomonsson and Sandell, in press a, b).

Studies of Effects on Physical Health and Health Care Consumption

A classic in this genre is the follow-up study by Dührssen (1962) in a group of patients registered with a public insurance company in Berlin, Germany. However, her estimation of the average number of hospital days posttreatment used a highly speculative procedure, and a biased sample was used. Dührssen and Jorswieck (1965) later reported a more convincing study in a new sample of so-called neurotic patients in long-term psychoanalytic therapy, who were compared with a corresponding group of untreated patients and a sample from the general population of insureds. Comparing the last 5 years before treatment and the first 5 years after termination, the authors found that the annual number of days in hospital decreased in the first group by almost 80%, to a level significantly below that of the “normal” population. No corresponding change occurred in the two comparison groups.

Whereas Breyer et al. (1997) analyzed patients’ self-reported retrospective levels of health care utilization as well as number of work days lost, Keller et al. (2002) reported somewhat more convincing results based on their sample of Jungian analysands. Information on number of days absent from work and days hospitalized before and after treatment, in real time, was collected through health insurance companies. Unfortunately, data were unavailable for almost half the final sample. According to the information available, 1 year after treatment, the average annual absence from work due to illness decreased by about 50%, whereas the number of days hospitalized decreased by almost 90%.

Like Keller’s (2002) group, Leuzinger-Bohleber’s was able to secure data on health care consumption from the health insurance companies. Unfortunately, as in Keller’s study, information was unavailable for a significant number of patients. From a high level of 19 days of absence the last year before treatment, there was an almost steady decline during the following years (Beutel et al. 2004).

Lazar and colleagues [2006, 2007a, 2007b], in the STOPP group, used patient reports on health care parameters like medical and psychiatric consultations, medication, inpatient weeks, and absence from work. In contrast to much previous research, they found virtually no significant changes during or after treatment and no differences between the psychoanalysis and the psychotherapy groups. This was in stark contrast to the very positive changes on self-rated measures of psychological distress and general health, and Lazar et al. (2007a) demonstrated that changes in health care utilization and changes in measures of psychological well-being were more or less unrelated. In still another study, the patient sample was divided in homogeneous subgroups based on health care consumption outcomes (Lazar et al. 2007b).

It was clear from the study that the vast majority consumed very little health care to begin with and hence had very little room for improvement, whereas smaller clusters developed in different directions, with some increasing and some decreasing their consumption.

In the study at the Netherlands Psychoanalytic Institute, Berghout and colleagues (2010b) collected self-reported figures related to health care utilization and then used them to calculate annual direct costs per case. These costs were reduced by 45% from pre- to posttreatment, and especially contributing to the reduced costs was the decreased use of outpatient mental health resources. Annual indirect costs due to absence from work were estimated to have decreased by 71% from pretreatment to follow-up. Berghout et al. (2010a) performed a cost-utility analysis and concluded that psychoanalysis was more costly but also more effective from a health-related quality-of-life perspective than psychoanalytic psychotherapy. This conclusion may be debatable—the extra cost for each additional quality-adjusted life-year (QALY) by delivering psychoanalysis instead of psychoanalytic psychotherapy was as much as 50,000 euros.

The TRANS-OP group, according to a conference report by S. Kraif, B. Puschner, and H. Kordy (personal communication, March 2002), found substantial increases in health care costs in the years before treatment start and steady but rather unpredictable reductions during the first 2 years in treatment.

A large German multicenter study of relatively brief psychodynamic therapy with patients with eating disorders, Project TR-EAT [Kächele et al. 2001], yielded
mixed results: a low success rate for anorexia patients at discharge; a considerably higher success rate at follow-up after 2.5 years; and a higher success rate for bulimia patients at discharge, but a reduced rate at follow-up. Overall, the outcomes were not very encouraging, reaching less than 40% recovery at follow-up.

In another specific patient population, children with insulin-dependent diabetes mellitus, Moran and Fonagy (1987; Moran et al. 1991) compared psychoanalytical high-intensity, though relatively brief, psychotherapy with medical treatment. Twenty-two children and adolescents were allocated to one of two clinics on the basis of their home address. In one clinic, the children were offered three or four sessions per week under inpatient conditions; in the other clinic, the children received routine medical treatment. The level of diabetic control (blood sugar concentration) was significantly improved in the psychoanalytic group and remained at the same level 1 year after discharge, in contrast to virtually no change in the comparison group. A time-series analysis of a single case showed that the presence and interpretation of psychic conflict predicted an improvement in diabetic control.

Studies on Time Factors

Several of the earlier chart reviews summarized by Bachrach et al. (1991) noted a positive association between treatment duration and outcome, and that finding is in accordance with consistent findings in psychotherapy research (Orlinsky et al. 2004), although there are exceptions (e.g., Kächele et al. 2001). Seligman (1995), in the Consumer Reports Study, as well as Freedman et al. (1999), also found a positive, albeit not very strong, association for duration, although their retrospective design does not allow convincing conclusions. However, as Howard and colleagues (1986) found, the psychotherapy dose-effect is a negatively accelerated function, implying that the marginal benefit of each additional session will gradually decrease—at least up until about 100 sessions. It may so happen, however, that there is a qualitative upward shift as the number of sessions approaches 1,000.

There has been much less research on the effect of session frequency on outcome (Orlinsky et al. 2004), although Heinicke’s study (Heinicke 1965; Heinicke and Ramsey-Klee 1986) offered a strong case for high-frequency psychotherapy. Also, of course, the superiority of psychoanalysis that has been found in several comparisons with psychotherapy may be taken as support for the value of high-frequency therapy, although there are several competing explanations for this finding. The IPTAR study (Freedman et al. 1999), specifically focusing on the frequency issue, reported a significant outcome rise when frequency was increased from one to more sessions per week.

Now, one may have to consider the possibility that the effect of increased duration, whether in terms of number of sessions or in terms of weeks or months, depends on the frequency or density of sessions, and vice versa. Surely, 150 sessions during 3 years, for instance, is a very different regimen than 150 sessions during 1 year. An analysis by Kordy and colleagues (1988) of the outcomes of 76 psychoanalyses or long-term psychotherapies indicated just that: the success of these treatments depended not only on the quantity of time, counted in sessions or years, but on the distribution of the sessions across time. Outcome increased with the total number of sessions as well as with duration in years, but when session frequency was increased, there was no clear trend in outcomes, whether positive or negative. Further, surprisingly and in contrast to clinical lore, the data suggested that a variable number of weekly sessions, as well as occasional interruptions of 3–6 weeks, was not only not adverse but even beneficial to treatment outcome.

Even more to the point, a study in the STOPP project could establish a statistical interaction between duration and session frequency, yielding both very positive changes during follow-up under high-frequency, long-duration conditions and relatively positive changes with low-frequency, short-duration ones (Sandell et al. 2002).

Limitations and Problems of Psychoanalytic Outcome Research

Even a psychoanalytic partisan will have to admit that the evidence base for psychoanalysis and long-term psychoanalytically informed therapy is limited in both quantity and quality. Thus, the number of studies is indeed small. The studies that do exist, especially earlier ones, have serious methodological shortcomings, as noted by several reviewers (Bachrach et al. 1991; Fisher and Greenberg 1985, 1996; Luborsky and Spence 1978; Meltzoff and Kornreich 1970). Two recent studies have systematically rated the quality of studies on psychodynamic therapies. De Maat et al. (2009) found that 9 of 27 reviewed fell short of their quality criterion, and Gerber et al. (2011), in a review of 96 randomized studies of psy-
Psychodynamic therapies (mostly of short or moderate length), rated the studies as of only moderate quality, at an average, in terms of the description, execution, or justification of their methods. The main problem, possibly insurmountable, is the randomization issue. Several scholars (e.g., de Maat et al. 2007; Leichsenring 2005; Seligman 1995) have pointed out that randomized controlled studies are not appropriate with long-term treatments for several reasons: practically unfeasible, unethical, and nongeneralizable. It is probably futile to expect patients to comply for years with a randomized assignment they do not appreciate, and it is nevertheless meaningless to evaluate a treatment with patients who did not want it or for whom it was not suitable in the first place. Also, what is often not recognized is that randomized designs fail to control for differences in suitability base rates in the patient population, with the consequence that the treatment condition that is the most suitable to more patients has an undue advantage. One should hope that the preference design (Brewin and Bradley 1989) or some similar “suitability design,” where patients are allocated to treatments according to their differential suitability, would gain wider acceptance.

An even more fundamental problem may have to do with disinterest, unwillingness, fear, and other emotional factors among psychoanalysts to expose their work to serious testing against some sensible well-being criterion. It is probably at least one good thing about the evidence movement, in many respects so dogmatic or fundamentalistic, that it has raised awareness among clinicians of the need for such tests.

### Study Outcomes and Their Interpretation

The studies there are, again, do show that psychoanalytically informed therapies are viable alternatives in the psychological treatment assortment. This is true for short-term therapies, especially with depressed patients and probably for patients with more complex anxiety disorders than simple phobias and patients with substance abuse. Moderate-length psychodynamic therapies have been shown to be quite effective for patients with complex and chronic psychiatric and personality disorders.

The few convincing studies there are of psychoanalysis proper suggest that psychoanalysis in general is superior to psychodynamic therapy, although exceptions exist. It is not at all clear which is the critical contrast in these comparisons, however. To what extent are the differences due to the initial selection of patients, to duration, to session frequency, and to factors of technique? First, several studies have shown that patients clinically assigned to psychoanalysis are different from patients assigned to psychotherapy in very many respects (e.g., Berghout and Zevalkink 2009; Rudolf et al. 1994; Weber et al. 1985). To the extent that these assignments are correlated with true suitability differences, it makes little sense to compare the treatments as if they were interchangeable. Second, as noted earlier, almost all studies show that increasing treatment duration, whether in terms of number of weeks or in terms of months or sessions, does influence outcome in the positive direction.

Third, in-session technique may be different in psychodynamic psychotherapy compared with psychoanalysis—and to the extent it is not, it may be non-optimal or even dysfunctional, becoming “as-if analysis,” applying psychoanalytic technique under nonpsychoanalytical conditions (Grant and Sandell 2004). On the basis of an inventory of therapeutic attitudes, Sandell and colleagues (2004) were able to identify a group of treatment providers with significant overrepresentation of persons with psychoanalytical training, also including a sizable number of psychodynamic therapists. Persons in this group devalued support, caring, and kindness in the therapeutic relation but valued neutrality and emphasized the irrationality of man and the intuitive or artistic component in the therapeutic enterprise. This pattern of attitudes was interpreted as classically, even orthodox, psychoanalytical. Whereas the outcomes of the psychoanalyses run by the analysts in this group were quite positive, on average, the outcomes in psychotherapy provided by therapists, and some analysts too, were clearly inferior to those run by therapists with less classical, more eclectic patterns of attitudes (Grant and Sandell 2004; Sandell et al. 2007). The conclusion was that the classical psychoanalytic attitude, while quite prevalent among psychodynamic therapists and analysts, is generally dysfunctional in psychotherapy. Further support for this conclusion was offered in a study in which patients with therapists and analysts with very long personal or training analyses were found to do poorly in psychotherapy (Sandell et al. 2006). The finding was interpreted in terms of the modeling function of personal or training analysis. The longer the therapist’s analysis, the stronger she or he will identify with the approach of her or his psychoanalyst, and the more likely she or he will unwittingly adopt it working with patients in psychotherapy.
Valued and Responsive Outcome Criteria

A further issue in comparisons between psychoanalysis and psychotherapy, whether psychodynamic or not, has to do with the measurement of outcome. Psychoanalysts in general show a depreciative attitude toward symptom or distress criteria of outcomes. Nevertheless, changes in levels of symptoms or distress have almost always been found to be larger than changes in measures of social relations or personality (Berghout et al., in press; de Maat et al. 2009). The specificity of psychoanalysis and the justification for its extra demands on the patient are believed to lie in the greater “depth” of its effects, and there are indeed a few systematic procedures suggested to estimate so-called structural change (DeWitt et al. 1991; Grande et al. 2006). Apart from these operational definitions, the concept has become rather a slogan, a rhetoric, or what Weinshel (1990) has called “a psychoanalytic shibboleth.” In a conceptual analysis, Sandell (2005) suggested that “mental structure” is an intervening construct to explain stability, consistency, and predictability of behavior and that change in mental structures can only be established on the basis of repeated, long-term observations. The crux of the matter of structural change is whether change is stable, creating different consistencies (restructuring), creating new ones where there were none (structuring), or breaking up maladaptive ones (destructuring). It therefore speaks for the “depth” of change in psychoanalysis that studies generally have found it stable and—very important—increasing even under extended follow-up. In a qualitative study, Falkenström and colleagues (2007) found that continued improvement after termination was significantly associated with the patient’s self-analytic function. The finding that outcome is a process, changing after termination in ways that are not always individually predictable, is another argument for the vital importance of extended follow-up in outcome research.

Variation in Outcomes Among Cases

In their eagerness to establish evidence for the efficacy of their own brands of treatment, outcome researchers are almost exclusively preoccupied with means or averages, and that has certainly been the focus in this review, too. The average is a misleading statistic, however, all too easily suggesting that it represents the true value in the population, as if the dispersion around it were occasional or random. Yet one has to realize that outcome usually varies widely and systematically in study reports, though such variation is seldom analyzed or accounted for. Yet, also, one of the great contributions of psychological science is the study of systematic individual differences. People are different, as the whole world has noticed, but outcome researchers in general go on as if the patients were replicas of one another (Sandell 2007, 2009). It was in an attempt to promote an individual differences perspective on change in psychotherapy that Lazar et al. (2007b) displayed the great diversity of health care consumption outcome.

Is Psychoanalysis “a Method”?

In view of such vast heterogeneity, one may well question whether psychoanalysis should be considered a method at all, taking “method” to mean some standard tool or procedure. As clinical discussions typically
reveal, psychoanalysts are quite diverse in looking at, and understanding, cases. Maybe one should consider psychoanalysis not as a method but as a process evolving between two persons in a fairly standardized and strict setting, where one of the persons, the psychoanalyst, thinks and acts as a “process manager” on the basis of her or his familiarity with psychoanalytic thinking and writing. And, correspondingly, “thinking psychoanalytically” within the frames of a more or less standardized protocol may be what distinguishes the psychoanalytically informed therapies, both short and moderate term. Although we do not know for sure the critical and effective ingredients (whether time, setting, relationship, technique, or whatever), evidence indicates that the end result is most often quite useful to patients, especially in the long run.

**KEY POINTS**

- Establishing effectiveness or efficacy for any kind of treatment, psychoanalysis included, requires measures of control that are more difficult to implement the longer and the more open-ended the treatment is.

- Good-quality studies of the outcome of psychoanalysis proper are still rare. Most studies are naturalistic cohort studies, with insufficient controls for confounding factors.

- The studies that do exist suggest that psychoanalysis in general is an effective form of treatment. Effects on the level of symptoms or distress are generally larger than on personality or social relations variables, and the effects tend to improve during follow-up.

- Moderate-length psychoanalytically informed therapy has been shown to be effective with patients with severe, complex, long-standing clinical conditions, at least as much as—if not more than—its nonpsychodynamic alternatives.

- Comparisons between psychodynamic psychotherapy and psychoanalysis proper have tended to favor psychoanalysis, although the interpretation of these comparisons is far from straightforward. For instance, a robust body of research has shown that positive outcome is a function of treatment duration.

- Psychoanalytically informed short-term therapy has been shown to be as effective as its nonpsychodynamic alternatives for miscellaneous clinical conditions—and especially for depression.

- A most striking, and generally neglected, fact is that outcome in psychoanalysis and psychotherapy in general, whichever the brand, varies widely and systematically across cases. To what extent this heterogeneity is due to variability among patients or therapists remains to be explored.

**References**


Research on Outcomes of Psychoanalysis and Psychoanalysis-Derived Psychotherapies


Breyer F, Heinzel R, Klein T: Kosten und Nutzen ambulante Psychoanalyse im Deutschland [Cost and benefit of outpatient psychoanalytical psychotherapy in Germany]. Gesundheitsökonomie und Qualitätsmanagement 2:59–73, 1997


