Psychotherapy Research 11(4) 361–382, 2001

LONG-TERM OUTCOME OF LONG-TERM PSYCHOANALYTICALLY ORIENTED THERAPIES: FIRST FINDINGS OF THE STOCKHOLM OUTCOME OF PSYCHOTHERAPY AND PSYCHOANALYSIS STUDY

Johan Blomberg
Department of Behavioural Sciences, Linköping University
Anna Lazar
Department of Behavioural Sciences, Linköping University and Stockholm County Council Institute of Psychotherapy
Rolf Sandell
Department of Behavioural Sciences, Linköping University

The scarcity of findings on the effectiveness of psychodynamic long-term psychotherapy and psychoanalysis is a problem, at least in countries where these treatments are the general treatment of choice. In this study, a sample of 405 patients in various stages before, during, and after treatment was drawn from a population of 1,200 patients who had received or were on a waiting list for public subsidized long-term psychoanalytically oriented psychotherapy or psychoanalysis. The average duration was 40 months for the psychotherapies and 51 months for the psychoanalyses. All patients completed a questionnaire, including the Symptom Checklist-90 (SCL-90), Sense of Coherence Scale (SOCS), and Social Adjustment Scale (SAS), on 3 occasions with 12-month intervals. Using a partly within- and between-subjects design, group means were regressed on an ordinal time scale covering 7 stages in the treatment process: before, early during, during, late during, soon after, after, and late after treatment. The slopes indicated small to moderate effect sizes (d = 0.4–0.6) for psychotherapy and moderate to very large effect sizes (d = 0.4–1.5) for psychoanalysis. The largest effect sizes were on the SCL-90 and the lowest on the SAS for both groups. The proportions of patients with more "healthy" scores compared with the worst scoring 10% in a nonclinical norm group increased from roughly 30% to 55% in the psychotherapy group and from 10% to 75% in the psychoanalysis group. The results underscore the importance of long-term follow-up when evaluation of long-term treatments is concerned.

Correspondence concerning this article should be sent to Johan Blomberg, Heleneborgsgatan 7C, 11731 Stockholm, Sweden. E-mail: johan.blomberg@provinsfem.se.

Background

The Status of Outcome Research on Long-term Treatments

Psychoanalysis has now endured almost 100 years of heated intellectual discussion regarding its clinical and scientific value. Despite its indisputable clinical and theoretical impact, there is a peculiar lack of systematic empirical research on its clinical value. With few exceptions, research on outcome of psychoanalysis has come from anecdotal case studies (see Bachrach, Galatzer-Levy, Skolnikoff, & Waldron, 1991; Doidge, 1997; Fonagy, 1999; Kantrowitz, 1997, for reviews and Fisher & Greenberg, 1996, for a methodological critique). There is, indeed, a general skepticism among psychoanalysts toward the value of systematic empirical outcome research. In fact, some influential psychoanalysts still refute it altogether by arguing that it is incompatible with the basic psychoanalytical tenets (Green, 1996). Furthermore, among "mainstream" psychotherapy researchers, there has been equally weak interest in evaluating long-term psychological treatments. In Grawe, Donati, and Bernauer's (1994) review of 796 studies on psychotherapy, no more than 8 (1%) focused on therapies lasting more than 2 years, and 5 of these concerned so-called psychodynamically oriented psychotherapies, including group therapies and therapies with psychotic patients.

One reason for this underrepresentation of long-term therapies in psychotherapy research may be that, by definition, research on long-term treatment takes time and may not pay off careerwise (e.g., in terms of publication rates). Also, time tends to exacerbate ethical, practical, financial, and, most of all, methodological problems, making so-called gold standard research extremely difficult to implement (Sandell, Blomberg, & Lazar, 1997; for an extensive discussion see also VandenBos, 1996).

Another important issue of time has to do with follow-up. In a meta-analysis by Shapiro, Harper, Startup, Reynolds, and Suokas (1994) of 33 process outcome studies, the mean duration of follow-up was 1.9 months, but this figure was heavily influenced by a small number of long-term follow-ups: More than 90% of the studies had no follow-up at all. The same pattern could be seen in Smith, Glass, and Miller's (1980) review: 33% of the studies measured effects immediately after termination; only 2.5% had follow-up intervals of 1 year or more. In the Grawe et al. (1996) review, 37% of the 796 studies had no follow-up at all and another 53% had follow-up after less than 1 year. Whereas Nicholson and Berman (1983) found no relationship between follow-up time and outcome (and consequently concluded that follow-up is not necessary), Smith et al. (1980) found a negative relation between effect size and follow-up until 2 years after termination. Kazdin (1994) argued that effects may vary across follow-up time, and Shapiro et al. (1995) reported that outcome difference among therapy modes and duration schedules changed with length of follow-up from 3 months to 1 year. Assuming that outcome is a process rather than a stable state, and that different therapeutic orientations may promote different processes, the vicissitudes of the outcome process become an issue of great importance from both a clinical and a theoretical point of view. Therefore, from a methodological perspective, long-term follow-up becomes mandatory.

We conclude that credible research on long-term effects of long-term psychotherapies and psychoanalyses is almost nonexistent. Thus, systematic outcome research so far has little to offer in terms of guidance for psychotherapeutic practice and decision making, especially in the Nordic countries, where long-term psychotherapy is generally the treatment of choice both from patients' and therapists' point of view.

Aims and the Field Conditions of the STOPP Project

In 1988 the Swedish health care authorities launched a project to subsidize psychoanalysis and long-term psychotherapy with private practitioners. The main purpose was to reduce sick leave and increase general well-being among persons in need of psychiatric treatment. On the basis of a referral from a licensed therapist, a program supervisor decided on the subsidization of an analysis or a therapy. The subsidies were limited to 3 years, but treatment itself was not: Patients were free to apply even if they were in ongoing therapy and were also free to continue financing it in other ways after expiration of the subsidy. Between 1990 and 1993, a waiting list of nearly 1,200 patients accumulated. In 1993, a grant was approved for this follow-up study. The background, general design, and methodological challenges encountered in this study have previously been described in detail by Sandell et al. (1997).

Our main question was whether it was possible to discern any beneficial effects of the treatments patients underwent. Specifically, our purpose was to study (a) whether long-term psychoanalytically oriented psychotherapy and psychoanalysis had any beneficial effects on patients' subjective well-being in terms of symptoms, social relations, and morale and (b) how the effects of these treatments developed over time after termination.

Method

Design and Procedures

Our design was quasi-experimental, partly cross-sectional, and partly longitudinal. It was based on a three-wave panel questionnaire survey under "Caucus race" conditions (Carroll, 1865/1981).[1] It turned out that patients did not behave as expected: Most patients had already started their treatment at the time of referral. Consequently, the treatment status in each of the individual participant's wave was uncontrolled; some patients had started treatment, some were waiting to start, and some had already terminated. In analyzing the observations in the panel, we unfolded, so to speak, the panel along a time scale, distributing the panel members on the time scale in accordance with their treatment status for each wave, thus creating an "unfolded panel" design. Because we realize that our design is quite complex, we describe it rather meticulously in terms of the following five steps.

First, in 1994, we selected a sample of 756 persons from the total of 1,200 who had been referred to the project. The decision not to use all 1,200 was an economic one. We selected participants so as to ensure that the sample consisted of people

[1]The Caucus race during Alice's adventures in Wonderland was arranged by the Dodo as that totally uncontrolled race in which "*Everyone* has won, and *all* must have prizes": "Alice had wept violently, creating a pool of tears, crowded with birds and animals that had fallen into it. The Dodo then suggested a Caucus race to get them dry and declared that the best way to explain it was to do it: "First it marked out a race-course, in a sort of circle (the exact shape doesn't matter, it said), and then all the party were placed along the course, here and there. There was no 'one, two, three, and away,' but they began running when they liked, and left off when they liked, so that it was not easy to know when the race was over. However, when they had been running half an hour or so, and were quite dry again, the Dodo suddenly called out 'The race is over!' and they all crowded round it, panting, and asking 'But who has won?'" (Carroll, 1865/1982, p. 33).

who had terminated their treatments as well as those who were in the midst of it or had not yet started. Thus, we selected all persons (n = 202) who had been subsidized for the periods 1991 to 1993 and 1992 to 1994. To obtain as many patients as possible who had terminated treatment, we also selected the first 554 persons on the waiting list for subsidization. It turned out that most patients were not passively waiting but were already in psychotherapy or psychoanalysis at the time of the referral. Consequently, because it was treatment itself, not subsidization, that was our focus of interest, we made no distinction between those who had received subsidy and those who were on the waiting list for subsidy in 1994. (In addition, the timing of the subsidy differed between patients: Some started treatment when they received subsidy, others received it after a while, and still others did not receive any subsidy at all.)

Second, the Well-being Questionnaire (WbQ; see later discussion) was distributed in 1994 to these 756 persons and, again, in 1995 and 1996 to all who had responded the previous year. During each of these years participants were given reminders to return the Questionnaire. Returns from 78%, 86%, and 89%, respectively, for each year produced a panel of 450 people, which was 60% of the initial sample of 756. Twenty individuals with incomplete or inconsistent data on the WbQ and 12 who never started treatment were eliminated, resulting in a sample of 418 persons who had returned complete data all three panel waves. An analysis of the attrition (two-tailed t tests and chi-squares), with response/no response as the dependent variable and social, demographic, and diagnostic variables as independent variables, showed that patients with a higher education level and higher current level of functioning were more likely to respond and stay in the study (ps < .05). Also, the pattern of attrition did not differ between patients in psychotherapy and those in psychoanalysis.

Third, the 418 persons were then divided into six subgroups on the basis of their treatment status at each panel wave (Figure 1). Group A had not started treatment in the first (1994) or in the second (1995) panel wave but was in treatment in the third wave (1996; i.e., patients started treatment at some point between Wave 2 and Wave 3). Group B had not started treatment at the first wave (1994) but did so in 1995 and was still in treatment in 1996 (i.e., patients started treatment at some point between Wave 1 and Wave 2). Group C had started treatment before the first wave (1994) and was in ongoing treatment all 3 years (hence, we do not know when they terminated treatment). Group D was in treatment the first (1995) and second (1995) waves but had terminated at the third wave (1996). Group E was in treatment the first wave (1994) but not in during the second (1995) and third (1996) waves. Group F had finished treatment already at the first wave (1994).

We considered Group F as having reached a later stage in the treatment process than the rest of the groups. Similarly, we considered Group A as being at an earlier stage than the others. Correspondingly, we could order the rest of the six subgroups along an ordinal or relative time scale, defined by the relations "before" and "after" or "earlier than" and "later than." In this way, each position on the scale is later than or after all positions to the left and earlier than or before all positions to the right. When treatment durations were recorded, up to the third panel wave, the mean durations, as given in Figure 1, differed significantly as expected between the Groups A to F, $F(5, 394) = 19.92$, $p = .000$. There was no difference between patients in the two main treatment modalities, psychotherapy or psychoanalysis, $F(1, 394) = 2.29$, $p = .13$, nor was there any interaction, $F(4, 394) = 0.73$, $p = .57$.

Fourth, we aligned the three panel waves in each of these subgroups on the assumption that the last panel wave before treatment was at the same relative time,

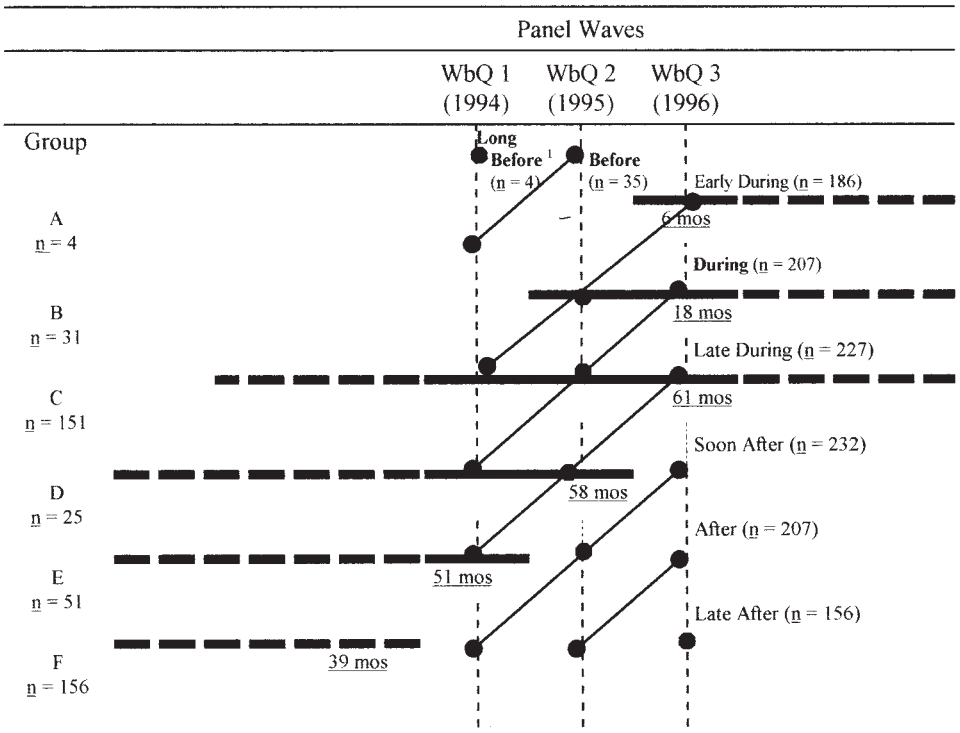

FIGURE 1. The unfolded panel design. Horizontal bars indicate approximate average durations of treatment in each subgroup up to Well-being Questionnaire (WbQ) 1996; solid bars indicate approximate average durations during the panel; dashed bars indicate approximate average durations before the panel. The diagonals and their nodes indicate which observations, in which waves and in which groups, were pooled to represent the respective steps on the relative time scale. The steps are labeled as indicated to the right of the rightmost nodes. The *n*s to the left refer to number of individuals in each subgroup and the *n*s to the right the number of observations in each step on the time scale.

whether it was the 1995 wave (Group A) or the 1994 wave (Group B), and, correspondingly, that the first wave in treatment was at the same time, relative to treatment, whether it was in 1996 (Group A), in 1995 (Group B), or in the 1996 wave (Group C). In this way, we were able to produce a sub-grouping of all 1,254 observations (three waves × 418 persons) according to the diagonal lines in Figure 1. In effect, we were able to position each observation in one of the following eight stages on a relative time scale:

1. Long before treatment[2]: observations from Wave 1 with patients in Group A (four observations).

[2]In principle, we had indeed three stages before treatment, one with a number of observations on patients who never started any treatment at all and one with observations only from 4 patients (Wave 1, Group A). These two groups were, therefore, discarded, the first since it was highly unrepresentative for the sample, the latter on account of its small number of participants. However, the observations in Group A second and third waves were used to estimate means in the next two steps on the time scale.

2. Before treatment: observations from Wave 1 with patients in Group B and from Wave 2 in Group A patients (35 observations).
3. Early during treatment: observations from Wave 1 in Group C, Wave 2 in Group B, and Wave 3 in Group A (186 observations).
4. During treatment: observations from Wave 1 in Group D, Wave 2 in Group C, and Wave 3 in Group B (207 observations).
5. Late during treatment: observations from Wave 1 in Group E, Wave 2 in Group D, and Wave 3 in Group C (227 observations).
6. Soon after treatment: observations from Wave 1 in Group F, Wave 2 in Group E, and Wave 3 in Group D (232 observations).
7. After treatment: observations from Wave 2 in Group F and from Wave 3 in Group E (207 observations)
8. Late after treatment: observations from Wave 3 in Group F (156 observations)

By pooling observations from different groups in different waves who were at the same relative stage of treatment, we had 1,254 observations distributed along eight steps on a relative time scale. In each step, the observations were partly from the same people as the observations in the steps before and after (i.e., Wave 2, in which cases in the steps were 1 year apart), partly from individuals without prior observations (i.e., Wave 1, in which cases they contributed observations for the following two steps/years as well), and partly from patients without subsequent observations (i.e., Wave 3, in which cases they contributed observations for the previous two steps/years as well). Thus, the variation across time was partly within subjects and partly between subjects, with complete independence (only between subjects) three steps apart (e.g., between Steps 8 and 5, 7 and 4, 6 and 3, and so on). Thus, for instance, the differences between steps before treatment or early during treatment and after treatment or later after are completely between subjects, respectively.

Having tested the correlations of our time scale with more than 30 different variables (patient and therapist characteristics), we have concluded that the distribution of the observations along the time scale is independent of obvious confounds.

Fifth, to compare treatments under different conditions, the observations in each step of the time scale were quasi-experimentally split into subgroups. For example, of the 418 persons in the panel, 331 had long-term psychotherapy as their only treatment or, in case they had been in more than one, their main treatment in terms of number of sessions. They accordingly generated 993 observations spread over different stages of psychotherapy. Correspondingly, 74 individuals had psychoanalysis as their main treatment (thus, 222 observations spread out along the time scale) and 13 patients had various kinds of low-dose treatments (39 observations in low-frequency individual therapy, group therapy, family therapy and so on). The latter observations were too few to yield stable statistical estimates, however, and are excluded. Also, as noted before, the $n = 4$ group before therapy was discarded because of its small size. This leaves 1,211 observations in all.

In summary, using the Caucus race analogy from Alice in Wonderland, we had another kind of race with three intermediate checkpoints (at which point we administered the WbQ) with 1-year intervals. At each checkpoint, we observed patients' well-being (just as the Dodo might have checked the animals' pulse if he'd been interested enough). Then we grouped each of the 1,211 checkpoint observations (331 therapy patients and 74 analysands times three checkpoints or observations each minus four observations before therapy) according to where the patient was in his or her race/treatment at the time: before, during, or after treatment and, more spe-

cifically, in various subdivisions of during and after treatment. Thus, we were able to position each checkpoint observation in one of seven stages on a relative time scale. Technically speaking, we had a quasi-experimental design that is partly cross-sectional (e.g., across stages of treatment; across treatment modalities; across treatment provider characteristics) and partly longitudinal (across three successive stages of treatment).

Assessment Procedures

Patients' pretreatment status. Because many patients were already in treatment at the time of the first panel wave, pretreatment status could not be assessed in any way other than on the basis of the referrals. To complicate matters further, diagnoses had to be made retrospectively for those patients who had already begun treatment. Therefore, more precise diagnostics were not feasible. On the assumption that something is better than nothing, each patient was grossly diagnosed by research assistants (licensed psychologists) for the presence of *Diagnostic and Statistical Manual of Mental Disorders* (fourth edition; *DSM-IV*) Axis I or II disorders (American Psychiatric Association, 1994) before treatment. Patients were also assessed on the Global Assessment of Functioning Scale (GAF; *DSM-IV,* Axis V). In addition to the current state, a rating was made of the lowest level of functioning after age 18 years. Ratings were also made on a specially designed scale of vocational impairment, the Vocational Impairment Scale (VIS), a 5-point scale with ratings ranging from 0 (*no impairment*) to 5 (*totally unable to work for more than 1 year*). Despite the difficulties with chart-based diagnostics, interrater reliabilities, tested with three judges on 20 referrals, were acceptable to very good: Intraclass correlations (ICC) = .69 for Axis I diagnoses, .51 for Axis II diagnoses, .69 and .88 for both current and lowest GAF, respectively, and .80 for the VIS.

Patient outcome measures. The WbQ was designed to explore the patients' symptoms, social relations, and morale (Frank, 1974). Several standard self-rating scales were included. The SCL-90 (Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974; Derogatis & Lazarus, 1994) contains 90 items representing various psychological and somatic signs of distress. The patient rates the extent to which he or she has been troubled by each during the last 7 days. The scales contain five points with ratings ranging from 0 (not at all) to 4 (very much). The ratings are scored in various combinations. There are ten subscales: Depression, Anxiety, Obsessive–Compulsive, Psychoticism, Phobic Anxiety, Hostility, Paranoid Ideation, Interpersonal Sensitivity, Somaticization, and Additional, which has no decisive meaning. The General Symptom Index (GSI) is calculated as the mean rating across all 90 items. On the basis of a suggestion from Karterud et al. (1995), a Personality Severity Index (PSI) was computed as the mean of the Hostility, Paranoid Ideation, and Interpersonal Sensitivity subscales. A corresponding Symptom Severity Index (SSI) was formed as the mean of the remaining scales (Obsessive–Compulsive, Depression, Anxiety, Phobic Anxiety, Psychoticism, and Additional). Internal consistency was .92 for the PSI, .93 for the SSI, and .97 for the GSI.

The Social Adjustment Scale (SAS; Weissman & Bothwell, 1976; Weissman, Prusoff, Thompson, Harding, & Myers, 1978) contains 39 items divided in six sections: Work, Friends and Leisure Time, Relatives (parents, siblings, and brothers- and sisters-in-law), Partner, Children, and Family (partner and children). The original items were translated and revised to suit Swedish users in the 1990s. The task is to rate to what

extent, during the last 2 weeks, various kinds of contacts have been satisfying or unsatisfying. One of the important revisions was to introduce the same 5-point rating scale for all items, ranging from *every or almost every day (or every or almost every time)* to *not one day (or never/not once)*. A general score is computed as the mean across all items, and six subscale scores may be computed as the mean across all items in each of the sections. The internal consistency (i.e., α) was .87 for the total scale.

The Sense of Coherence Scale (SOCS; Antonovsky, 1987) is a 29-item self-rating instrument designed to measure sense of coherence, a pervasive, enduring, and dynamic feeling of confidence that life is manageable, comprehensible, and meaningful. Manageability refers to the feeling that sufficient resources are at one's disposal for meeting internal and external stimuli, a feeling that the world makes sense, and that information about the environment is structured, ordered, and consistent. Finally, meaningfulness refers to the feeling that different areas of life are worthy of emotional investment. The items are questions or phrases about life experiences, and the rating scales are seven-step bipolar scales, the poles of which are alternative responses to the item. A general score is calculated as the mean across all items (α = .92). On the basis of factor analyses (Sandell, Blomberg, & Lazar, 1998), three subscale scores were also calculated: Joy (α = .91), Harmony (α = .81), and Trust (α = .71). Joy refers to a feeling of vitality and zest for life. Harmony was originally interpreted as an indication of splitting but has been reinterpreted as the experience of a conflict-free state. Trust indicates a feeling of confidence in the benevolence and loyalty of others.

Besides these instruments, the WbQ contained the following sections, with standard items/questions on (a) demographic and socioeconomic, familial, vocational, and financial situation; (b) ongoing psychotherapy; (c) previous treatments (including psychotherapy) for psychological distress; (d) current health status and health care utilization in the past 12 months; (e) current and prior severity of psychological problems; (f) occupational activities (including studies) in the past 12 months.

Therapist characteristics. In the fall of 1995, the Therapeutic Identity (ThId) was distributed to all 294 therapists and analysts involved in treatment with the patients in this sample. After four reminders, 209 (71%) had returned their questionnaires. Analyses of the attrition showed no systematic patterns. The ThId had been standardized on a random sample of 350 licensed psychotherapists throughout Sweden, of whom 227 had responded. In the national sample, there was significant nonresponse from therapists in the higher age categories who claimed that they had retired from work.

The ThId contains about 150 questions or items divided into six sections. Sections a (basic education and professional training), b (professional experience), and c (personal therapy or training analysis) were partly designed for the project and partly adopted from a questionnaire used in an international network (Orlinsky, Ambühl, et al., 1999; Orlinsky, Rønnestad, 1999). Section d of the ThId included a set of six scales to rate one's allegiance to each of the major schools of psychotherapy; ratings ranged from 0 (*not at all*) to 6 (*very strong*). Section e of the ThId had two sets of items. One set contained 33 items rating one's belief in the curative value of each of a number of ingredients of psychotherapy ("What do you believe contributes to long-term and stable change in psychotherapy?" with items like "Helping the patient avoid anxiety-provoking situations"). The second set had another 31 items to describe one's own manner of conducting psychotherapy, in the general case ("What

are you like as a therapist?"). In both sets, the items were rated on 5-point Likert scales from 0 (*not at all*) to 4 (*a lot*). Both sets of scales contained items designed by Sundland (1977) and items of our own design (e.g., "I do not answer personal questions from the patient"). Finally, Section f contained a series of 16 items relating to one's more basic assumptions about the nature of psychotherapy and the nature of the human mind ("What are your general beliefs about the human mind and about psychotherapy?"). The items were partly inspired by Hjelle and Ziegler (1981). The ratings scales were seven-step bipolar scales, with each pole offering a completion of the item stem (e.g., "Psychotherapy may be described . . . as a science/as a form or art").

On the basis of factor analyses using data from the national sample, the items in Sections e and f had been combined into nine subscales: three to summarize the therapist's beliefs in the curative factors (Adjustment, Kindness, Insight), three to characterize the therapist's ideas about his or her own therapeutic style (Supportiveness, Self-Disclosure, Neutrality), and three to indicate basic beliefs (Therapy as a Form of Art [Art], Developmental and Epistemiological Pessimism [Pessimism], Man as an Irrational Being [Irrationality]). These scales are more fully described in other publications (Blomberg & Sandell, in press). We call this set of nationally standardized scales the Therapeutic Attitudes Scales (TASC).

Treatment Characteristics

In the referrals, psychotherapy was defined as once- or twice-a-week treatment with a licensed psychotherapist and psychoanalysis as three- to five-times-a-week treatment with a fully trained psychoanalyst or a member of any of the two psychoanalytic societies in Sweden. All therapists and analysts were certified by the National Board for Health and Welfare. Because all treatments were done in a naturalistic setting, they were not manualized or standardized with respect to duration, session frequency, technique, and so on.

The mean duration of the terminated treatment was 40 months for the psychotherapy cases (SD = 21) and 51 months for the psychoanalyses (SD = 18), both adjusted with respect to regular time-outs. The mean session frequency per week in psychotherapy was 1.4 (SD = 0.5) and in psychoanalysis 3.5 (SD = 0.7). As already noted, the mean durations of all treatments did not differ significantly between psychotherapy and psychoanalysis, $F(1, 394) = 2.29$, $p = .13$, nor was there any interaction, $F(4, 394) = 0.73$, $p = .57$. The average dose for all treatments in our sample, including those who were not yet terminated at the third wave, was 650 (SD = 265) for psychoanalysis and 240 (SD = 150) for psychotherapy ($p < .001$, two-tailed t test).[3]

Without a protocol, further specification of the treatments had to be ex post facto in terms of provider characteristics from information in the ThId. Thus, all treatment providers claimed to have a psychoanalytic or psychodynamic theoretical orientation. The majority (75%) were women, and their mean age was 56 years. Number of years of therapeutic experience was high (M = 19.5 years, SD = 5.0 years), and the median number of patients seen in psychotherapy during these years exceeded 50.

[3]Dose was calculated as the number of years × 44 weeks (the average number of weeks therapists and analysts work in Sweden) × number of sessions/week.

Comparing the psychoanalysis cases and the psychotherapy cases on the basis of provider characteristics, the psychoanalyses were more often provided by someone who was more experienced (a few years older, longer experience with psychotherapy after licensing or graduation, longer experience of doing psychotherapy in private practice, more frequent formal training as supervisor, more frequent active offering of supervision during the last year).

Considering therapeutic values and attitudes on the basis of the TASC, and assuming that the therapeutic attitudes are reflected in actual practice, the treatment providers in general highly valued insight as a curative factor (e.g., "Helping the patient see the connection between his or her problems and childhood") but not adjustment (e.g., "Helping the patient adjust to prevailing social conditions"). When therapeutic style and technical factors are concerned, they valued neutrality highly (e.g., "I keep my personal opinions and circumstances completely outside the therapy") but not self-disclosure (e.g., "I always communicate the therapeutic goals to the patient in the beginning of a therapy").

Comparing the psychotherapy and psychoanalysis cases, the profile of the psychotherapy providers across these scales was significantly more similar to a group of therapists with a more behavioral or cognitive orientation in the standardization sample than were the psychoanalysis providers. The *t* tests revealed that they relied significantly more strongly on curative factors like adjustment and kindness (e.g., "Being warm and kind to the patient"). Further, they described their therapeutic style as significantly higher on Self-Disclosure and lower on Neutrality (all $ps < .05$).

Patient Characteristics

The typical patient was a woman, single, divorced or unmarried, with children. The majority (87%) had at least some university education and typically worked in health care, education, or the social sector. The mean age was 36.4 years ($SD = 8.1$ years).

Table 1 shows a sociodemographic breakdown of the patients in the two treatment groups. As can be seen, there were relatively more men in the psychoanalysis group. Analysands were also somewhat older, had higher education, and were more often married or divorced. There were no (or only minor) differences between the two groups with respect to *DSM-IV* diagnoses. However, analysands had more prior experiences with psychiatric treatments in general and with psychotherapy in particular, whereas psychotherapy patients had more often been hospitalized.

Norm Groups

To establish a standard for evaluating patient outcome in relation to the nonclinical norm group, the WbQ was also distributed to (a) a random community sample of 400 persons between 20 and 69 years of age in Stockholm County and (b) a sample of 250 psychology students demographically very similar to the clinical sample, according to pilot analyses of the referrals. The norm groups took the questionnaire only once, in May 1994. Without any reminders, the response rates in the two groups were 37% and 79%, respectively. Analysis of the attrition (Lazar, 1995) showed that both groups were representative samples of their respective populations. Also, the respondents in the two groups had almost identical mean values and standard deviations on all three outcome scales and were, therefore, treated as one group.

TABLE 1. Sociodemographic and Diagnostic Breakdown of Psychotherapy Patients and Analysands

Variable	Patient *n*	Analysands *n*	*p*
Sociodemographic characteristics			
Men (%)	20	37	**
Mean age (years)	35.6	40.0	**
Married or divorced (%)	38	60	**
Has children (%)	49	66	*
College education (%)	76	94	**
Cohabiting (%)	45	47	
DSM-IV categories			
Had an Axis I diagnosis (%)	58	54	
Had an Axis II diagnosis (%)	12	11	
V code (%)	33	36	
Mean level of functioning			
GAF, current	59.5	61.1	*
GAF, lowest	51.9	53.7	
VIS	1.3	1.2	
Previous treatments for psychiatric problems			
Any at all (%)	79	91	*
Psychotherapy (%)	63	75	*
Drugs (%)	55	56	
Outpatient (%)	56	45	
Psychiatric emergency room (%)	38	26	
Hospitalized/inpatient (%)	21	10	*

Note. Statistical chi-square tests for categorical data and two-tailed *t* tests were used for age and level of functioning. *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders* (fourth edition); GAF = Global Assessment of Functioning Scale; VIS = Vocational Impairment Scale.
*$p < .05$. **$p < .01$.

Results

In the following analyses, self-ratings were first averaged across all respondents at the same stage on the time scale and in the same treatment modality. As already noted, the first group on the time scale, early before treatment, with only 4 respondents was discarded. Then, the means were plotted across time as growth or decay curves. The change rate in each group was then estimated by the linear trend (unstandardized *b*) found when the means were regressed on the time scale. For the sake of convenience, time was assumed to be an equidistant scale. The regression parameters are presented in Table 2.

Symptoms

Figure 2 shows the development on the GSI of the SCL-90. For comparison purposes, the two horizontal reference lines indicate the mean in the norm group and the dividing line between the 90% best scoring and the 10% worst scoring persons in the norm group. The latter division corresponds to 1.28 standard deviations above

TABLE 2. Linear Trend Intercepts and Slopes (Unstandardized *b*) Across Time in the Two Treatments

	Long-term psychotherapy					Psychoanalysis					Differences[a]	
	Intercept[b]			Slope		Intercept[b]			Slope		Intercepts	Slopes
Scale	*M*	*SE*	*b*	*SE*	*p*	*M*	*SE*	*b*	*SE*	*p*	*p*	*p*
SCL-90; GSI	1.08	.032	–.071	.009	.000	1.10	.028	–.121	.008	.000	> .20	< .01
PSI	1.00	.034	–.060	.009	.001	1.18	.042	–.146	.012	.000	< .01	< .01
SSI	1.08	.037	–.072	.010	.001	1.07	.031	–.117	.009	.000	> .20	< .01
SOC total	4.10	.071	.078	.020	.011	4.22	.065	.150	.018	.000	> .20	< .05
Joy	4.41	.058	.100	.016	.002	4.57	.080	.162	.022	.001	> .20	< .05
Harmony	4.08	.177	.051	.049	.350	3.81	.160	.225	.044	.004	> .20	< .05
Trust	4.08	.074	.096	.021	.005	4.45	.096	.102	.026	.012	< .01	> .20
SAS total	2.35	.040	–.058	.011	.004	2.17	.052	–.055	.014	.012	< .05	> .20
Work	2.00	.082	–.034	.023	.200	1.97	.029	–.087	.008	.000	> .20	< .05
Friends	1.67	.042	–.001	.012	.920	1.68	.052	–.025	.014	.139	> .20	> .20
Extended family	2.24	.021	–.048	.006	.000	2.02	.092	–.000	.025	.990	> .20	> .20
Partner	2.23	.057	–.054	.016	.019	2.22	.120	–.083	.033	.056	> .20	> .20
Children	2.00	.117	–.028	.033	.437	2.06	.196	–.076	.054	.219	> .20	> .20
Family	2.51	.064	–.098	.018	.003	2.36	.088	–.101	.025	.009	.20	> .20

[a]Between groups.

[b]All intercepts were significantly different from 0 ($p < .001$).

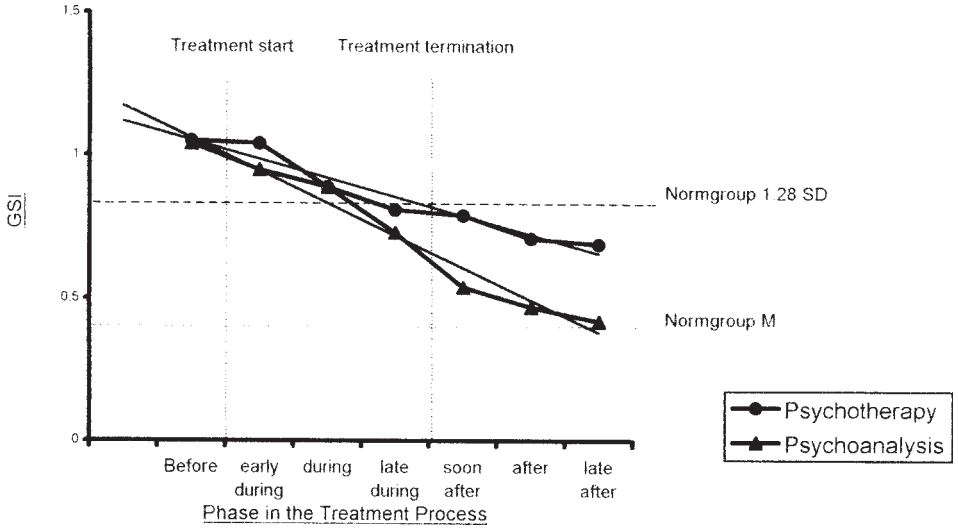

FIGURE 2. Symptom Checklist-90 (GSI = General Symptom Index; n = 1,211): decay curves and best linear fits for patients in psychotherapy and psychoanalysis in relation to the norm group (lower horizontal line) mean and the "caseness" criterion (norm group SD = 1.28; upper horizontal line).

the mean in a normal distribution and has been suggested by Derogatis and Lazarus (1994) as the most appropriate "caseness criterion" for the SCL-90. As can be seen, the psychotherapy and the psychoanalysis groups started off at almost exactly the same level, far above the caseness criterion, and then followed similar paths during treatment. From termination of treatment and beyond, the groups diverged rather sharply. The linear trends across the entire time span are indicated in Figure 2 by the straight sloped lines, the intercepts and b coefficients of which were both significantly different from zero in both groups, $t(6) > -8.07$, $p < .01$. Also, the difference between the slopes in the two groups is significant, $t(12) = 4.08$, $p < .01$. The effect size, computed on the differences between the before treatment and the late-after treatment observations, were 1.55 in the psychoanalysis group and 0.58 in the psychotherapy group.

Analyzing the SCL-90 subscales, we found no significant pattern, except for a floor effect. Thus, in both groups, there was a very strong negative correlation (–.92) between intercepts and slopes across subscales, such that scales with high initial values (e.g., Depression) had steeper (more negative) slopes than subscales with low initial values (e.g., Phobic Anxiety). Also, there was a very strong correlation (r = .96) across subscales between the intercepts in the two treatment groups and a strong correlation between the slopes (r = .79). On both the PSI and SSI, as shown in Table 2, the psychoanalysis group improved at a higher rate. Also, there was a tendency for more improvement on the PSI than on the SSI in the psychoanalysis group, $t(12) = 1.93$, $p < .10$, two-tailed, whereas there was no such difference in the psychotherapy group. The PSI intercept was also significantly higher in the psychoanalysis group than in the psychotherapy group, $t(12) = 3.15$, $p < .01$, two-tailed.

Morale

Figure 3 shows the growth curves on the SOCS. Again, both groups started off at approximately the same level. The psychotherapy group deteriorated at the beginning of treatment and then improved gradually, and the same pattern, although weaker, was found in the psychoanalysis group. Again, there was an increasing differentiation after treatment termination. Although the trends in both groups had an obvious quadratic component, the linear regression has been estimated for the sake of convenience. Both slopes were significantly different from zero, $t(6) > 3.93$, $p < .01$. Again, the difference between the two slopes was significant, $t(12) = 2.67$, $p < .05$. The effect sizes, again computed on the differences between the before-treatment and the late-after-treatment groups, were 1.18 in the psychoanalysis group and 0.40 in the psychotherapy group.

Analyses of the SOCS components (meaningfulness, comprehensibility, and manageability) failed to show any significant deviations from the total score pattern. However, the subscales suggested by Sandell et al.'s (1998) factor analyses revealed important differences among them, as shown in Table 2. The Trust subscale developed with equal, moderate improvements in the two treatment groups but at a significantly higher level throughout in the psychoanalysis group, $t(12) = 3.05$, $p = .01$, for the intercepts difference. In contrast, the Harmony subscale differed sharply between the groups, with a nonsignificant slope in the psychotherapy group and a strong slope in the psychoanalysis group, $t(12) = 2.64$, $p < .05$, for the slopes difference. The Joy subscale occupied an intermediate position in these respects, with small to moderate improvement in the psychotherapy group and significantly greater improvement in the psychoanalysis group for the slopes difference, $t(12) = 2.30$, $p < .05$.

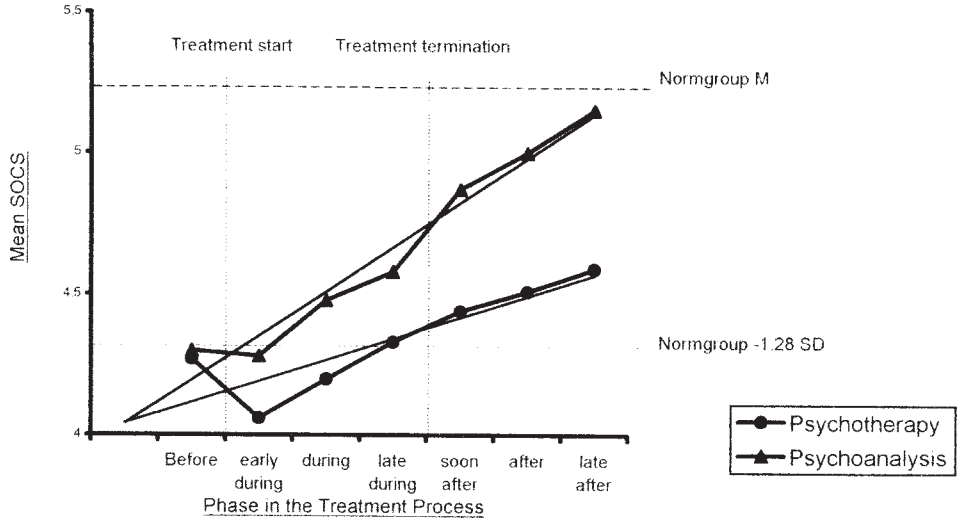

FIGURE 3. Sense of Coherence Scale (SOCS; n = 1,211): growth curves and best linear fits for patients in psychotherapy and psychoanalysis in relation to the norm group mean (upper horizontal line) and the "caseness" criterion (norm group SD = –1.28; lower horizontal line).

Social Adjustment

On the SAS, the outcome pattern was quite different (Figure 4). The psychotherapy group started off at a significantly higher level than the psychoanalysis group, and the groups then decayed at roughly the same rate. The slopes were both significantly different from zero, $t(6) < -3.85$, $p = .012$. The before and after effect sizes were small to moderate: 0.44 in the psychotherapy group and 0.40 in the psychoanalysis group.

Analyses of the subscales showed an intriguing pattern in both groups: a sharp increase at the beginning of treatment on most of the scales. In general, a cubic function appeared to yield the best fit. In Table 2, a large standard error (e.g., > .015), together with a small effect size (e.g., < .05), suggests such a pattern. The only clear exception was the Work subscale in the psychoanalysis group, which had a rather linear decay; there was also a clear, although somewhat quadratic, improvement on the Family subscale in the psychotherapy group.

Clinically Significant Outcomes

We define a clinically significant outcome (CSO) as more "healthy" scores than the least "healthy" 10% in the norm group on all three scales: the SCL-90-R, the SOCS, and the SAS. We found an increase in the proportion of CSOs in both groups across the time scale. Figure 5 shows the proportions of patients in the two groups who had a CSO at each stage of treatment. The before-treatment to late-after-treatment increase was 61% in the psychoanalysis group and 25% in the psychotherapy group.

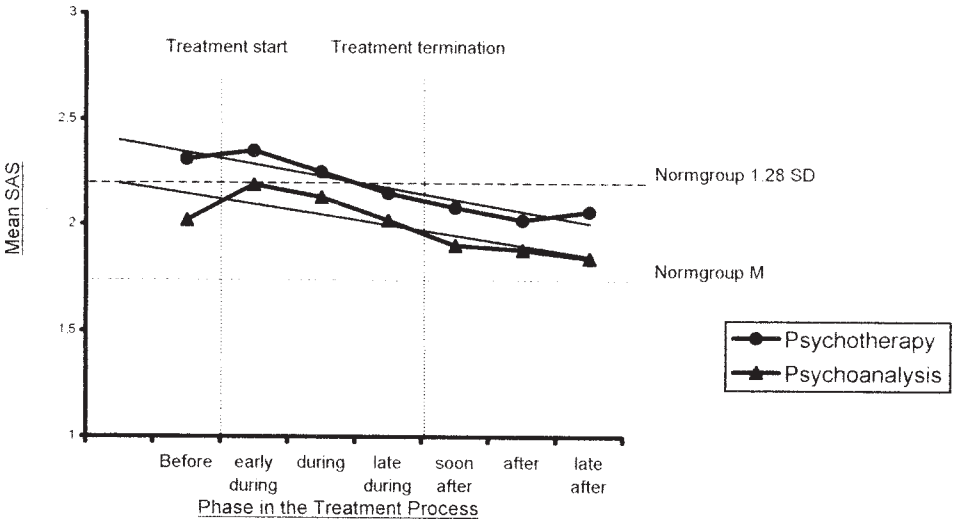

FIGURE 4. Social Adjustment Scale (SAS; $n = 1{,}211$): decay curves and best linear fit for patients in psychotherapy and psychoanalysis in relation to the norm group mean (lower horizontal line) and the "caseness" criterion (norm group $SD = 1.28$; upper horizontal line).

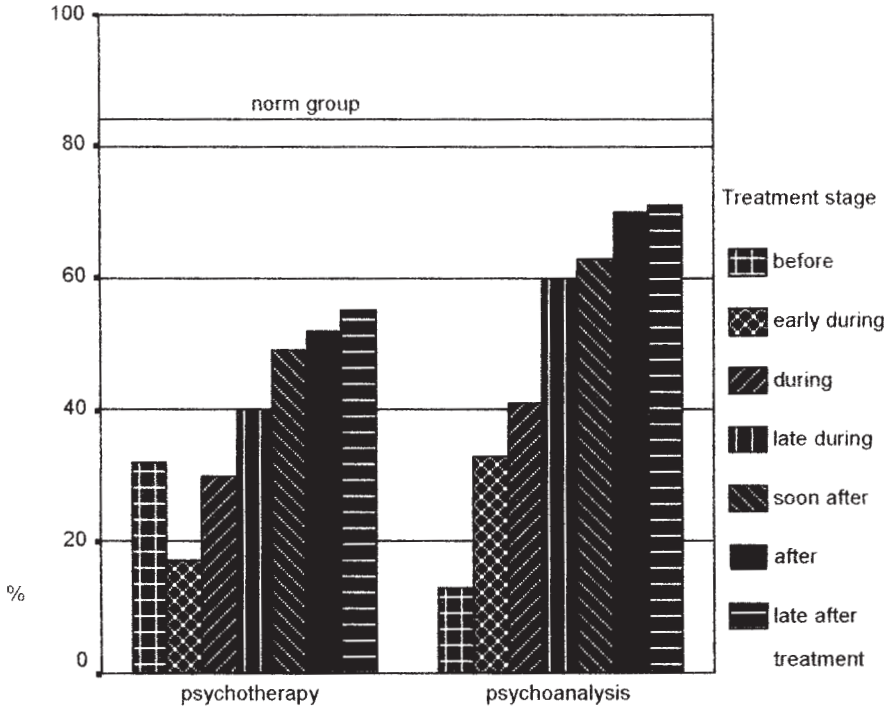

FIGURE 5. Percentages of patients in psychotherapy (n = 331) and psychoanalysis (n = 74) with clinically significant outcomes (scores "healthier" than the "caseness" criterion on the Symptom Checklist-90/Global Severity Index, Sense of Coherence Scale, and Social Adjustment Scale).

Discussion

There is a notion that psychoanalysis and long-term psychoanalytic psychotherapy are for "worrywarts." This certainly does not seem to be the case in this study. According to the lowest GAF, which is an innovation for this study, the average patient in both groups has had, at some point in life, severe psychiatric problems (Ms = 51.9 and 53.7, respectively); about 25% of the patients have had serious occupational problems according to the VIS (M = 1.2); and almost 90% had subnormal, or clinical, scores on the SCL-90, SAS, and SOC before starting treatment. Compared with this benchmark, 73% of the patients who had been in psychoanalysis ended up on the "healthy" side of the very strict CSO criterion (by way of comparison, 84% in the combined norm group were CSOs). This is, we believe, strong support for the clinical relevance of psychoanalysis.

Assuming these to be genuine treatment effects, one might, on theoretical grounds, have expected that psychoanalytically oriented treatments, with their emphasis on relational aspects, both in terms of cause and curative factors, would have had stronger effects on social relations than on symptoms. Therefore, the modest effect of both treatments on the SAS is indeed surprising. Of course, this may simply indicate that the SAS is not valid or sufficiently sensitive. We do not believe that this is the case, however. First, the SAS is a well-established instrument. Second, we have invested a great deal of effort in adapting it to the current Swedish conditions. Third,

on the basis of a pilot study, we do know that the SAS scores are clearly correlated with the quality of internal objects (Sörhus & Hylén, 1995). Fourth, the same pattern on the SAS was found on the SOCS Trust scale, which is an independent indicator of much the same qualities as the SAS. Fifth, most of the SAS subscales were certainly sensitive enough in the first stages of treatment to indicate a sharp deterioration. At this point, we are not able to explain this deviation from theoretical expectations.

Another pattern that is difficult to interpret is the initial deterioration on almost all of the SAS subscales. To speculate, this finding, which also has a parallel on the SOCS, may reflect an initial narcissistic withdrawal of object relations, a distancing from, primarily, people deemed too close. It appears as a more robust finding than would be accounted for by chance, and we suggest more fine-grained research on the initial reactions in psychotherapy.

As noted early in this article, there are few studies with which we can compare our findings. In view of the finding in the Menninger project (Wallerstein, 1989), the large outcome differences between psychotherapy and psychoanalysis proper in the current study are surprising. Of course, they have to be interpreted in light of the large dose differences. Thus, insofar as the outcome differences are due to dosage factors, it is more likely a frequency effect than a duration effect. Assuming a frequency effect, it is certainly not a simple or direct one, as if the final outcome were an accumulation of the minute effects of each single session. If so, there should have been a cumulative differentiation from the very first session on. In fact, on the SCL-90, which is actually the most sensitive of our outcome measures, the critical differentiation appears only as the treatments are terminated. That implies a more complicated effect of frequency, if any. We can only infer that something is occurring during the more frequent sessions that prompts a posttreatment process that has to be qualitatively different, because there are no quantitative differences between the treatment regimens after termination. The difference between the posttreatment processes in the two groups emphasizes the importance of length of follow-up. With follow-up intervals of typically less than 1 year (Grawe et al., 1994), many important long-term effect differences may have gone unnoticed in earlier studies. We submit that Nicholson and Berman's (1983) conclusion that long-term follow-up is not necessary is weakened when longer follow-ups are considered.

Certainly, one might argue that the outcome differences between treatments could be an effect of pretreatment differences between the two groups. We have two ways to analyze pretreatment status among patients. On the basis of data from the WbQ, we know that analysands had better pretreatment scores on the SAS but that there were no pretreatment differences in terms of symptoms (SCL-90) or morale (SOCS). An examination of the chart data reveals that differences were minor or marginal. Nevertheless, by regressing each of the outcome scales on pretreatment status variables (age; sex; education; cohabitation; earlier treatments; VIS; *DSM-IV* Axes I, II, and V [GAF]), we calculated the residual gains in each group. When the pretreatment variables were thus kept constant, we found no support for the pretreatment differences interpretation.

Conventional research wisdom would suggest randomizing patients to psychoanalysis and psychotherapy. However, our experiences with this study (Sandell et al., 1997) corroborate the common sense in the conviction that it is unrealistic (and maybe unethical) to expect people to accept, for such long stretches of time, treatment assignments contrary to their preparedness and willingness. Not only would we have had to enforce their compliance with the randomized assignments to begin with, but we would also have had to enforce that the patients stayed in their respective

treatments during the entire trial. As we have described elsewhere (Sandell et al., 1997), patients actively seek their therapies, interrupt those they are not satisfied with, and seek alternative ones. So, along with Seligman (1995) and other writers (e.g., Mumford, Schlesinger, Glass, Patrick, & Cuerdon, 1984), we believe self-selection to be part and parcel of long-term psychoanalytically oriented therapies, not only in choosing treatment but all the way through; that is, in choosing to stay in treatment, to go to today's session, day after day. Also, for treatments with such heavy contributions from expectation and motivation, randomization may indeed create more problems than it might solve. If one treatment has higher credibility or is generally more suitable with more randomized patients than another, a differential effect may be due to the credibility factor rather than to the treatments as such. In the perspective of the Aptitude × Treatment interaction (ATI) paradigm, one might indeed consider patients' favorable expectations with a certain treatment (and less favorable with another) as an aptitude factor in interaction with the treatment techniques. From that perspective, self-selection would be a way of optimizing the interaction for both treatments in a two-treatment comparison. Whereas the outcome of a randomized trial is a confound of treatment technique and patient expectations, the outcome of a self-selection trial would yield a purer effect of treatment, with optimal expectations in each treatment group.

It is true that controls for so-called spontaneous change—or, as it is sometimes called, regression—would have increased the credibility of our conclusions about therapeutic change. This would have required keeping prospective analysands and therapy patients in nontreatment for as long as 6 years. Again, it is questionable whether it would have been possible or ethical to control the adherence of nontreated patients to that condition. Nevertheless, regression toward the mean is a factor that has frequently been suggested to account for psychotherapeutic changes across time. This suggestion is often based on a stereotypic misunderstanding of the mechanism of regression. Without being too technically detailed (but see Campbell & Kenny, 1999; Hsu, 1995; Nesselroade, Stigler, & Baltes, 1980), it's safe to say that regression of individual scores toward the group mean will occur in any series of measurements on the same individuals unless the measurements are perfectly reliable. However, it will affect the mean and variance of these individuals as a group only if they have been selected to the group on the basis of a t_0 measurement and to the extent that this selection has been level biased, that is, by picking them from the top or from the bottom of the distribution of that measurement. Then the variance in this select group will increase, and the mean will decrease (if they were picked from the top) or increase (if they were picked from the bottom) from t_0 to the next measurement t_1, but there will be no further change on any measurements $t_{2\ldots n}$ as an effect of the t_0 selection. In this study, there could not have been any mean change as a result of regression from the first-wave to the second-wave measurement simply because the first-wave measurement was not used for any selection. Insofar as participants had, indeed, been level selected (or level selected themselves) to our study on the basis of some unreliable measurement or judgment t_0 before the study, their mean would have decreased (or increased) to the next measurement t_1, which might have been the first wave in the study. This decrease (or increase) would have been toward their true mean value, the estimate of which had been inflated (or deflated) on the t_0 measurement as a result of unreliability. The important lesson is that t_0 measurements should not be used in measures of change or outcome, and they were not in this study. If there were indeed any self-selection in the study, such that the sickest individuals would have tended to drop out from the first wave (t_0) to the second (t_1),

this would have produced a selection from the healthy end of the t_0 distribution. To the extent that our measurements in that wave were related to the self-selection criterion ("health sickness"), the t_0 means in the remaining group of participants would then have been "healthier" than the t_1 means, thus creating an apparent increase of "sickness" in the group, contrary to what was observed. We believe this effectively rules out regression toward the mean as an explanation for the across-time trends in our outcome variables.

The reader who is still not convinced may want to reconsider our research design. The design was such that the A to F subgroups (see Figure 1), although different in relative time, were, in fact, measured at the same time in the same 3 consecutive years. Whereas one might believe that internal validity threats like history, maturation, regression, or simply time would have helped to produce change across the three years/waves in each group, they could not explain the differences between the A to F subgroups in the outcome variables. If we regress an outcome variable like the GSI on real time/years/waves and test the intercept and slope differences between the subgroups, we find a significant linear trend across time for both, $F(1, 398) = 20.99$, $p = .000$, for the intercept, and $F(1, 398) = 8.58$, $p = .004$, for the slope. Thus, across the same real time, the subgroup means changed linearly as a function of relative time, that is, time in relation to treatment. This trend suggests other explanations aside from regression to the mean.

Like any other study, this one leave some questions unanswered. Clearly, our study does not "prove" the efficacy of long-term psychoanalytically oriented treatments for particular diagnostic groups. Nor does it give any "once-and for-all" proof that psychoanalysis outperforms psychotherapy. With respect to the efficacy versus effectiveness debate (VandenBos, 1996), ours is clearly an effectiveness study whose main asset is high ecological validity. We believe that it shows that long-term psychoanalytic therapy and psychoanalysis in particular, as practiced in the field, are not a waste of time for those patients who actively choose these kinds of treatments.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.).Washington, DC: American Psychiatric Association.

Antonovsky, A. (1987). *Unraveling the mystery of health*. San Francisco: Jossey-Bass.

Bachrach, H. M., Galatzer-Levy, R., Skolnikoff, A., & Waldron Jr., S. (1991). On the efficacy of psychoanalysis. *Journal of the American Psychoanalytic Association, 39,* 871–916.

Blomberg, J., & Sandell, R. (in press). Close family or mere neighbours? Some empirical data on the differences between psychoanalysis and psychotherapy.

Carroll, L. (1982). Alice's adventures in wonderland. In E. Guiliano (Ed.), *Complete illustrated works of Lewis Carroll*. London: Chancellor Press. (Original work published 1865)

Campbell, D. T., & Kenny, D. A. (1999). A primer on regression artifacts. New York: Guilford Press.

Consumer Reports. (1995, November). Mental health: Does therapy help? pp. 734–739.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Science, 19,* 1–15.

Derogatis, L. R., & Lazarus, L. (1994). SCL-90-R, Brief Symptom Inventory, and matching clinical ratings scales. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 217–248). Hillsdale, NJ: Erlbaum.

Doidge, N. (1997). Empirical evidence for the efficacy of psychoanalytic psychotherapies and psychoanalysis: An overview. *Psychoanalytic Inquiry, 17*(Suppl.), 102–150.

Fisher, S., & Greenberg, R. P. (1996). *Freud scientifically reappraised*. New York: Wiley.

Fonagy, P. (Ed.). (1999). *An open door review of outcome studies in psychoanalysis. Report prepared by the Research Committee of the IPA at*

the request of the President [On-line]. Available: www.ipa.org.uk.

Frank, J. (1974). Psychotherapy: The restoration of morale. *American Journal of Psychiatry, 131,* 271–274.

Grawe, K., Donati, R., & Bernauer, F. (1994*). Psychotherapie im Wandel. Von der Konfession zur Profession (4. Aufl.).* Göttingen, Germany: Hogrefe.

Green, A. (1996). What kind of research for psychoanalysis? *International Journal of Psychoanalysis, 5,* 10–14.

Hjelle, L. A., & Ziegler, D. J. (1981). *Personality theories: Basic assumptions, research, and applications*. New York: McGraw-Hill.

Hsu, L. M. (1995). Regression toward the mean associated with measurement error and the identification of improvement and deterioration in psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 141–144.

Karterud, S., Friis, S., Irion, T., Hehlum, L., Vaglum, P., & Vaglum, S. (1995). An SCL-90 derived index of the severity of personality disorders. *Journal of Personality Disorders, 9*, 112–123.

Kantrowitz, J. L. (1997). A brief review of psychoanalytic outcome research. *Psychoanalytic Inquiry, 17*(Suppl.), 87–101.

Kazdin, A. E. (1994). Methodology, design, and evaluation in psychotherapy research. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 19–71). New York: Wiley.

Lazar, A. (1995). Det är väl bara så'na som har problem själva . . . –Psykologistudenters livskvalitet och hälsotillstånd. [Isn't it just a bunch of basket cases . . . ? Quality of life and general health status among psychology students]. Unpublished master's thesis, Department of Psychology, University of Stockholm.

Mumford, E., Schlesinger, H. J., Glass, G. V., Patrick, C., & Cuerdon, T. (1984). A new look at evidence about reduced cost of medical utilization following mental health treatment. *American Journal of Psychiatry, 141,* 1145–1158.

Nesselroade, R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin, 88,* 622–637.

Nicholson, R. A., & Berman, J. S. (1983). Is follow-up necessary in evaluating psychotherapy? *Psychological Bulletin, 93,* 261–278.

Orlinsky, D., Ambühl, H., Rønnestad, M. H., Davis, J., Gerin, P., Davis, M., Willutzki, U., Botermans, J.-F., Dazarod, A., & Cierpka, M. (1999). Development of psychotherapists: Concepts, questions, and methods of a collaborative international study. *Psychotherapy Research, 9,* 127–153.

Orlinsky, D., Rønnestad, M. H., Ambühl, H., Willutzki, U., Botermans, J.-F.,Cierpka, M., Davis, J., Davis, M., & the SPR Collaborative Research Network. (1999). Psychotherapists' assessments of their development at different career levels. *Psychotherapy, 36,* 203–215.

Sandell, R., Blomberg, J., & Lazar, A. (1997). When reality doesn't fit the blueprint: Doing research on long-term psychodynamic therapies in a public health service program. *Psychotherapy Research, 7,* 333–344.

Sandell, R., Blomberg, J., & Lazar, A. (1998). The factor structure of Antonovsky's Sense of Coherence Scale in Swedish clinical and nonclinical samples. *Personality and Individual Differences, 24,* 701–711.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The *Consumer Reports* study. *American Psychologist, 32,* 752–760.

Shapiro, D. A., Harper, H., Startup, M., Reynolds, S., & Suokas, A. (1994). The high-water mark of the drug metaphor. A meta-analytic critique of process-outcome research. In R. L. Russell (Ed.), *Reassessing psychotherapy research* (pp. 1–35). New York: Guilford Press.

Shapiro, D. A., Rees, A., Barkham, M., Hardy, G., Reynolds, S., & Startup, M. (1995). Effects of treatment duration and severity of depression on the maintenance of gains after cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 378–387.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy.* Baltimore, MD: Johns Hopkins University Press.

Sundland, D. M. (1977). Theoretical orientations of psychotherapists. In A. S. Gurman & A. M. Razin (Eds.), *Effective psychotherapy: A handbook of research* (pp. 189–218). Elmsford, NY: Pergamon Press.

Sörhus, K., & Hylén, C. (1995). Objektrelationer och sociala kontakter [Object relations and social relations]. Unpublished master's thesis, Department of Psychology, University of Stockholm.

VandenBos, G. (Ed.). (1996). Outcome assessment of psychotherapy [Special issue]. *American Psychologist, 51,* 1005–1072.

Wallerstein, R. S. (1989). The psychotherapy research project of the Menninger Foundation: An overview. *Journal of Consulting and Clinical Psychology, 57,* 195–205.

Weissman, M. M., & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. *Archives of General Psychiatry, 33,* 1111–1115.

Weissman, M. M., Prusoff, B. A., Thompson, W. D., Harding, P. S., & Myers, J. K. (1978). Social adjustment by self-report in a community sample and in psychiatric outpatients. *Journal of Nervous and Mental Disease, 166,* 317–326.

Zusammenfassung
Dass es so wenig Ergebnisse über die Wirksamkeit von psychodynamischen Langzeittherapien und Psychoanalyse gibt, ist ein Problem, zumindest für die Länder, in denen diese Therapien die im allgemeinen bevorzugte Behandlungsmethode sind. In dieser Untersuchung wurde eine Stichprobe ausgewählt von 405 Patienten 'vor', 'während' und 'nach der Behandlung' aus einer Gesamtheit von 1200 Patienten, die durch öffentliche Mittel Unterstützung für ihre Langzeittherapien (n = 331), beziehungsweise Psychoanalyse (n = 74) empfangen hatten oder sich auf einer Warteliste für diese Behandlungen befanden. Die durchschnittliche Behandlungsdauer war 40 Monate (SD = 21) für Langzeittherapie und 51 Monate (SD = 18) für Psychoanalyse. Alle Patienten haben zu drei Zeitpunkten mit einem 12-Monatsintervall Fragebögen ausgefüllt unter Einschluss der Symptom Check List (SCL-90), Sense of Coherence Scale und Social Adjustment Scale (SAS). Mit einem gemischten 'innerhalb' und 'zwischen' Personen Design wurden Regressionsanalysen der Gruppenmittelwerte auf sieben Stadien im Behandlungsprozess durchgeführt: vor Beginn der Behandlung, in einem frühen Behandlungsstadium, während der Behandlung, in einem späten Stadium der Behandlung, kurz nach der Behandlung und mit einigem Zeitabstand nach dem Ende der Behandlung. Kleine bis mittlere Effektstärken (d = 0,4 und 0,6) resultierten für Langzeittherapien und mittlere bis sehr große (d = 0,4–1,5) für die Psychoanalyse. Die größte Effektstärke bezog sich bei beiden Gruppen auf die SCL-90 und die kleinste auf die SAS. Die Positionierung von Patienten mit Scores im 'gesunden' Bereich im Vergleich zu den 10% schlechtesten Werten einer nicht-klinischen Normgruppe verzeichnete bei den Langzeittherapien einen Anstieg von 30% auf 55% und bei der Psychoanalyse von 10% auf 75%. Dieses Ergebnis ist auch geeignet, die Bedeutung einer 'Langzeit-Katamnese' bei Langzeitbehandlungen zu unterstreichen.

Résumé
La rareté de données au sujet de l'efficacité de la psychothérapie psychodynamique de longue durée et de la psychanalyse est un problème, au moins dans des pays où ces thérapies sont les traitements généralement choisis. Dans cette étude, un échantillon de 405 patients en phase variée du traitement—avant, pendant, après—a été recruté d'une population de 1200 patients ayant reçu ou étant sur la liste d'attente pour une subvention publique pour une psychothérapie de longue durée (n = 331) ou une psychanalyse (n = 74). La durée moyenne était de 40 mois (SD = 21) pour les psychothérapies et de 51 mois (SD = 18) pour les psychanalyses. Chaque patient a rempli un questionnaire contenant la Symptom Distress Checklist (SCL-90), la Sense of Coherence Scale, et la Social Adjustment Scale (SAS) à 3 reprises et 12 mois d'intervalle. Sur la base d'un design partiellement intra- et intersujets, les moyennes des groupes étaient régressées sur une échelle de temps ordinale regroupant 7 phases du processus thérapeutique : avant, tôt pendant, pendant, tard pendant, peu après la fin, après la fin, et tard après la fin du traitement. Les pentes indiquent des taux d'effet petits à modérés (ds = 0.4 et 0.6), et modérés à très grands (d = 0.4–1.5) pour la psychanalyse. Pour les deux groupes, on constate les taux d'effet les plus grands à la SCL-90 et les plus faibles à la SAS. Le taux de patients à des scores plus « sains » que les 10% du groupe de comparaison nonclinique au score le plus mauvais a augmenté d'environ 30% à 55% dans le groupe de psychothérapie, et de 10% à 75%, dans le groupe de psychanalyse. Les résultats soulignent l'importance d'une évaluation à long terme quand il s'agit de traitements de longue durée.

Resumen
La escasez de datos sobre la efectividad de la psicoterapia y el psicoanálisis prolongados es un problema, por lo menos en los países en que estas terapias son los tratamientos de elección. Para este estudio, se tomó una muestra de cuatrocientos pacientes de una población de mil doscientos que habían recibido (o estaban en lista de espera para obtenerlo) un subsidio público para realizar psicoterapia (n = 331) o psicoanálisis (n = 74) prolongados. Los pacientes elegidos estaban en varias etapas del tratamiento (antes, durante y después). La duración promedio de las psicoterapias fue de cuarenta meses (SD = 21) y de cincuenta y un meses (SD = 18) para el psicoanálisis. Todos los pacientes completaron, en tres oportunidades, con intervalos de doce meses entre sí, la Lista de Síntomas de Distrés (Symptom Distress Checklist, SCL-90), la Escala de Sentimiento de Coherencia y la Escala de Adaptación Social (SAS). Con un diseño parcialmente dentro y entre sujetos, se hizo una regresión con las medias de los grupos en una escala de tiempo ordinal que cubrió 7 etapas: antes, al empezar, durante, avanzado, al terminar, después y mucho después de terminado el tratamiento. Las pendientes de los gráficos mostraron efectos entre pequeños y moderados (ds = 0.4 y 0.6) para la psicoterapia y entre moderados y muy grandes (d

= 0.4–1.5) para el psicoanálisis. Para ambos grupos, los mayores efectos se registraron con el SCL-90 y los menores con el SAS. La proporción de pacientes con puntajes más "saludables" que el peor (10% en un grupo no clínico control) aumentó aproximadamente del 30% al 55% en el grupo de psicoterapia y del 10% al 75% en el de psicoanálisis. Los resultados subrayan la importancia del seguimiento sostenido en la evaluación de tratamientos prolongados.

Received April 13, 2000
Revision received November 19, 2000
Accepted April 20, 2001