Psychodynamic Psychotherapy:
An Effectiveness Study

David J. Roseborough

The College of Saint Catherine and The University of Saint Thomas

Objective: Both the National Institute of Mental Health and the American Psychological Association have called upon psychodynamic practitioners to start demonstrating their outcomes. This effectiveness study attempted to begin to answer these calls. Method: The study was a secondary analysis of data from a multidisciplinary, psychodynamic mental health clinic. It used a single-group, within-subjects longitudinal design. The psychometrically validated Outcome Questionnaire was used as a measure of change. A linear mixed and random effects model was used to analyze the data. The aims of this study were (a) to look at whether subjects improve and (b) if so, at what variables moderate outcome. Results: Findings suggest that psychodynamic treatment, provided within this practice configuration, is effective over time, producing moderate effect sizes, and points to the particular importance of the first three months. Conclusions: Findings suggest a common course of recovery, with some between-group variability.

Keywords: psychodynamic; psychoanalytic; intervention research; outcome; effectiveness; efficacy; randomized clinical trial

Most research regarding psychotherapy outcomes is done in highly controlled settings. It uses structured, manualized treatments applied to people with a single psychiatric diagnosis. Although strong in design (i.e., providing strong internal validity), questions have been raised about the applicability of the findings from these efficacy studies to real-life clinical settings. This is a concern for many clinicians in that the majority of potential research subjects in such efficacy studies, especially those with more than one diagnosis, are often screened out up front (Nathan, Stuart, & Dorn, 2000). These studies have, to date, most often evaluated the outcomes of cognitive-behavioral (i.e., cognitive behavioral therapy) interventions, as can be seen in a review of the American Psychological Association's (APA) Division 12 listing of empirically supported treatments, of which only a few are psychodynamic in nature (APA, 1993).

The National Institute of Mental Health ("Exploratory/developmental grants for psychosocial treatment research," 1993) and others (APA, 1993; Gabbard, Gunderson, & Fonagy, 2002) have lamented the lack of empirical outcome data provided by psychodynamic therapists and have called for such practitioners to start demonstrating their outcomes. Gabbard, Gunderson, and Fonagy (2002) wrote recently about the threat posed to dynamic therapy by not demonstrating such outcomes: "Psychoanalytic psychotherapy is at risk of being sacrificed if scientific methods cannot be developed that will further test its practitioners' claims of efficacy" (p. 505).

This study was an effectiveness study: using a research design with fewer controls, done in the field, to begin answering this call. It looked at the outcomes of psychodynamic psychotherapy as it is actually practiced: within an agency setting and with different groups of clients. It looked at how effective psychodynamic psychotherapy is and at what its change process looks like.

Specific aims of the study: The study asked the following questions.

> Do clients receiving this type of treatment make significant improvement over time?
> If so, what variables appear to moderate its outcome?

Effectiveness research in this area has been, to date, quite limited. In the mid-1990's, *Consumer Reports* published a study based on client self-report. That study involved thousands of people self-reporting on their satis-

faction with their own psychotherapy but did not look at symptom change per se. In fact, its author (Seligman, 1995) called for the next research in this area to look beyond such self-report and to use more empirical measures of symptom change. This study attempted to do just that.

Psychodynamic Psychotherapy Evaluation

Psychodynamic psychotherapy is one of the most commonly practiced forms of therapy (Anderson & Lambert, 1995; Doidge, 1999; Gunderson & Gabbard, 1999; Svartberg & Stiles, 1991; Vaughan et al., 2000) and yet one of the least researched. Despite its being widely practiced, this form of therapy has historically avoided evaluation (Doidge, 1999; Gabbard, Gunderson, & Fonagy, 2002; Sandell, Bloomberg, Lazar, & Carlsson, 2000). Dynamic practitioners have historically claimed efficacy for their treatments by referring to case studies (i.e., Wallerstein's *Forty-two lives in treatment: A study in psychoanalysis and psychotherapy* in 1986), often pointing to changes observed exclusively from the analyst's perspective. Others have asserted that dynamic therapy's outcomes are beyond measurement (i.e., involving internal, personality or structural change not easily or at all able to be quantified). Still others have called their practice an art more than a science and thus not open to empirical evaluation (Stone, 1997). Doidge puts the sentiment of many of his colleagues concisely: "We reject empirical outcome research" (p. 674).

This historic stance, however, is gradually changing, with psychoanalytic leaders themselves calling for more empirical research. Freedman, Hoffenberg, Vorus, and Frosch (1999) note, "Psychoanalysis finds itself not in a situation of crisis, but surely in one of reorganization" (p. 741). Kernberg (1991) raises the concern that by having avoided research, "our psychoanalytic institutes are characterized by an atmosphere of indoctrination rather than of free scientific exploration" (p. 55). By this he means that psychoanalysis is at risk of being taught as an ideology, as a closed system not open to the challenges or correctives scientific inquiry might bring. In so doing, he is echoing an older criticism of psychoanalysis brought originally by philosopher of science Karl Popper, who rejected psychoanalysis as a closed or circular system not open to the scrutiny of science (Popper, 1963).

Kernberg also calls for psychoanalytic schools not to be free-standing but to join with universities in the spirit of holding psychoanalytic concepts up to scientific scrutiny (p. 57-58) and points to the risk of the field's stagnation by failing to do so (p. 61). He ultimately hopes for the field to become more scientific: to test and to ground its

assertions in science (p. 61). He is not alone in this desire. Other leaders within psychoanalysis such as Gabbard and Gunderson (1999) are acknowledging the need now to evaluate its outcomes empirically (see also Barber & Lane, 1995; Doidge, 1999; Sandell et al., 2000).

METHOD

Sample

Participants (*n* = 164) were all adult outpatients in psychodynamic psychotherapy, as offered by this clinic. Thirty different treating therapists were represented in this sample. Treating therapists were primarily master-level social workers and psychologists. Most subjects were seen for just longer than a year, averaging 64 sessions, although with a large degree of variation (*SD* = 44.6). The total number of sessions ranged between 4 and 237 sessions. Most subjects were seen somewhere between 20 and 100 sessions, usually with one time a week as the standard of care. Approximately a third (31%) of these clients reported a planned termination, whereas 44% of these clients simply withdrew. Sixteen percent ended because of their therapist's internship ending. The rest ended for other reasons, including a group ending, being referred out, or for unidentified reasons.

Clients ranged in age between 20 and 83. The mean age for a client in this sample was 38 (*SD* = 12.00). The majority of clients were female (60%). In keeping with national samples, the majority were on medication at some point during their therapy. Total, 68% used medication during their therapy, whereas 32% did not. This is a higher percentage than the national average of 62% (Olfson, Marcus, Druss, & Pincus, 2002) and may be accounted for by the clinic having psychiatrists on staff, in house, and by its strongly interdisciplinary model.

Just under one half (45%) of subjects had one or more comorbid disorders, whereas just more than half (51%) were diagnosed with a single presenting diagnosis (4% of this data was missing). The vast majority of these diagnoses were Axis I diagnoses (i.e., clinical syndromes), with only a small number of people presenting with an adjustment disorder or V-Code (*n* = 19) or personality disorder (*n* = 3) as primary diagnoses. Fourteen subjects (8.5%) presented, although, with a personality disorder as part of their overall diagnosis. Initial presenting diagnoses can be broken down according to major depression (*n* = 49), dysthymia (*n* = 43), mood-other (*n* = 19), anxiety (*n* = 22), adjustment disorders or V-codes (*n* = 19), and other (

Table 1:
OQ Sample Scores in Relation to National Norms

	Clinic's Mean	National Mean
OQ total	82.09	83.09
OQ symptom distress	47.62	49.40
OQ interpersonal relationship	20.09	19.68
OQ social role	14.46	14.01

NOTE: OQ = Outcome Questionnaire.

consisting of alcohol dependence, eating disorders, etc., with $n = 6$).

Clients' initial presentations were also in keeping with national norms for adult outpatients entering mental health settings (see comparison table below). The presenting scores were as follows.

Treatment Conditions

Setting. Hamm Memorial Psychiatric Clinic is an adult, outpatient mental health center located in St. Paul, Minnesota. It is a multidisciplinary clinic, employing eight social workers, five psychologists, and three psychiatrists. The clinic provides both psychodynamic psychotherapy as well as psychiatric consultation, with or without medication management. It does not offer medication management apart from psychotherapy. That is, all clients receiving medication at the clinic are currently clients in therapy at the clinic. Services are offered along a sliding scale. The clinic also provides clinical training for graduate level psychologists, psychiatrists, and social workers as an integral part of the clinic's practice structure (i.e., its service and training milieu). It also provides continuing education in psychodynamic topics for mental health practitioners already in practice in the metropolitan area.

Procedures

Staff therapists consist of graduate level psychologists and social workers, all of whom are educated at either a master's or doctoral level and all of whom have significant training and experience in this relationship-focused model of psychotherapy, including continuing training and education through the clinic itself. Many are graduates of the clinic's training program and several teach the cohort of 8 to 10 interns and residents within the clinic at any given time, in addition to their practice. Treating therapists are Caucasian ($n = 10$), Hispanic ($n = 2$), and African American ($n = 1$). Most of the practitioners are women ($n = 11$), whereas five are men. The clinic employs three psychiatrists, who are trained as psychotherapists. Clinic administrators are similarly clinicians and carry a case load as part of their clinic duties. The clinic's work is strongly cross-disciplinary. Clients are considered clinic clients. Social workers, psychologists, and psychiatrists have input into cases during weekly group supervision.

Practitioners use an approach that is generally consistent with the principles described by Gabbard (2000) in his book *Psychodynamic Psychiatry*. This model is based on both a descriptive (i.e., DSM-IV) and relational diagnosis (understood as the impact of both past and present relationships), the fostering of a strong working alliance, the use of the therapeutic experience as a corrective one, the exploration of transference, and the positive use of the therapy relationship in general. This therapeutic relationship is seen as something that is gradually internalized by the client. It occurs along a supportive-expressive continuum, and careful attention is given to both resistance and to the client-therapist interaction (Gunderson & Gabbard, 1999). The clinic draws heavily on both (a) attachment and (b) object relations as its core theoretical underpinnings. The clinic estimates the cost of this intervention at $2,645.00 per course of therapy (this estimate assumes a 23 session intervention, the clinic's average, at a cost of $115.00 per session).

Outcome Measure

Measure. The study used the Outcome Questionnaire (OQ-45.2) (Lambert, Gregersen, Brulingame, & Marush, 2004; Lambert et al., 1996). The OQ is a 45 item client-administered questionnaire developed specifically to measure outcomes relevant to dynamic therapy. It uses a 5-point scale. The OQ provides both an overall score as well as three subscales or domain scores: (a) symptom distress (how the person feels inside, the level of distress symptoms cause), (b) interpersonal relations (how the person gets along with others), and (c) social role (how the person functions in terms of important life tasks). The instrument has been normed on several community samples, including the psychiatrically well, clients of employee assistance programs, clients in outpatient mental health therapy, and with psychiatric inpatients. It has also been tested across gender, race, and with various ethnicities. It includes a cut-off score that differentiates normal community samples from symptomatic and impaired clinical samples as well as a way to calculate clinically significant change (i.e., versus a full recovery) on all four scales. The OQ has been tested for both internal

consistency and for test-retest reliability (see Lambert, Finch, & Maruish, 1999; Mueller, Lambert, & Burlingame, 1998).

Data Analysis

The statistical package SPSS 12.0 was used to perform the inferential statistical analyses. A mixed effects model was chosen largely because of the nature of the data itself. For instance, repeated measures ANOVA could not be performed because of the amount of missing data in earlier clinic cases. Although most cases included a baseline, the other data points were spread out during the course of 3 or more years of therapy. There was a high degree of attrition: 127 cases had a baseline, 66 had a 3-month score, 64 had a 6-month score, 43 had a 9-month score, and only 35 had a score at 12 months. From 15 to 24 months, the number of respondents ranged around the low 20s at each 3-month interval, and only 7 respondents had data points at 3 or more years). Because of this degree of attrition (i.e., drop off), the decision was made to analyze a year's worth of data, which would be in keeping with existing studies (i.e., Asay, Lambert, Gregersen, & Goates, 2002; Kopta, Howard, Lowry, & Beutler, 1994) and offer a good basis for comparisons with existing literature.

The mixed effects model used is a regression-like approach. This approach allows the researcher to create a trend line, which shows both (a) the direction and (b) the statistical significance of a change. It also allows both within and between group comparisons. The model looks for trends or changes over time, by plotting residuals. As in regression, it assumes that a random or nonsignificant line is a straight and horizontal one. Time is seen as a fixed variable and each subject at each time point (in this case, each 3-month interval) varies randomly. Nonrandom changes move this line up or down to significant or nonsignificant degrees. Treating subjects as random effects also saves a number of degrees of freedom and allows for later between group comparisons.

Power Analysis

A power analysis was conducted using software from SSI.com (Congdon, 2001) called "Optimal Design", which provides a power analysis uniquely suited to (a) a mixed model, which uses both (b) repeated measures and (c) linear, quadratic, or cubic data. This model does so by factoring in the duration of the study, the expected effect size, the frequency and spacing of observations, the number of subjects, and the expected degree of variation across persons (Raudenbush & Xiao-Feng, 2001). It uses all of this information to create a multidimensional power analysis (relying on vectors and matrices vs. a single value the way a unidimensional power analysis might in the case of ANOVA). Using this software and its underlying model (see Raudenbush & Xiao-Feng, 2001), it was determined that the 164 people in this study would provide a 72% probability of correctly identifying significant versus nonsignificant findings (a power or d value of .72).

RESULTS

Research Question 1: Will people show statistically significant change between beginning and ending this treatment?

Effect Size

The overall effect size for this treatment was found to be moderate. The effect size for 1 year of treatment was calculated by subtracting the score at 1 year (OQTOT5) as the intervention mean from the baseline OQ (OQTOT1) and dividing that number by the baseline (OQTOT1) standard deviation. This yielded a d index of .41. This value is squarely in the middle of a medium effect size, as defined by Lipsey (1990) in a review of more than 6,700 studies on treatment effectiveness research. Similarly, an effect size was calculated using the same approach for the first 3 months of treatment, which had appeared as the only 3-month interval during which statistically significant change happened as part of, or defining, a unique treatment period. This analysis yielded an effect size of .44. This treatment effect was also thus moderate. The treatment effect was found to be, as expected, moderate or medium. The actual strength of the effect, however (.4) was less than anticipated and somewhat lower than the range in the existing literature, in which estimates have ranged from .6 to 2.02. Effect sizes for each of the subscales averaged .33.

OQ Total Score

Respondents showed statistically significant change on the OQ total score (OQTOT) between beginning and ending this treatment. The mixed effects model looked at OQ total scores during the course of a year of treatment. OQ total scores were defined as the dependent variable. Time was identified as a fixed factor. A high degree of correlation between the testing periods was assumed and

was factored into the model, as Lambert et al. (1996) point to strong correlation coefficients between weeks on the OQ during a 10-week period ranging from .66 to .82 (p. 10). This assumption was also verified for this data set by running correlations between each of the OQTOT scores at each 3-month interval, which revealed correlations between OQTOT scores ranging from .312 to .875 (with most in the .5 to .7 range). The model constructed as such produced an *F* value of 4.67 (210.124), significant at *p* < .01, identifying OQ scores as changing significantly over time—in this case, during the course of a year. The nature of this change (in OQ total scores during a year) can be graphed as follows.

Change on this overall score (OQTOT) produced an effect size of .41.

Figure 1:   Overall Change

OQ Symptom Distress Subscore

Statistically significant change was also made during the course of a year on the symptom distress subscale, which produced an *F* value of 4.1 (211.5), significant at < .01. That is, subjects did make symptomatic improvement during the course of a year, which was statistically significant. This held true across gender (*F* = .007, [140.4], sig. = .93) and whether someone had a personality disorder (*F* = 2.3, [142.6], *p* = .13. Medication did make a difference here, with a large *F*(31.9 [145.02, *p* < .01]). Change on this subscale produced an effect size of .38.

OQ Interpersonal Relationships

A similar pattern held true for the interpersonal relationship (IR) subscale. Statistically significant change was made during the course of a year, for both genders and for people with and without personality disorders (IR = *F* 3.1 [204.7], *p* < .016; SR = *F* 2.4 [276.6], *p* < .05). Change on this subscale produced an effect size of .29.

OQ Social Role Subscale

Statistically significant change was made during the course of a year, for both genders and for people whether they had a personality disorder. Similarly, the presence or absence of comorbidity did not affect outcome (*F* 1.450 [144.0], *p* = .23), whereas medication did (*F* 17.8 [155.2], *p* < .01). Change on this subscale produced an overall effect size of .31.

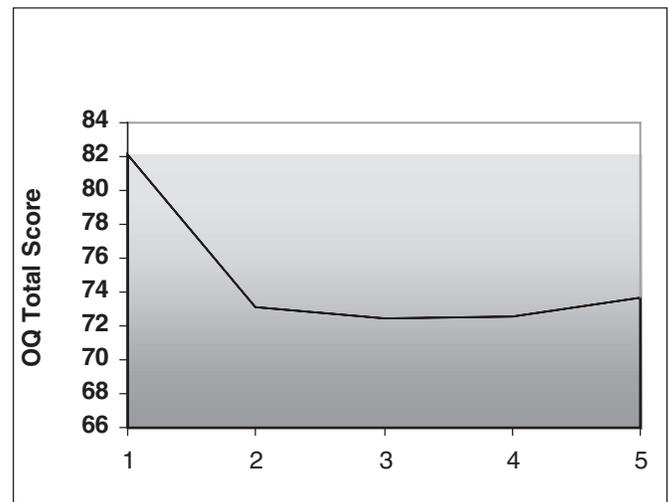Research Question 2: What are the variables significantly moderating outcome?

Results for each hypothesis individually tested are given below. An alpha level of .05 was used for all statistical tests.

Medication. Medication did exert a moderating influence in this study. Overall, people who were medicated during treatment began as more symptomatic than those who did not use medication. Those who were medicated tended to begin as more symptomatic and made a quicker initial recovery (had a stronger initial slope). This difference in slope (i.e., the nature of the recovery between the two groups), however, was not statistically significant. This held true across diagnoses. Interaction effects were explored between medication and depression (*F*[136.2] = 1.717), anxiety (*F*[139.5] = .093), and adjustment disorders and V-codes (*F*[127.5] = .61). None were found to be specific.

That is, although medication exerted a strong moderating effect, this effect was broad and not unique to any of these three broad diagnostic groups. This is substantiated, too, by the finding that medication accounted for only 13.8% of variance explained, when added as a covariate and compared against recovery in general. Those who did not use medication also made statistically significant change (i.e., their scores trended downward) but more evenly and gradually.

Comorbidity. The presence or absence of a comorbid disorder did moderate outcome in this model, when comorbidity was specified as a covariate and treated as a fixed effect. It produced a strong effect, with an

associated F value of 4.630 ($df = 133.0$) and was significant at $p < .05$. As might be predicted, people with comorbidity tended to begin treatment more symptomatic than those without. They did not, however, vary significantly in their recovery trajectory (i.e., slope) ($F[200.4] = .7$) from those without it. The presence of comorbidity accounted for only 3.9% of explained variance.

Personality disorder. Similarly, the presence of a personality disorder did not moderate outcome, when specified as a covariate and treated as a fixed effect ($F[141.5] = 2.8$). It accounted for only 3.9% of explained variance. Thus, the findings in this study suggest that although the presence or absence of comorbidity might predict a different starting point (i.e., intercept), the presence of a personality disorder did not necessarily carry the same prognostic or predictive power. In this study, subjects with comorbidity tended to begin treatment more symptomatic than those without it but followed a similar treatment recovery path (i.e., slope) to those without it.

Demographic variables. As can be seen below, demographic variables as a group did not tend to moderate outcome, with the exception in this study of household income. As predicted, gender did not moderate outcome in the mixed model, when specified as a covariate and treated as a fixed effect. It produced an F value of .01 (139.27 *df*). It accounted for a negligible .006% of explained variance. Age did not moderate outcome in this model, when specified as a covariate and treated as a fixed effect. The model produced an F value of 1.1 (348 *df*). Race did not moderate outcome in this model, when specified as a covariate and treated at a fixed effect. The model produced an F value of 3.01 (138.7 *df*), when dummy columns were created to compare White versus non-White in terms of treatment response over time.

Household income. Household income did moderate outcome when specified as a covariate and treated as a fixed effect ($F[134.3] = 10.3, p < .05$). People with greater incomes began treatment as more symptomatic than those without, although again, their treatment path followed that of the broader clinic population. Income did not correlate significantly with length of treatment (i.e., number of sessions) in a bivariate correlational analysis. People with greater incomes did not tend to stay any longer (or shorter) than others in treatment. Although statistically significant, household income itself also only accounted for 4.5% of explained variance when added as a covariate.

There was no interaction effect between improvement over time and income: $F(197.9) = 1.24$.

Education. Educational level did not moderate outcome when specified as a covariate and treated as a fixed effect in this model. Two categories were compared. First, those who attended college were compared with those who had not and no significant differences were found: $F = .13$ (139.2). Also, those who had attended high school only were compared with those who had pursued any form of postsecondary education. No significant differences emerged here either: $F = .32$ (144.2).

Diagnostic cluster. Diagnostic cluster was broken down into three diagnostic subgroups that were compared: (a) mood disorders, (b) anxiety disorders, and (c) adjustment disorders or V-codes (i.e., problems in living). Both mood disorders ($F = 3.5$ [136.4]) and anxiety disorders ($F = .53$ [141.16]) failed to moderate outcome when compared against other diagnoses. This held true as well when depression and anxiety were compared against each other ($F = .1$ [59.7]). Only adjustment disorders emerged as a cluster significantly moderating or affecting outcome over time. People with adjustment disorders began treatment as significantly less symptomatic than the broader clinic population (with a mean starting OQ total score of 64.88) but followed a very similar recovery path.

Length of treatment. The length of treatment defined as the number of visits did moderate outcome in this model when number of sessions was specified as a covariate and treated as a fixed effect. It produced an F value of 5.4 (1, 176.3), $p < .05$. This took the form of a general effect and was not diagnosis-specific. It was also a small effect, accounting for only 2.5% of explained variance. Similarly, no specific interactions were found between number of sessions over time and anxiety, depression, or adjustment disorders and V-codes. Interestingly, those people who would stay longest in treatment began somewhat less symptomatic than those who would end treatment earlier. The treatment path, however, had a similar slope. There was no unique recovery path that emerged for people who stayed longest in treatment. It thus seems that number of sessions exerts a significant influence but only insomuch as it demarcates where people begin treatment (as less symptomatic). It does not, from these findings, predict a unique recovery pattern, a differential effect on any particular diagnostic group, or correlate with higher or lower scores after a year of treatment.

Thus, although their beginning score is unique, their treatment path and ending point, are not.

Initial symptom severity. Initial symptom severity did moderate outcome in this model when specified as a covariate and treated as a fixed effect. It produced a strong effect, with an F (280.8) = 3.3, $p < .05$, and accounted for 57% of explained variance when added as a covariate versus recovery in general. It was also one of the few variables that produced not only a different intercept or starting point but a distinctive slope that was distinct from the general clinic population over time. Those people who began treatment with greater symptom severity had a slower and less consistent recovery path. This can be seen especially in the statistically significant bump or increase in symptoms at time 4 (i.e., 9 months).

Interpersonal functioning at baseline did moderate outcome in this model, when specified as a covariate and treated as a fixed effect. Interpersonal functioning in this case was operationalized as the IR subscale on the OQ at baseline (Time 1). An analysis of interaction effects was done between OQTOT change over time and OQ IR scores at baseline. First, interpersonal functioning at baseline did moderate outcome ($F[1, 170.4] = 38.5$, $p < .05$), accounting for 31.8% of explained variance. Second, a small interaction effect was discovered that approached, but did not reach, statistical significance: $F[4, 232.7] = 2.2$. This means that although those people with poorer initial relationships (in the form of higher scores on the IR subscale at baseline) tended to begin treatment more symptomatic (having higher OQTOT scores at baseline) than those without such IR subscores at baseline, their recovery path (i.e., slope) was not unique. The small uptick at time four (9 months), evident visually below, was not statistically significant.

DISCUSSION AND APPLICATIONS TO RESEARCH AND PRACTICE

The findings suggest that clients receiving psychodynamic psychotherapy make both clinically and statistically significant change over time as measured by the OQ 45.2. These findings add evidence that this historically under-researched and widely practiced form of psychotherapy is associated with clients reporting improved outcomes when examined broadly across practitioners and across heterogeneous diagnoses. It also suggests that a common course of recovery may exist across a large number of variables. This study adds important information to dynamic literature in the area of moderating variables of

outcome that have been limited in previous studies. It does so within a longitudinal context and uses an empirical measure. It also points to the importance of the first 3 months of treatment. All of this is significant in that many previous studies of psychodynamic treatment have been criticized for lacking the above: a longitudinal component, an empirical measure, and attention to moderating variables. This study adds to the existing literature by responding to the limits of these previous studies.

The findings point to statistically and clinically significant change on both the OQ total score and all three subscale scores (*symptom distress*, *IRs*, and *social role* functioning) over time. This change held true across broad diagnostic clusters (i.e., depression, anxiety, and adjustment) and whether someone had a personality disorder. This is in keeping with existing literature that psychodynamic psychotherapy is effective in the treatment of depression (Leichsenring, 2001), anxiety (Svartberg, Seltzer, & Stiles, 1998), and personality disorders (Leichsenring & Leibing, 2003).

Change over time on one particular subscale merits special attention. The OQ's Symptom Distress subscale allowed this study to speak to Seligman's (1995) call for future research to look at symptom change apart from clients' self-reporting on their more subjective sense of satisfaction with therapy. Subjects in this study did make statistically significant symptom change over time. This held true for not only the OQ total score (i.e., overall score) but for the symptom distress subscale. The Symptom Distress subscale produced an effect size of .38, the highest among the three subscales (which ranged from .28 to .38). Only the overall or OQ total score produced a stronger effect size (.41). In both cases, the majority of change occurred during the first 3 months. The 0 to 3 month interval in both cases proved to be the only discrete period in which statistically significant change occurred in relation to any other 3-month treatment interval.

This finding is similar to a recent finding by Lambert et al. (2003), who point to the importance of the first five sessions in forming an alliance and in producing some symptom change. Lambert et al. speaks to the need for early feedback on symptom change on the OQ, concluding that clients do not wait for an alliance to develop or for symptom change. His study used 1,020 clients and 49 therapists. Based on his findings, he concludes that (a) "ongoing feedback leads to better outcomes" (Cooper, 2003, p. 16) and that (b) "research tells us that clients who deteriorate in the first sessions don't eventually get better. . . . they leave therapy" (Cooper, 2003, p. 16). The finding in this study that the most change appears to happen in the first 3 months of therapy substantiates Lambert

et al.'s finding as to the importance and prognostic value of this early period. Finally, this study's findings also support the existing literature that suggests that demographic variables such as gender, race, and age do not tend to moderate outcomes (i.e., Barkham, Rees, & Shapiro, 1996; Sandell et al., 2000; Warner, 1998).

This study benefited from those traits associated with effectiveness studies. For instance, the data are from a naturalistic clinic setting, and uses 30 different treating therapists, with subjects (clinic clients) who come and go with a multitude of diagnoses. As can be seen in the descriptive analysis, nearly half (44.5%) of subjects in this sample presented with more than one diagnosis. Subjects were thus not screened out for people with single diagnoses in the way they often are in more traditional efficacy studies. This is important because it speaks to the potential generalizability of these findings to other clinic settings that are often characterized by clients with similar features (comorbidity, multiple treating therapists, etc.). Because this data collection is part of routine clinic procedures, subjects also were not paid and did not receive extra attention for participating in a study. This in itself may soften or minimize any expectancy effects such as change because of a desire to please the clinician or simply because of increased contact with providers.

A fairly uniform intervention was delivered in that this clinic provides a large amount of weekly supervision and ongoing training that are strongly dynamic in nature. Thus, although the study suffers from a lack of manualization or other treatment adherence measures, the clinic's strongly and uniquely psychodynamic identity, with its accompanying training and supervision, may help to buffer against the lack of manualization and adherence issues. This impression has also borne out recently in this author's interviews with all clinic staff about what constitutes the treatment, which revealed a great deal of uniformity in terms of how clinicians report they practice generally and in how they would approach treating a given diagnosis: major depression.

Limitations include a lack of specificity in the form of attention to individual diagnoses. This study did look at diagnostic clusters (i.e., mood, anxiety, and adjustment disorders). It did not look specifically at major depression or other distinct disorders. Such attention to individual diagnoses would be a useful direction for future research to see if individual diagnoses produce similar recovery trend lines, in studies with larger sample sizes, more frequent measures, and more power. This analysis was limited by the number of data points used (five). More data points would likely provide a more nuanced or detailed picture of the nature of this change.

The amount of missing data presented a particular set of challenges and limitations. The high degree of middle-phase missing data and the infrequent measures resulted in a loss of power. The power analysis was ex post facto or after the fact. See Wilkinson and the Task Force on Statistical Inference (1999) for discussions about the validity of and alternatives to such an analysis. Some have called for not using a power analysis in this way and for relying strictly on the confidence intervals themselves. There is not, however, uniform agreement on this.

The possibility of expectancy and placebo effects certainly exist as additional threats to internal validity. It is possible that clients sought to please their therapist with improving scores over time. However, this is somewhat tempered in that support staff administered the OQ, it was computer scored, and at this point in the clinic's research, the OQ was not formally reviewed with clients in most cases. It is possible, although, that clients would wish to see their experience over time as positive and to thus report improved outcomes. Placebo effects may thus exist and are even likely. These, however, are not limited to this type of therapy. They have been reported as a common factor in both medical and psychiatric outcomes broadly (Benson & Starke, 1996; Hubble, Duncan, & Miller, 1999). Additional threats include phenomena such as maturation and regression to the mean: the notion that people often improve or simply revert to their baseline, irrespective of any treatment or formal intervention. This is especially as risk in a time series design such as this, as Campbell and Stanley (1963) note.

The next steps for research in this area are related. They include the use of empirical measures such as the OQ at more frequent intervals (monthly measures) and with less missing data. This would allow for greater power, a more detailed picture of what this change looks like over time, and closer comparisons with existing literature on recovery trends (Lambert et al., 2003). Future comparisons and replications would be aided by researchers using a common or shared battery of measures (Miller, Luborsky, Barber, & Docherty, 1993). This would allow studies using the OQ to compare results with studies using the OQ versus comparing the OQ to instruments such as the social adjustment scale.

Future studies would also benefit from the use of treatment fidelity measures. Although supervision and a fairly theoretically homogenous approach was used in this study, future studies would benefit from an even better defined treatment such as the use of either or both (a) a psychodynamic treatment manual and (b) measures of fidelity to the treatment itself. This could be done through supervision during the study itself or thorough independent

assessors rating the degree of adherence via video or audio tape. If the above could be done in relation to particular diagnoses, the design would be especially strong, reaping the benefits of both (a) effectiveness research in a naturalistic setting with (b) tighter design controls. This would also speak to remaining questions in the literature as to which treatments are best for which people with which diagnoses.

This study similarly lends support for a more brief and focused dynamic treatment as a viable or legitimate form of treatment. Of course, other literature exists speaking to the length of treatment as being diagnosis-specific. Whereas diagnoses such as borderline personality disorder (Beck, 1995; Linehan, 1993) often call for longer term treatments (i.e., greater than a year), this study lends support to the idea that for more acute diagnoses, a time-limited dynamic therapy can be effective. This is a finding in line with both Strupp and Binder (1984) and with the American Psychological Association's list of empirically supported treatments, both of which discuss brief or time-limited dynamic therapies with good research behind them. Agencies, too, can use these convergent findings to inform their own internal policy decisions as to how they think about limiting or dispensing therapy as a resource.

Last, the findings from this study have several implications for how psychodynamic theory is conceptualized. First, psychodynamic theory has often historically claimed its authority by way of attributed authority. It has stayed in the realm of theory, without its tenets being empirically tested (Stone, 1997). This study, which used an empirical measure, lends weight to the claim that dynamic therapy is effective, when tested. Second, the finding that the most significant change (on all OQ scales) happens within the first 3 months of a year-long therapy points to the idea of an initial therapeutic window. This finding has several important implications itself. Not only does it point to the importance of initial improvement but it challenges the psychodynamic idea that people need to get worse before they get better. That is, in this study, improvement happens up front versus later. This challenges the idea of an incubation theory, in which change is hypothesized to begin later in or even after the therapy ends. A quantitative study by Huber, Klug, and von Rad (2002) reached a similar conclusion (they did not find evidence for an incubation period in dynamic therapy).

The idea that longer is better is also not necessarily supported in that although length of treatment did moderate outcome in this study, it did so by affecting where clients began treatment on the OQ versus how they recovered. No interaction effects were found between length of treatment and improvement at Time 5 (1 year of treatment). This may argue for a more targeted, more brief intervention, consistent with what this clinic offers (the average number of sessions for this clinic's entire population is 23; the modal number for this data set was 27). However, the limits of this design, along with the degree of attrition, should temper any final conclusion on this question based on this data. The findings do begin to suggest, although, that psychodynamic therapy can be offered in both an effective and cost-effective way in a real-life clinic setting and that it may not be the black hole (i.e., expensive therapy without benefit) that many have feared it might be.

REFERENCES

American Psychological Association. (1993). *A report adopted by the Division 12 Board*. Washington, DC: Division 12 Office of the American Psychological Association.

Anderson, E. M., & Lambert, M. J. (1995). Short-term dynamically oriented psychotherapy: A review and meta-analysis. *Clinical Psychology Review*, *15*(6), 503-514.

Asay, T., Lambert, M., Gregersen, & Goates, M. (2002). Using patient-focused research in evaluating treatment outcome in private practice. *Journal of Clinical Psychology*, *58*, 1213-1225.

Barber, S., & Lane, R. C. (1995). Efficacy research in psychodynamic therapy: A critical review of the literature. *Psychotherapy in Private Practice*, *14*(3), 43-69.

Barkham, M., Rees, A., & Shapiro, D. (1996). Outcomes of time-limited psychotherapy in applied settings: Replicating the second Sheffield psychotherapy project. *Journal of Consulting & Clinical Psychology*, *64*, 1079-1085.

Beck, J. S. (1995). *Cognitive therapy: Basics and beyond*. New York: Guilford.

Benson, H., & Stark, M. (1996). *Timeless healing: The power and biology of belief*. New York: Scribner.

Campbell, D. T. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Congdon, R. (2001). *Optimal design: Freeware/shareware program for calculating power within a mixed or hierarchical linear model*. Retrieved on April 20, 2004, from www.jiscmail.ac.uk/lists/multi-level.html

Cooper, G. (2003). Clinician's digest. *Psychotherapy Networker*, March/April, 16.

Doidge, N. (1999). How to preserve psychoanalysis: Introduction to Gunderson and Gabbard. *Journal of the American Psychoanalytic Association*, *47*, 673-676.

Freedman, N., Hoffenberg, J., Vorus, N., & Frosch, A. (1999). The effectiveness of psychoanalytic psychotherapy: The role of treatment duration, frequency of sessions, and the therapeutic relationship. *Journal of the American Psychoanalytic Association*, *47*, 741-771.

Gabbard, G. (2000). *Psychodynamic psychiatry*. Washington, DC: American Psychiatric.

Gunderson, J. G., & Gabbard, G. O. (1999). Making the case for psychoanalytic therapies in the current psychiatric environment. *Journal of the American Psychoanalytic Association*, *47*, 679-704.

Gabbard, G. O., Gunderson, J. G., & Fonagy, P. (2002). The place of psychoanalytic treatments within psychiatry. *Archives of General Psychiatry*, *59*, 505-510.

Gunderson, J., & Gabbard, G. (1999). Making the case for psychoanalytic therapies in the current psychiatric environment. *Journal of the American Psychoanalytic Association*, *47*, 679-704.

Huber, D., Klug, G., & von Rad, M. (2002). The Munich psychotherapy study: A process-outcome comparison between psychoanalysis and psychodynamic psychotherapies. In M. Leuzinger-Bohleber & M. Target (Eds.), *Outcome of psychoanalytic treatment: Perspectives for therapists and researchers*. New York: Whurr.

Hubble, M. A., Duncan, B. L., & Miller, S. D. (1999). *The heart and soul of change: What works in therapy.* Washington, DC: American Psychological Association.

Kernberg, O. F. (1991, May 11). *The current status of psychoanalysis*. Key note address presented at the Annual meeting of the American Psychoanalytic Association, New Orleans, Louisiana.

Kopta, S., Howard, K., Lowry, J., & Beutler, L. (1994). Patterns of symptomatic recovery in psychotherapy. *Journal of Consulting & Clinical Psychology,* 62, 1009-1016.

Lambert, M. J., Gregersen, A. T., Burlingame, G. M., & Marush, M. E., (Ed.) (2004). *The use of psychological testing for treatment planning and outcomes assessment: Vol. 3: Instruments for adults* (3rd ed). Mahwah, NJ: Lawrence Erlbaum.

Lambert, M. J., Finch, A. E., & Maruish, M. E. (Eds.). (1999). *The use of psychological testing for treatment planning and outcomes assessment* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Lambert, M. J., Hansen, N. B., Umpress, V., Lunnen, K., Okiishi, J., & Burlingame, G. M. (1996) *Administration and scoring for the OQ-45.2*. Wilmington, DE: American Professional Credentialing Services.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science & Practice*, *10*(3), 288-315.

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

Leichsenring, F. & Leibing, E. (2003). The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders: A meta-analysis. *American Journal of Psychiatry*, *160*, 1223–1232.

Linehan, M. M. (1993). *Cognitive behavioral treatment of borderline personality disorder*. New York: Guilford.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage.

Miller, N. E., Luborsky, L., Barber, J., & Docherty, J. (1993). *Psychodynamic treatment research: A handbook for clinical practice*. New York: Basic Books.

Mueller, R. M., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment*, *70*, 248-262

Nathan, P. E., Stuart, S. P., & Dorn, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla & Charybdis. *Psychological Bulletin*, *126*, 964-981.

Exploratory/developmental grants for psychosocial treatment research. (1993). Program Announcement No. 93-093. *National Institute of Mental Health Guide*, *22*(21), **PAGES**.

Olfson, M., Marcus, S. C., Druss, D., & Pincus, H. A. (2002). National trends in the use of outpatient psychotherapy. *American Journal of Psychiatry*, *159*, 1914-1920.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge.

Raudenbush, S. W., & Xiao-Feng, L. (2001). Effects of study duration, frequency of observation, and sample size on power studies of group differences in polynomial change. *Psychological Methods*, *6*, 387-401.

Sandell, R., Bloomberg, J., Lazar, A., & Carlsson, J. (2000). Varieties of long-term outcome among patients in psychoanalysis and long-term psychotherapy: A review of findings from the Stockholm outcome of psychoanalysis and psychotherapy project (STOPPP). *International Journal of Psychoanalysis*, *81*, 921-942.

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

Stone, A. (1997). Where will psychoanalysis survive: What remains of Freudianism when its scientific center crumbles? *Harvard Magazine*, January-February, 35-39.

Strupp, H., & Binder, J. L. (1984). *Psychotherapy in a new key: A guide to time-limited dynamic psychotherapy.* New York: Basic Books.

Svartberg, M., Seltzer, M., & Stiles, T. (1998). The effects of common and specific factors in short-term anxiety-provoking psychotherapy: a Pilot process-outcome study. *Journal of Nervous and Mental Disease*, *186*, 691-696.

Svartberg, M., & Stiles, T. C. (1991). Comparative effects of short term psychodynamic psychotherapy: A meta-analysis. *Journal of consulting and clinical psychology*, *59*, 704-714.

Vaughan, S. C., Marshall, R. D., Mackinnon, R. A., Vaughan, R., Mellman, L., & Roose, S. P. (2000). Can we do psychoanalytic outcome research? A feasibility study. *International Journal of Psychoanalysis*, *81*, 513-527.

Wallerstein, R. S. (1986). *Forty-two lives in treatment: A study in psychoanalysis and psychotherapy*. New York: Guilford.

Warner, L. K. (1998). *A comparison of dose-response curves in cognitive behavioral and psychodynamic psychotherapy*. Doctoral dissertation, St. John's University.

Wilkinson, L., and the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals. *American Psychologist*, *54*, 594-604.