Psychodynamic Psychotherapy: A Quantitative, Longitudinal Perspective

David J. Roseborough,[1,2] Jeffrey T. McLeod,[3] and William H. Bradshaw[4]

Abstract

This effectiveness study examined the course of treatment longitudinally and outcomes associated with psychodynamic psychotherapy for a sample of 1,050 people undertaking this treatment in a community setting, over the course of 4 years, at 3-month intervals, using the Outcome Questionnaire (OQ)-45.2. The authors used multilevel modeling to look at the nature of change over time and at potentially meaningful moderating variables. Results show a robust general improvement, though a more moderate one than described in recent meta-analyses including primarily prospective studies. The treatment was followed by broad improvements, over time, with a general trend and few notable interaction effects. The treatment involved little deterioration, particularly in the first year. Subgroup analysis suggested that (a) clients with more initial symptom severity showed greater improvement and a unique course of recovery with (b) clients who stayed over a year constituting a potentially unique subgroup.

Introduction

Psychodynamic psychotherapy is currently enjoying something of a renaissance. Leichsenring and Rabung's (2008) meta-analysis evaluating the effectiveness of long-term psychodynamic psychotherapy, published in the *Journal of the American Medical Association* (JAMA), garnered national media attention and was referenced in publications such as *Newsweek* and the *Los Angeles Times*. It was commented on by leaders within medicine and psychiatry including Dr. Aaron Beck. More recently, another comparable review of meta-analyses, published by Shedler (2010), appeared in *American Psychologist*. Both analyses pointed to favorable and relatively strong treatment effects associated with psychodynamic treatment, both short- and long term.

These studies have met with some criticism. Authors such as Beck and Bhar (2009) and Littell and Shlonsky (in press) have questioned their methodology and have criticized these authors on methodological grounds as overstating the strength of their findings. Both meta-analyses referenced above reviewed predominantly prospective studies. Neither emphasized psychotherapies as they are naturally practiced in "real life" community settings. And yet, as Schilling (2010) has noted, "methodological requirements often result in studies that bear little resemblance to the circumstances of agency-based practice in human service settings" (p. 550). In contrast, Schilling also notes that "the bulk of intervention outcome findings in realms of interest to social workers (and other mental health practitioners) will be derived from non-experimental studies.

Fortunately, design and statistical advances have enhanced the interpretability of such intervention studies" (p. 550). This exploratory study was a collaborative effort between university-based researchers and practicing clinicians in a community mental health setting. It sought to further contribute to knowledge about the effectiveness of psychodynamic psychotherapy, conducted in the tradition of effectiveness and intervention research, done (a) in a naturalistic, community setting (in contrast to recent published meta-analyses done with often brief, prescribed treatments, more in the tradition of efficacy studies), (b) with an empirical measure (in this case, the OQ-45.2), (c) using a quantitative, longitudinal design, and (d) to offer the results as a point of comparison with recent published reports mentioned above, particularly in relation to the effect sizes (ES) associated with this treatment. Whereas in randomized clinical trials, participants are often treated for equivalent lengths of time and risk dropout due to treatment incompatibility, a design such as this one allows each participant to encounter treatment on their own terms and without a

[1] School of Social Work, St. Catherine University, St. Paul, MN, USA
[2] The University of St. Thomas, St. Paul, MN, USA
[3] Argosy University, Eagan, MN, USA
[4] University of Tennessee Knoxville, Knoxville, TN, USA

Corresponding Author:
David J. Roseborough, School of Social Work, University of St. Thomas, 2115 Summit Avenue, St. Paul, MN 55105, USA
Email: djroseboroug@stthomas.edu

predetermined length of treatment. This study also took advantage of repeated measures in order to get a clearer picture of the nature and effects of attrition over time.

Literature Review

Some Preliminary Context

Calls for this sort of research have come not only from external audiences such as the National Institute of Mental Health and the American Psychological Association (APA, 1993), but from the psychoanalytic community as well. As early as 1993, stalwart psychoanalysts such as Kernberg called for an increased attention to research by the psychoanalytic community. Kernberg lamented what he described as "its lack of interest in empirical research" (p. 46) and called for more interdisciplinary scholarship. He wrote, "too many psychoanalysts know too little about current psychoanalytic empirical research; the fertilizing effects of clinicians and researchers working jointly on research projects cannot be overestimated" (p. 49). Kernberg (1993) and Nobel laureate Kandel have called for psychoanalysts to give a portion of their time to research (Rose, 2007). Another prominent analyst, Doidge, similarly spoke to the ways practitioners in the psychoanalytic community have benefited from research, while at the same time discounting its importance. He wrote, concisely, "we reject empirical outcome research" (1999, p. 674) as somehow superficial. While many in the psychodynamic community have argued that this dialogic practice (Fraser & Galinsky, 2010) seeks "structural" or character change that does not lend itself well to or is somehow beyond measure, Doidge offered a more research-friendly perspective writing "just as analysts went from being 'id analysts' to 'ego analysis,' from starting at the depths to starting at the surface, so too must we be willing to do more work at the surface in our research." He noted the importance of beginning with even crude measures such as "number of sessions, visits to the emergency room (and) incidents of self-harm, both before and after treatment" (pp. 675–676). This claim of dynamic therapy being difficult to measure has been made more recently as well, with a September, 2010 issue of the *Harvard Mental Health Letter* reporting "randomized controlled studies are the ideal way to evaluate treatments in medicine, but psychodynamic therapy, with its individualized technique and complex aims, has not lent itself readily to this type of study. It is not surprising that it has taken longer for researchers to develop and validate rigorous methods for studying the treatment" (p. 3).

Findings, to Date

While psychodynamic therapy is the first and oldest among the most established forms of psychotherapy, its evidence base has historically relied on case studies, patient narratives, and impressions from the analyst's perspective (i.e., Wallerstein's *42 Lives*). More recent and better controlled quantitative studies have, however, pointed to its potential efficacy with a wide range of presenting problems (Leichsenring & Leibing, 2003; Milrod et al., 2007). A few recent systematic reviews and

meta-analyses here are particularly noteworthy. Leichsenring and Rabung's 2008 meta-analysis reviewed and summarized 23 studies of long-term psychodynamic psychotherapy (defined as therapies with over 50 sessions), included an overall sample size of 1,053 with a mean of 151 sessions (a median of 73), and used both manualized and nonmanualized treatments. This review found a large average ES of 1.8 in between-group comparisons. This study called for future studies to give greater attention to specific diagnoses and to comorbidity. It included only prospective designs.

Other reviews of long-term psychodynamic psychotherapy have pointed to its effectiveness as well. A more recent meta-analysis by de Maat, De Jonghe, Schoevers, and Dekker (2009) reported similar findings with an even larger sample. These authors conducted a systematic literature review including 27 studies ($N = 5,063$) pertaining to long-term treatment (defined akin to Leichsenring & Rabung, 2008, as 50 or more sessions and one or more years) and found an average ES of .78 at termination and an even stronger ES (.94) at follow-up. In their review, psychoanalysis and psychodynamic therapies did differ in their overall effectiveness, with analysis yielding a stronger ES (1.18 at follow-up) than psychodynamic treatment. Symptom change was found to be more amenable to change than personality or character change, which reportedly took longer and changed less (ES = .76). This study added attention to the question of whether randomized controlled trials versus effectiveness studies would yield different results, finding they did not in a statistically significant way.

Other authors have looked specifically at brief dynamic treatment. Shedler's (2010) review of brief dynamic therapy (40 or fewer sessions) pointed to ESs nearly as large as those found by Leichsenring and Rabung, but with a smaller number of sessions and also began to delineate some of the core features central to this treatment. Driessen et al. (2010) gave specific attention again in the form of a meta-analysis ($N = 23$ studies, with a sample size of 1,365, spanning 40 years) looking specifically at the treatment of major depression. In keeping with both meta-analyses above, Driessen et al. (2010) found large ESs for within-group (change over time) estimates (ES = 1.34), and smaller, but still robust ESs for major depression in between-group comparisons (.69 at posttest, finding short-term psychodynamic psychotherapy (STPP) superior to control conditions). While some treatments appeared superior at posttreatment, these differences were negligible at 3 months' follow-up. Interestingly, supportive and expressive treatments were found to be equally effective (p. 25). This review called for more attention to open-ended treatments and to subgroup analysis for potential treatment moderators. It also suggested that treatment gains appear to be maintained, finding a small, but statistically insignificant improvement after treatment, describing "a very small and non-significant decrease in depression scores at follow up when compared to posttreatment" (p. 32).

These findings have not been without controversy. Cuijpers, van Straten, Bohlmeijer, Hollon, and Andersson (2010) conducted a meta-analysis and found that ESs for the psychotherapeutic treatment of depression, specifically, tend to be

overstated, and inferior to medication, regardless of theoretical approach. In relation to psychodynamic treatment, Beck and Bhar (2009) and Bhar et al. (2010) have criticized several of the methods associated with Leichsenring and Rabung's (2008) study, in particular, the age of many of the studies, the loose, heterogeneous nature of many of the comparison groups along with other methodological flaws, including what they deem the miscalculation of ESs and thus overstating the significance of their findings. Others such as Littell and Shlonsky (in press) have argued that the standards used to conduct and report results from their meta-analysis do not meet current, accepted conventions as outlined by assessment of multiple systematic reviews (AMSTAR). These available studies have similarly often neglected other emerging research standards such as an intent to treat analysis, a posttreatment assessment of diagnosis, and attention to the nature of attrition.

While studies such as those referenced above have begun to give attention to moderating variables in this form of psychotherapy, that is, looking at treatment outcomes specific to diagnoses such as major depression (for a similar review in relation to dysthymia—see Svanborg, Wistedt, & Svanborg, 2008), in the way called for by Leichsenring and Rabung (2008), few have looked explicitly at *the longitudinal* nature of psychodynamic psychotherapy, or at what happens during, versus simply before and after treatment, using a quantitative methodology (for an exception, see Bond & Perry, 2004, regarding the impact of psychodynamic psychotherapy on personality defense structure). A recent German, longitudinal study published by Puschner, Kraft, Kachele, and Kordy (2007) did report on the longitudinal course of change in psychodynamic psychotherapy ($N = 932$), looking at privately insured outpatients, using the Symptom Checklist (SCL)-90 and a subscale of the OQ-45.2 over five measurement points. This study similarly used a mixed model analysis. It found that the majority of change occurred early in the treatment, with an average 30% reduction of symptoms occurring even before the first session. It did not find significant differences in outcome between psychodynamic and long-term psychoanalytic treatment. It did not provide formal ESs. The authors gave some attention to moderating (predictor) variables by way of regression analysis and found that initial symptom severity exerted significant influence on outcome with people reporting more initial symptom severity showing the most significant and sustained progress in therapy. The present study has several parallels with Puschner et al.'s (2007) study. We sought to add to this emerging literature by giving specific attention to longitudinal change, with more and more proximate time points, to subgroup analyses (i.e., attention to particular diagnoses and to attrition), to potential treatment moderators (those variables affecting outcome), and to associated ESs.

Method

Research Design

This study used a preexperimental time-series design, including a baseline and 15 subsequent time points (x-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0; Campbell & Stanley, 1963). Participants were given the OQ-45.2 at baseline (intake) and at 3-month intervals thereafter. All participants with two or more OQ scores on record were included in this sample.

Procedure

Measure. The OQ-45.2 is a 45-item client-administered questionnaire developed by Lambert et al. (2004) specifically to measure outcomes particularly relevant to psychodynamic therapy, though it can and has been used more broadly, including mandated use statewide in Utah by state-funded programs. It uses a 5-point Likert-type scale. The OQ provides both an overall score (ranging from 0 to 180) as well as three subscales (measuring symptom distress, interpersonal relations, and social role functioning). Lower scores represent less severity and higher scores represent more psychiatric distress, with 63 representing a clinical cutoff or measure of caseness. Clients scoring above 63 are seen as warranting treatment in contrast to community norm scores, which average 45. The instrument has been normed on psychiatrically well community populations, as well as on students in college counseling centers, employee assistance program (EAP) clients, clients in outpatient mental health centers, and with psychiatric inpatients. It has also been tested across gender, race, and with various ethnicities. Its alpha coefficients for internal consistency range between .84 and .93 for OQ Total scores (Lambert et al., 2004). Test–retest reliability for the OQ Total score was reported as ranging from .78 to .84 by Umphress, Lambert, Smart, and Barlow (1997). Concurrent validity ranged from .53 to .86 when correlating the OQ-45 Total score with a wide variety of psychological outcome measures including the Beck Depression Inventory, the Zung Depression Scale, the Taylor Manifest Anxiety Scale, the State-Trait Anxiety Inventory, and the Social Adjustment Scale (Lambert et al., 2004; Umphress et al., 1997). Vermeersch, Lambert, and Burlingame (2000) have provided evidence of the construct validity of the OQ-45 Total for measuring psychotherapy outcomes. The OQ is able to speak not only to recovery (a score in the range of community norms or below caseness) but to clinically reliable change and to deterioration (i.e., a 14 or greater point increase in total score). The OQ was administered regularly at 3-month intervals to each person remaining in treatment. It was distributed by a support staff person to the client at the session corresponding most closely to 3 months since the administration of their last OQ. This time frame was identified for each participant using a computer program called Therapist Helper. The OQ was reviewed in session with each client by their individual therapist.

The setting. The clinic chosen is an adult, outpatient mental health clinic located in St. Paul, Minnesota. Founded in 1954 as one of the first community mental health centers in the country, it is a multidisciplinary clinic, employing master level clinical social workers, master and doctoral level psychologists, and psychiatrists. The clinic is an APA-approved internship training site. It provides psychodynamic, relationally based psychotherapy, and the judicious use of medication. The clinic,

at the time of study, did not offer medication management apart from therapy. That is, all clients receiving medication at the clinic were concurrently clients in therapy at the same clinic. The clinic had 16 clinicians and provides graduate training for an average of 15 psychologists, psychiatrists in residency, and clinical social workers. It has a central training mission and provides continuing education on psychodynamic topics for mental health practitioners in the Twin Cities metropolitan area.

It has, since the beginning of its inception, operated with a clinic-wide psychodynamic, relationship-based practice orientation. Much akin to the Mayo model of care, clients are considered *clinic* clients. The clinic operates from a strongly collaborative, interdisciplinary model. Psychodynamic psychotherapy is seen as a common language spoken at the clinic and cases are teamed by way of weekly interdisciplinary supervision and monthly individual supervision. The clinic is relatively unique in its lateral structure. Psychiatrists are on-site and are seen as an integral part of the treatment team. They see clients for both medication and psychotherapy. Clients are primarily self-referred. Sources of referral include insurers, primary care providers (including visiting nurses), case managers, and former clients. The clinic is well known as a long-standing community mental health clinic with a sliding scale. Nearly 70% of clients are insured, while 30% are eligible for clinic-subsidized care, using a sliding scale.

The treatment. Gabbard (2008) has described psychodynamic treatment in training texts, such as the *Textbook of Psychotherapeutic Treatments*. In this text, he points to key components of the treatment such as an emphasis on understanding avoidance, the identification of relational patterns, a focus on relationships and relational styles, an emphasis on affect, attention to history, and the use of free association. All of this is done in the context of a warm and supportive therapeutic alliance, which Gabbard refers to as the "envelope" (Harvard Mental Health Letter, 2010, p. 2) in which this treatment occurs. This can be in some ways contrasted with traditional notions of analytic neutrality. We have described this particular clinic's treatment in more detail in publications such as Bradshaw, Roseborough, Pahwa, & Jordan (2010) and in a 20-page unpublished, beginning treatment manual. The clinic's treatment is very akin to models described recently in works such as Summers and Barber's (2009) book *Psychodynamic therapy: A guide to evidence-based practice.*

While the clinic treatment here was not formally manualized, a Swedish study by Vinnars, Barber, Noren, Gallop, and Weinryb (2005) casts some doubt about the necessity of doing so. Vinnars et al. offered a very comparable study to this one in several ways. It evaluated a year of treatment with an average of 21 (vs. 24) sessions provided by experienced dynamic practitioners, in an open-ended fashion, in a community setting, using a similar method (mixed model analysis of variance, or ANOVA), and found "no difference in effect between (manualized vs. nonmanualized) treatments" (p. 1933). "Manualized supportive-expressive psychotherapy was as effective as nonmanualized community-delivered psychodynamic therapy conducted by experienced dynamic clinicians," in their study in relation to treating people with personality disorders. Their study also spoke to the difficulty of implementing control groups in community settings for ethical reasons and due to practical limitations. This comparison is not meant to diminish the value of manualized treatment studies, but simply speaks to the importance of studying both manualized and nonmanualized community-based treatments.

Sample (N = 1,050) and inclusion criteria. The sample consisted of the clinic population between 2003 and 2008. It included all adults (ages 18–83) undertaking psychodynamically oriented psychotherapy at the clinic during this time period, who completed at least two complete OQs. "Complete" is defined according to the *OQ Scoring Manual* (Lambert et al., 2004) as a measure with at least 90% of items completed. The scoring manual also advises that missing data be treated by substituting the mean for that domain if 90% of the measure is complete. While mean substitution is certainly not the only statistical strategy for missing data, nor is it necessarily the best option, (i.e., replacement with the mean decreases the variability of outcome measures, which artificially deflates the standard errors of parameter estimates). The decision to use mean substitution was made as the best practical approach and was considered a defensible compromise especially considering the advice of the OQ-45 scoring manual. Three month measures were instituted for all such clients beginning in January of 2003 and are part of routine data collection for the clinic. All data were aggregated and deidentified for this study in keeping with Health Insurance Portability and Accountability Act, 1996 (HIPAA) requirements. The study also received formal institutional review board (IRB) approval.

Statistical analysis. The authors chose a two-level, linear mixed model for the primary analysis, where scores were seen as nested within people, with attention to potential moderator variables such as medication as interclass effects. Multilevel modeling was chosen due to the longitudinal design and the unbalanced nature of the data. Time was treated as a fixed and random effect while other variables (medication, gender, etc.) were treated as fixed effects. People were allowed to vary in both their intercept or where they began treatment (all participants had a baseline score), and in their slope, or how they recovered over time. This resulted in the best model fitting, with lower log likelihood values than competing models (likelihoods are measures of the goodness of fit of parameter estimates in a specified model to the raw data; smaller log likelihoods scaled negatively, indicate a better fit when comparing one model to another). Stepwise comparisons were run as well. We chose not to fit people to a curvilinear notion of change (linear, quadratic, or exponential), but wanted to see what happened uniquely at each time point, concluding based on previous research by Lambert et al. (2004) and others that change in psychotherapy is not linear and that wide variation over time is often the norm rather than the exception. We sought to model the specifics of this change.

Table 1. Descriptive Statistics for Number of Sessions

M		20.67
Mdn		9.00
Mode		3
SD		34.103
Range		560
Min		0
Max		560
Percentiles		
	25	3.00
	50	9.00
	75	24.00

Note. M = mean; Mdn = median; min = minimum; max = maximum; SD = standard deviation.

Statistical Analysis

The growth curve parameters were estimated using multilevel modeling. Following the notation style of Raudenbush and Bryk (2002), the multilevel model used in the primary analysis was:

Level 1 Model

$$Y_{it} = \beta_{0i} + \beta_{1i}t + r_i.$$

The Level 1 model states that the OQ Total score for person i at time t is a linear sum of an individual's unique intake severity modeled by the intercept term, their rate of improvement, modeled by the slope term, the elapsed time t, and the residual error term for person i denoted by r_i.

Level 2 Model

$$\beta_{0i} = \gamma_{00} + u_{0i}$$

$$\beta_{1i} = \gamma_{10} + u_{1i}.$$

The Level 2 model states that the individual's intake severity β_{0i} and rate of improvement β_{1i} are random variables centered on fixed constants for the intercept and for the slope, plus random deviations from each of those fixed constants for person i, denoted by u_{0i} and u_{1i}, respectively. The random deviations establish the individual's unique location relative to the average parameter value.

The levels can be combined to reveal the full model:

$$Y_{it} = \gamma_{00} + u_{0i} + \gamma_{10} + u_{1i}t + r_i.$$

The model has four parameters consisting of two fixed effects $(\gamma_{00}, \gamma_{01})$ and two random effects (u_{0i}, u_{1i}) across 16 measurement time periods with 1,050 patients. As can be seen by its linear form, the growth model has the appearance of an ordinary least squares multiple regression model, but the mixed model has a more complex variance/covariance or error structure that must be taken into account in estimating the parameters.

Multilevel modeling was employed using the Mixed Modeling module of SPSS 17.0. SPSS implements hierarchical linear modeling (HLM) with a variety of model specification options, covariance structure options, and parameter estimation options.

The software offers data management modules that simplify requisite data manipulation from case sequence to time sequence. The model specification had two fixed and two random effects for the intercept and slope. Also, the researchers specified the repeated measures on the OQ Total series as a first order autoregressive heterogeneous structure. This structure is appropriate for time-series data since the causal structure of sequential data tends to render neighboring points correlated. Although the relative proportion of variance accounted for by the autoregressive component turned out to be small, it is nevertheless a reasonable choice.

Power analysis. Statistical power was analyzed using Optimal Design version 2.0 (Spybrook, Raudenbush, Congdon, & Martinez, 2009), written by Stephen Raudenbush and made available through the Scientific Software International website (2010). Statistical power in HLM growth curve models is influenced by the sample size, the number of repeated measures, and the anticipated ES. The sample size of 1,050 combined with a repeated measure with 16 periods yielded an estimated statistical power of .99 to capture a moderate ES. This result should be interpreted with some caution, though, in that Bickel (2007) and others have pointed to both the complexity of applying a power analysis to a multilevel design, particularly in light of differing sample sizes across levels and the presence of intraclass correlations (Bickel, 2007), and have urged caution in doing so. Others have pointed to the complexity of applying this method, in particular, to longitudinal designs, where a power analysis must "take into consideration within-subject variance" (Vinnars et al., 2005), p.1935).

Results

Descriptive Statistics

Average number of sessions per client. The mean number of sessions for the sample of 1,050 participants was 20.67, but this estimate is influenced by a strong positive skew in the distribution, as evidenced by the much smaller median of 9.0 sessions. Table 1 reveals that the 25th percentile of the sample is 3.0 sessions and the 75th percentile is 24 sessions. A window of 16 three-month periods (48 months total) adequately capture the full range of treatment effects including outliers.

Major diagnoses represented. Table 2 shows the breakdown of *DSM-IV* (*Diagnostic and Statistical Manual of Mental Disorders*, Fourth edition) major diagnostic groupings. The list is dominated by major depressive disorder (24.4%), dysthymic disorder (17.8%), adjustment disorder (21.1%), and anxiety disorder (15.8%).

Age and gender. The sample was made up predominantly of females (66%) and predominantly (both male and female) in the age range of 30 to 50 (54%). Table 3 gives a detailed breakdown of frequencies within gender by age, showing that the modal category consisted of females aged 30 to 40.

Table 2. Distribution of Cases by *DSM-IV* Groups

DSM IV Grouping	F	%
Schizophrenia, dementia, psychotic	17	1.6
Major depression	256	24.4
Bipolar	34	3.2
Anxiety disorders	166	15.8
Dysthymic disorder	187	17.8
Personality disorder	11	1.0
Sexual and substance	19	1.8
Adjustment disorders	222	21.1
Depression NOS	64	6.1
Impulse/identity disorder	6	0.6
ADHD	7	0.7
Deferred/unspecified	5	0.5
Missing	56	5.3
Total	1050	100.0

Note. ADHD = attention deficit hyperactivity disorder; NOS = not otherwise specified.

Nevertheless, there is a wide variety of participants crossing all levels of gender and age, which strengthens the generalizability of these findings.

Ethnicity. The distribution of self-reported ethnic status is Caucasian (73.6%), Hispanic (11.5%), Multiracial or not specified (7.8%), African American (5.5%), Asian (1.0%), and American Indian (0.6%; Table 4). This distribution reflects considerably more diversity than would be expected from the ethnic mixture of the State of Minnesota (2001), which shows the mixture to be Caucasian (91.2%), Hispanic (1.9%), African American (3.1%), Asian (2.7%), and Native American (1.2%).

Severity in comparison to national norms. Initial assessments on the OQ Total revealed an elevated mean and standard deviation ($M = 75.08$ and $SD = 24.04$). One can compare these figures to benchmark norms given by Mauish (2004, p. 203) for the General Community ($M = 42.5$, $SD = 17.3$), a Counseling Center ($M = 67.6$, $SD = 20.7$), a Community Mental Health Clinic ($M = 80.80$, $SD = 26.5$), and an Inpatient Unit ($M = 99.9$, $SD = 28.7$). The sample statistics on the OQ Total for this study fall nearest to the Community Mental Health Clinic norms, which is where they would be expected to fall. Both the mean OQ Total *and* the standard deviation are consistent with the clinical norm groups, less so with the counseling norms, and not consistent with the normal control sample.

Caseness. The OQ provides a measure of caseness, which serves as a clinical cutoff score (Lambert et al., 2004). The instrument is normed such that scores above or equal to 63 are seen as clinically significant and as warranting treatment. Sixty eight percent ($n = 717$) of this sample met criteria for caseness at intake. A smaller number and percentage ($n = 592$ of 1,050, or 56.4%) met criteria at 3 months. Each subsequent 3-month interval pointed to a similar and relatively stable percentage of remaining participants meeting criteria for caseness

Table 3. Age by Gender Composition

Gender	Age	f	%
Male			
	<30	47	13.2
	30–40	96	26.9
	40–50	92	25.8
	50–60	80	22.4
	Over 60	42	11.8
	Total	357	100.0
Female			
	<30	137	19.8
	30–40	210	30.3
	40–50	164	23.7
	50–60	118	17.0
	Over 60	64	9.2
	Total	693	100.0

Table 4. Ethnic Distribution

Ethnic Self-Report	f	%
Caucasian	773	73.6
Hispanic	121	11.5
Multiracial or unspecified	82	7.8
African American	58	5.5
Asian	10	1.0
Native American	6	0.6
Total	1,050	100.0

throughout the first 4 years of treatment, hovering between 56% and 59%. This is in keeping with the overall recovery trend, described below.

OQ Total Score

Severity at intake. The intercept estimate indicates the severity on the OQ Total instrument at intake, which is the base measurement. The *fixed* coefficient was 71.82, $t(1039.865) = 106.29$, $p < .001$, 95% confidence interval (CI) = [70.5, 73.1]. The variance of the *random* intercept was 384.96, the $SD = 19.62$. This suggests that the majority of participants, according to the normal reference distribution, would fall within one standard deviation of the average severity, or somewhere in the range of 52.2 to 91.44.

Rate of growth. Participants were measured at intake, and then once every 3 months subsequently using the OQ Total score. The rate of growth indicates how rapidly participants improved between measurement periods. Therapeutic growth, estimated from the γ_{10} fixed coefficient was -1.39, $t(239.242) = -9.28$, $p < .001$, 95% CI = [-1.69, -1.09]. This number is negative because a decrease in the OQ Total score indicates *improvement* in psychological functioning. Thus, the model stipulates that participants improve by -1.39 points on the OQ Total score every 3 months, so that after 9 months of psychodynamic psychotherapy or three measurement periods

Table 5. Aggregate Growth Trajectory on the OQ Total Scores

Period	N	M	SD	Mdn	Range
1	1,050	75.08	24.04	76.00	153
2	1,050	68.09	24.74	67.00	165
3	683	69.46	24.13	68.00	165
4	496	68.61	24.67	68.00	148
5	375	66.90	23.73	67.00	148
6	291	68.96	23.88	69.00	145
7	229	69.11	23.89	70.00	139
8	183	68.87	27.81	68.00	140
9	150	70.03	25.22	68.00	135
10	127	72.35	25.60	69.00	148
11	108	71.52	25.75	70.00	160
12	85	71.55	26.81	70.00	132
13	57	71.91	24.81	70.00	128
14	46	71.52	28.04	71.50	132
15	38	71.61	26.90	74.50	122
16	31	70.52	19.62	72.00	91

Note. M = mean; Mdn = median; SD = standard deviation.

from baseline, the expected improvement would be 4.17 points. The variance of the random coefficient was 4.97 (SD = 2.23). According to the normal curve reference distribution, the majority of participants (67%) would improve at a rate of -1.39 ± 2.23 (-3.62 to $+.84$) points in every measurement period. This suggests there is considerable individual variability in the rate of change: that some participants changed dramatically better than average and that some patients failed to change and possibly got worse over time.

On average, however, participants achieved significant gains immediately, after the first 3 months of treatment (measurement period 2). Table 5 shows that the change in the *average* OQ Total score from Time 1 to Time 2 was 75.08 – 68.09 = 6.99 points, almost a third of a standard deviation (SD = 24.04). The fairly large improvement in terms of raw points on the OQ Total may appear to conflict with a model which suggests a change of only 1.39 ± 2.23 points per measurement period, but since the statistical model is a *linear* growth curve, the rate of change is a weighted average of the *large immediate gains* and the *smaller later term gains* in the time series. Figure 1 depicts the actual data versus the fitted model; the bars represent sample size over time, clearly indicating a negative exponential attrition pattern. This figure shows that the first five measurement periods exerted the most weight on the rate of change parameter because the sample size was much larger in Year 1. A different dynamic occurs from Year 2 forward. A curvilinear growth curve model might or might not capture the dynamics more precisely. A piecewise linear model with two segments was run, one during Year 1 and a second segment after Year 1. This model pointed to a significant slope in Year 1, but not in Year 2. The dynamics of the second segment may be interpreted as maintenance of gains, but differential attrition effects may also account for the change seen here.

In summary, the slope of -1.39 for the full sample fits the observed data very well in the first year. After Year 1, it appears that a different outcome dynamic was at play, so that

the strict linear model may be extrapolating outside its useful range. The ES for Year 1 growth for the entire sample was computed using Cohen's *d*, using the square root of the pooled standard deviation of the baseline and the end-of-year measurements. The aggregate ES of treatment after 1 year, ES = (66.90–75.08)/23.96 = $-.34$ which is a considered a small to moderate ES. Cohen's *d* is sometimes interpreted using the normal curve reference distribution, which in this case suggests that a person who was in the 50th percentile (average OQ Total) of the Time 1 *untreated sample* would be found in the 63rd percentile of that same distribution after 1 year of psychodynamic psychotherapy.

Attrition. Table 5 depicts the average OQ Total scores over the 16 measurement periods. Anomalies appear in the time series. For example, the dramatic aggregate level improvement in the first 3 months (measurement period 2 on the OQ Total) appears to be followed by an unexpected backslide between Time 2 (68.09) and Time 3 (69.46), an *increase* in 1.37 points. However, attrition appears to be at work, because the sample size dropped precipitously from 1,050 to 683 in the period from Time 2 to Time 3 and continues to drop steadily thereafter. The 367 individuals who lapsed from Time 2 to Time 3 were actually those participants who achieved more dramatic gains at Time 2 compared to the 683 who persisted to Time 3. The lapsers had a Time 1 OQ Total score M = 72.81 (SD = 25.40) and a Time 2 score M = 63.81 (SD = 25.41), a 9.0 point improvement. In contrast, those who *stayed* in therapy had a Time 1 score on the OQ Total of M = 76.29 (SD = 23.21) and a Time 2 score of M = 70.39 (SD = 24.08), which is a gain of only 5.9 points. It therefore appears that the aggregate decline in functioning at Time 3 was due in part to the fact that *those who experienced the greatest initial treatment gains from Time 1 to Time 2 were more likely to exit treatment prior to Time 3 which had the statistical effect of raising the aggregate mean.* Indeed, looking only at the 683 nonlapsers, the Time 2 OQ Total score M = 70.39 (SD = 24.08), and the Time 3 M = 69.46 (SD = 24.13) appears to eliminate the appearance that the overall sample got worse in overall psychological functioning.

Another trend anomaly merits attention, one which again shows that the aggregate growth curve must be interpreted cautiously. This effect is seen in Table 5 where the improvement at Time 5 (M = 66.90) is followed by an apparent backsliding at Time 6 (M = 68.96). A careful examination of Figure 2, however, shows that a rather large segment of participants (N = 84) dropped out at measurement period 6.

The bar graph indicates that the soon-to-be dropouts had uniformly better average OQ Total scores across all age groups at Time 5, with a slightly greater difference for the older groups. In short, those participants who had improved most during the first year of treatment often *dropped out at the end of the year, at Time 5*, which changes the mixture of participants at measurement period 6. Again, since the aggregate OQ Total means in the second segment of the time series are now weighted toward participants *whose symptoms persisted after a year of treatment* it is not surprising to see an upswing
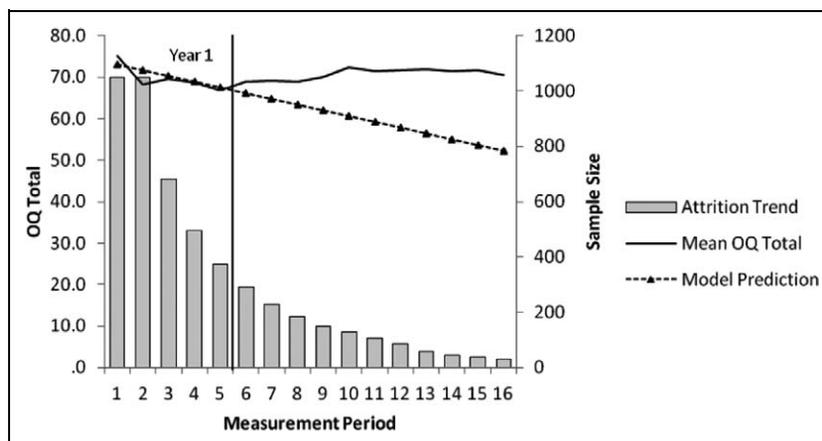
Figure 1. Treatment effect over time on the OQ Total scale with the fitted model score and the sample size at each measurement period. The linear model is a good fit for the treatment trajectory during Year 1 where the sample size is largest, but appears to be extrapolating badly starting in Year 2. This effect is potentially caused by differential attrition in the sample (differential in the sense that patients who experienced gains during the first year were more likely than other patients to dropout of treatment), and the effect might merit a piecewise growth curve to account for the different growth dynamics of the different populations at Year 1 and Year 2 forward.
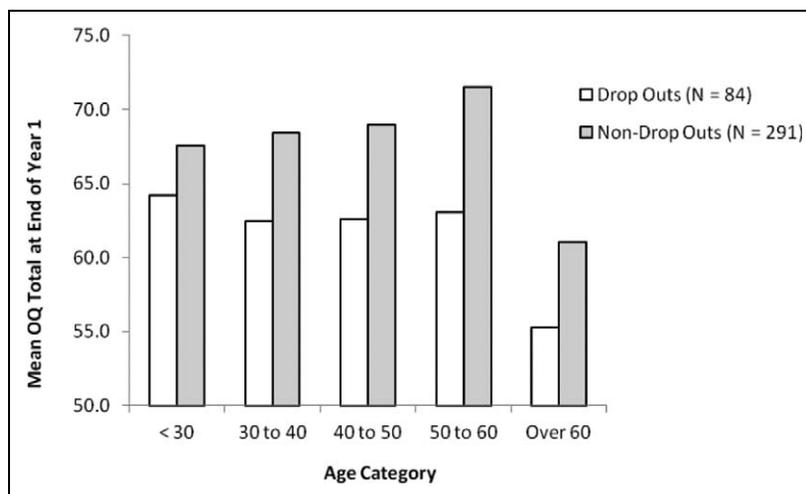
Figure 2. End of first year dropout. Dropout at Time 6 (beginning of second year) is strongly associated with quality of outcome at Time 5. This graph reveals that those who dropped out of the sample at Time 6 were the patients who had better outcomes at Time 5. Thus, one would expect that average OQ Total scores at Time 6 increase not due to backsliding, but because the lower scores dropped out of the sample, thus increasing the mean of the sample that persisted.

in OQ Total scores as indicating not necessarily a worsening of psychological functioning, but *a new mixture of participants*, a new growth curve dynamic. Those participants who remained after 1 year were perhaps a qualitatively different subset whose psychological functioning might be presumed to reflect a mixture of maintenance and slower treatment gains, after the dramatic first year gains of the recovering groups have taken leave of the equation. It should be noted, also, that the percentage of remaining clients deteriorating after a year increases. It is possible that a qualitatively different group stays longer than a year and that it accounts in part for this trend, but that at the same time another subset of participants are becoming more symptomatic in relation to their own baseline score.

This pattern of results suggests not only the value in using a random effect slope term to capture individual differences in

the rate of change in HLM but also suggests that there might be significant treatment by person interaction effects beyond the main effect of time. The researchers examined a series of contextual variables in the language of multilevel modeling, to see who might make up these subgroups that make up the aggregate improvement seen in the aggregate growth curve over time. We looked at if perhaps some personal factors, such as symptom severity and age, might influence the treatment effects.

Moderator Variables

Several context variables were added to the mixed model to identify higher order treatment effects as well as the possibility of moderator variables, which are variables that might cause or
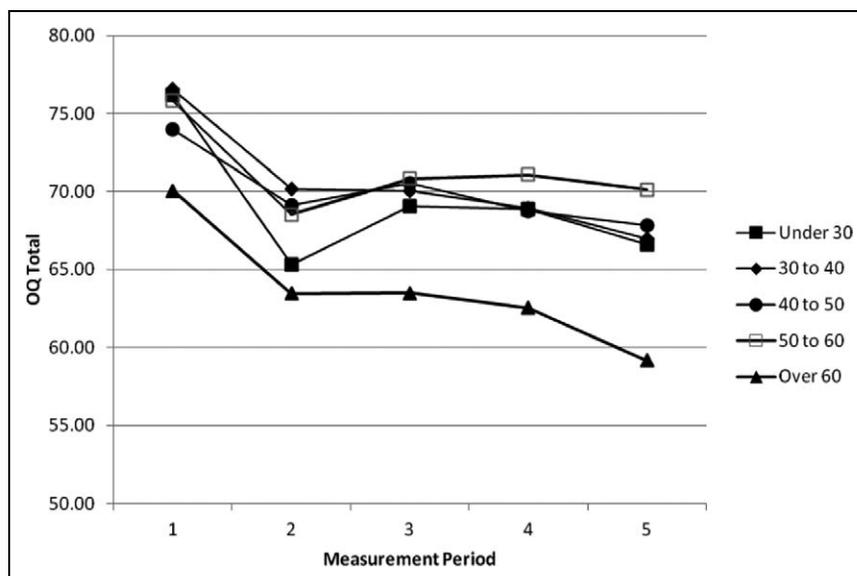
Figure 3. Average OQ Total scores during the first five measurement periods (first year) by participant's age. The thick solid line represents people age 60 and older.

reflect different "shapes" among individual growth trajectories. The moderator effects changed the initial model by adding a main effect and an interaction term. The main effect indicated whether the presence of the context variable would adjust the fixed intercept above and beyond the individual's unique random adjustment to the fixed value. The _variable by time interaction_ tested whether different levels of the moderator variable would have different slopes in their growth curves above and beyond the individual's unique random adjustment to the fixed value.

Significant context effects were found for age (those 60 and older), initial symptom severity, medication, income, major depression, and adjustment disorder.

Model effects for gender, number of sessions, Spanish speaking, bipolar, and anxiety were all found to be nonsignificant.

Age. It appears that older participants had a significantly different and more dramatic pattern of treatment outcome compared to other age groups. Figure 3 shows treatment trajectory in Year 1 broken down by age. The over 60 age group stands out, displaying a strong monotonically decreasing trajectory, but younger individuals (under age 30) also showed significant short-term improvement. The mixed model analysis of a Level 3 dummy variable designated _senior_ showed a significant main effect, $F(1, 1020.863) = 10.12$, $p = .002$, but not a significant interaction effect with time. The qualitative differences in trajectory of the age groups shown in Figure 3 suggest some identifiable differences that were not detected by dichotomizing the age groups into senior versus all others. The fixed intercept was 72.92, and the fixed effect of time was −1.43, both comparable to the primary two-level model defined earlier, but the fixed parameter estimate for the Level 3 _senior_ variable was −10.19 (significant because the main effect is significant) indicating that above and beyond the fixed intercept, OQ Total

scores for seniors can be expected to be slightly more than 10 points lower than the average (less symptom severity). The ES for persons over age 60 after 1 year of psychodynamic psychotherapy, using the calculation method described above, was found to be ES = −.49, which is considered a medium effect.

Initial symptom severity. This was defined as a score on the OQ Total scale greater than or equal to 93, as this is a cutoff noted by Lambert and was used in shaping a modified version of the OQ specific to people with severe and persistent mental illness (SPMI). As can be seen in Figure 4, this subsample ($N = 238$) demonstrated decisive, sustained change during the first year of psychotherapy. The main effect of greater initial symptom severity was significant, $F(1, 1020.052) = 178.857$, $p < .001$, as was the interaction of initial symptom severity with time, $F(1, 224.841) = 4.06$, $p = .045$. The effect of this variable was 29.11, which suggests that above and beyond individual differences in intercept, this more symptomatic group can be expected to be just over 29 points higher than the average slope. In addition, the growth rate for this group can be expected to be almost a full point larger than the average growth rate. When the three-level model parameters are estimated in HLM, the fixed intercept is 69.19, the slope is −1.30, and the interaction with greater initial symptom severity effect is −.97, which gets added above and beyond the fixed −1.30, resulting in a fitted growth rate for this subgroup of −1.30 to .97 = −2.27 points on the OQ Total. Again, Figure 4 shows that the outcome trajectory of the more symptomatic subpopulation is qualitatively different than the less symptomatic. This qualitative difference in shape is what is indicated by the significant interaction. The ES for this subpopulation, using the methods previously described, showed ES = −1.34 which is considered a large treatment effect, but is consistent with much recent literature
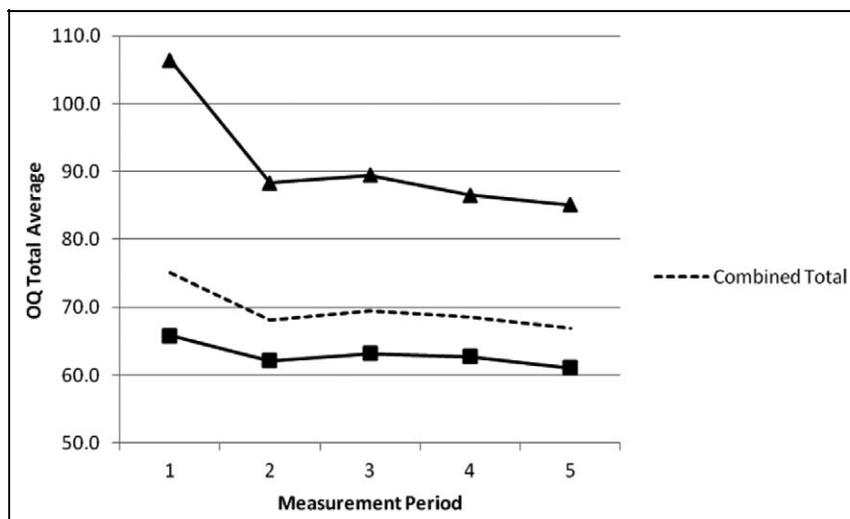
Figure 4. Outcomes for higher versus lower initial symptom severity patients. Results show dramatic and sustained improvement for the more symptomatic group during the first year of psychotherapy and moderate yet sustained improvement for the remainder of the sample. The combined scores gravitate more closely to the less symptomatic group due to the larger number of participants in that group.

on the effect of psychodynamic psychotherapy on people with more extreme symptom severity. Certainly there is always the specter of regression toward the mean effects with no-control pretest posttest designs. However, regression toward the mean is an artifact of random statistical error and random measurement error. These might be anticipated particularly in the period from the initial measurement to the first outcome period measurement (Time 1 to Time 2). Yet, Figure 4 reveals that the initial treatment effect was sustained for the remainder of the year.

Medication. Medication exerted an effect both in terms of intercept and slope. Participants using medication in this model presented, on average, as 8 points more symptomatic in their initial OQ Total score. They similarly made significant progress, in the form of a downward slope, early in treatment, but also showed additional symptom relief between 9 and 12 months, which was not as visible for people not using medication as part of treatment. There was a pronounced association between medication and age, with older adults more likely to be on medication from outside the clinic, thus the medication effect may be reflecting the senior/age effect.

Income. There was a significant effect for income, but it was very slight in effect in magnitude (income's effect on slope was $-.00002$). In addition, income was perhaps not surprisingly found to be strongly associated with age, $\chi^2(16) = 76.47$, $p < .001$, and therefore this particular result seems to be confounded with the age variable.

Major depression. There was a significant main effect for major depressive disorder, $F(1, 914.81) = 75.44$, $p < .001$, but no significant interaction with time. The effect of major depression on the intercept was 13.93, which suggests that those with major depression began treatment with an elevated OQ Total

score, and that this elevation persisted across measurement periods, though the rate of change is no different than other. In other words, with major depression showed comparable change on a parallel trajectory that was elevated by about 14 points.

Adjustment disorder. The main effect of adjustment disorder was significant, $F(1, 652.847) = 47.55$, $p < .001$, but the interaction with time was not. The effect of adjustment disorder was -16.19, which suggests that patients diagnosed with adjustment disorder began treatment on average -16.19 points lower than the average intake score on the OQ Total and that this difference remained constant across measurement periods. Again, patients with an adjustment disorder had a parallel trajectory compared to the average, but consistently lower in symptom severity, which might be expected.

Discussion

Findings, overall, were robust and pointed to a general trend of recovery over time. While participants in this treatment tended to begin with a baseline score in keeping with national norms for an outpatient community mental health sample (OQTOT = 75.08), clients improved, broadly, with few notable interaction effects. This trend toward recovery was particularly pronounced in the first 3 months of treatment, where the sharpest, most dramatic improvement tended to occur. This finding of significant, early symptom change, and of an inverse recovery curve is, in itself, very much in keeping with past studies regarding the nature of recovery in psychotherapy (i.e., Kopta, Howard, Lowry, & Beutler, 1994). Specifically, our finding pointing to the importance of the first 3 months is also very much in keeping with Lambert's finding that treatment response is most often evidenced by 18 sessions (Lambert, Whipple, Hawkins, & Vermeersch, 2003). Lambert, an author

of the OQ, has pointed repeatedly to the importance of measuring early treatment response, the absence of which often predicts a poorer prognosis. This has potential theoretical implications for dynamic therapists as well in that we found little evidence of a need to get "worse before better," even in longer-term treatments. To the contrary, we found evidence that without early clinically reliable improvement, clients were unlikely to improve later.

The strongest surprise finding was perhaps the degree to which this recovery trend was shared across subgroups. Subgroup analysis pointed to very few interaction effects, or to moderating variables that seemed to affect the slope or trajectory of recovery. Only a few variables emerged suggesting unique recoveries, such as those participants who began as more symptomatic and older adults. People who began as more symptomatic (OQ TOT above 93) showed an even stronger initial slope in recovery that was maintained for at least a year, in keeping with findings from Puschner et al. (2007). Older clients tended to begin with an intercept significantly lower than the general clinic population by an average of 10 points. While they did not demonstrate a necessarily unique course of recovery, they did show, as a group ($n = 106$) a higher than clinic average ES ($-.49$ vs. $-.34$ for the sample as a whole). This finding is also in keeping with other recent studies which suggest older adults may be able to make particularly good use of psychotherapy (James, 2008).

Also, while the general treatment effect, sample-wide, appears robust, it is more tempered than that reported by either Shedler (2010) or Leichsenring and Rabung (2008) in their meta-analyses of both short-term and long-term psychodynamic psychotherapy, respectively. We interpret our smaller ESs as potentially reflecting the actual practice of clinic-based, open-ended treatments with people versus shorter-term, prospective treatments of a predefined, prescribed length. That said, the overall ES of $-.34$ while small to moderate, was not weak. Shedler (2010) makes a similar point when he references the average ES for an antidepressant across trials as 0.3 or less.

We did find evidence of change after the initial 3 months. While 68% of participants met criteria for caseness (clinically significant symptomology) at intake, fewer met this criterion at 3, 6, and 9 months. We found that older adults (those at or over age 60) and people with greater initial symptom severity in particular showed additional improvement during the 9–12 month interval. Similarly, what appeared initially to be some deterioration (worsening scores between 3 and 9 months in the overall sample) in part reflected people who began as more symptomatic at intake. Subgroup analysis pointed to a phenomenon whereby participants who improved were more likely to end treatment by the following interval throughout the course of the first year.

Overall, we found a year to be a significant marker or "break point" in treatment. That is, while data pointed to a significant slope or recovery trajectory spanning the first year of treatment, this trend was not demonstrated broadly after a year. Supporting this finding, trends in deterioration increased

significantly after a year, and treatment responses appeared to become more heterogeneous. Stepwise comparisons did not demonstrate significant change, broadly after a year. However, the majority of people who stayed beyond a year appear to maintain their gains.

These findings point to several potential clinical implications: (a) Improvement is often demonstrated early. (b) People who improve in the first year tend to end. (c) People who stay for 9–12 months sometimes, but not always, get additional benefit., (d) The general trend after a year is one of the maintenance of gains rather than additional improvement. (e) After a year, treatment trajectories become more varied. What looks like deterioration may be accounted for, in part, by early responders ending and by more symptomatic clients continuing; however, we found evidence as well of the percentage of clients deteriorating *in relation to their own baseline* score increasing significantly after a year. All of this points to a central potential clinical implication: the wisdom in looking for early improvement and in reevaluating at a year. Data from instruments such as the OQ ought to serve in support of but not as replacements for clinical judgment. These findings also begin to challenge those by Leichsenring and Rabung (2008), with ours suggesting that, in actual practice, clients can often benefit earlier than their interpretation suggests, and that those who stay beyond a year (or 50 sessions by their definition) may constitute a unique subpopulation or in the language of multilevel modeling, a "population within a population." It also raises a question as to whether the lack of significant slope representing change after a year reflects an important period in order to encourage and to solidify the maintenance of gains or a "plateau" and a call to consider ending these treatments earlier.

Overall outcomes did not seem to correlate with number of sessions. Leichsenring and Rabung (2008) in some ways concede this, pointing out that the impact of number of sessions in their meta-analysis was not always clear. Unlike these authors, we did not find an inherent superiority of long-term treatment. People in this sample were able to often benefit well in advance of this (with a mean of 21 visits, a median of 9, and a mode of 3). A similar recovery trajectory was found, even in three visits.

Our findings were more in keeping with Shedler (2010), who found evidence of a treatment response in fewer than 40 sessions. This study adds some additional information that may even provide some caution regarding potential risks of longer-term treatments in that we found greater risk or evidence of deterioration after a year and of a plateau in recovery at a group level. In keeping with Vinnars et al. (2005), this study points to the importance of a year of treatment. In looking at personality disorders, Vinnars et al. (2005, p. 1938) found that "community delivered psychodynamic treatment for at least one year decreases the severity of personality disorders and psychiatric symptoms, as well as improving functioning" (p. 1938). While this study did not look at personality disorders, specifically, it found evidence of a benefit in a similar time frame, broadly, for a large clinic sample. Also, in keeping with this same study, Vinnars et al. point to smaller ESs in their clinic sample than in several existing prospective studies.

While they found that people with personality disorders responded well to a dynamic approach to treating personality disorders, they found that after a year, "(their sample) did not recover to the same level as in a non-clinical Swedish sample" (p. 1938). This is also comparable to what we found. Finally, we found, in keeping with Driessen et al. (2010), that gender did not correlate with outcome or act as a moderating variable. Men and women showed evidence of benefiting from this form of treatment. We also found that older adults could do well and showed potentially an even better response than younger participants in this study using a nonmanualized treatment.

Expectedly, this study had associated strengths and limitations inherent in its design. In terms of strengths, we were able to begin to offer some of what Driessen et al. (2010) requested, which is attention to both moderating variables and to subgroup analysis (i.e., looking at the course of specific diagnoses) in relation to psychodynamic treatment. We used open-ended, actual treatments, looking at them retrospectively versus the way Leichsenring and Rabung limited themselves to prospective studies (2008, p. 1552). While prospective and manualized studies tend to offer stronger internal validity, designs such as ours are often associated with greater external or "ecological" validity. They also critiqued their own study as relying on a small number of studies (23) of variable quality, many of which were older studies (dating back to 1980s) with "methodological shortcomings." This is, in contrast, a current data set.

With both a smaller mean number of sessions than Shedler (2010) and a longer course of treatment represented (45 months) than in the majority of studies included in Leichsenring and Rabung's 2008 meta-analysis, we were able to look simultaneously at both short- and longer-term treatments. We were also able to give attention to deterioration, a variable more recently being given attention in psychotherapy research (readers are referred to the 2010 issue 65 of *American Psychologist*, which was devoted to this topic). This study is able to begin to weigh in on some of the above questions, regarding optimal dosing or length of treatment, ESs in efficacy versus effectiveness studies, and even the role of number of visits and the accompanying risk of deterioration.

In terms of weaknesses, we lacked dimensionality (i.e., more than a single instrument) and specificity, in that we did not look at outcomes for particular diagnostic groups, beyond some attention broadly to diagnostic clusters such as major depression and anxiety disorders, looking at them for potential interaction effects. This study in many ways offers a "bird's eye view" of change for a broad group of people with heterogeneous diagnoses and lengths of treatment. It lacks some specificity and much of the nuance that case studies and qualitative methods can bring to the particularities of individual treatments with individual clinicians. Pinsoff and Wynne (2000) have warned of the "uniformity myth:" the notion that different clients with different therapists, even with a similar theoretical orientation are getting "the same thing" (quotes added). This study also lacked information about the nature of these often early endings. That is, are we witnessing formal, planned endings or attrition (i.e., dropout). Statistical attention to trends in this data set offers a partial, but not complete answer and invites further research.

The present study used a nonexperimental within-subjects design, which means that subjects were compared against their pretest scores to determine treatment efficacy. Rival hypotheses in such designs cannot be ruled out. The overall improvement documented here can plausibly be explained by the hypothesis that psychodynamic treatment caused patient gains. The therapeutic effectiveness of psychodynamic psychotherapy has been demonstrated by prior research, thus the hypothesis had a nonnegligible prior probability in our estimation. However, without random assignment of subjects to treatment or control groups, the following general classes of rival hypotheses must be at least considered as possible alternatives. First, the treatment gains might be explained by concurrent history, including any events of a public nature that might have caused participants to change for the better, or the gains might simply be explained by the passage of time. Statistical regression toward the mean is certainly a threat in a study such as this, particularly for participants who began treatment with clinically elevated scores. Selection of participants with elevated scores, in the context of statistical variability and measurement unreliability create a likelihood that follow-up measurements will be closer to the mean, creating the appearance of improvement. Maturation effects such as developmental change could also explain treatment gains. Repeated measurements using the OQ-45 could possibly play a role in these findings, as the sheer experience self-assessment over time might act as a motivating factor to improve. Attrition is an important threat to studies such as this one. The scenario is one in which the most pathological participants drop out of treatment over time, which in this study would produce a false picture of improvement over time. In this study, the researchers examined which patients dropped out of treatment after the first year and found that the attritors were actually those who improved more in the preceding year. While not eliminating the possibility of attrition as a confounding factor, an attempt was made here to consider it carefully. Obviously, there is considerable censoring of data, as a large number of participants dropped out in the first year of treatment. It would be useful for future researchers to model attrition explicitly, using recent techniques such as event history designs (Blossfeld & Rohwer, 2002).

Finally, while we also lacked follow-up here in at least two important ways. First, we had neither posttreatment confirmation of diagnosis nor the ability to continue to measure participants after they dropped out or otherwise ended. These are components of an intent to treat design and analysis. Second, we lacked a follow-up measure in order to find out whether gains are maintained. However, a smaller, unpublished pilot study at this same clinic ($n = 43$) points to evidence of the maintenance of gains posttreatment at 12–18 months using the OQ in a way consistent with Maina, Rosso, and Bogetto's (2009) study: a randomized controlled trial ($N = 92$) which showed that patients who have had brief psychodynamic

treatment (BDT) in addition to medication were less likely to relapse than those who had only had pharmacotherapy at even 4 years of follow-up. We recommend future studies build on this initial look by considering the use of a formal control or comparison group, in a community setting, if feasible. A replication would be strengthened by following participants with specific diagnoses, perhaps using a mixed method approach. A prospective, longitudinal, and even mixed method design would allow researchers to talk with people who end along the way and to ask about reasons for ending, their own sense of their success, and how their impressions do and do not align with their OQ scores.

Acknowledgement

Declaration of Conflicting Interests

Funding

References

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

Beck, A., & Bhar, S. (2009). Letters: Analyzing the effectiveness of psychodynamic psychotherapy. *Journal of the American Medical Association*, *301*, 931.

Bhar, S. S., Thombs, B. D., Pignotti, M., Bassel, M., Jewett, L., Coyne, J. C., & Beck, A. T. (2010). Is longer term psychodynamic psychotherapy more effective than shorter term therapies? Review and critique of the evidence. *Psychotherapy and Psychosomatics*, *79*, 208-216; doi: 10.1159/000313689

Bickel, R. (2007). *Multilevel analysis for applied research.* New York, NY: Guilford.

Blossfeld, H. P., & Rohwer, G. (2002). *Techniques of event history modeling.* Mahwah, NJ: Lawrence Erlbaum.

Bond, M. B., & Perry, J. C. (2004). Long-term changes in defense style associated with psychodynamic psychotherapy for depressive, anxiety and personality. *American Journal of Psychiatry*, *166*, 1665-1671.

Bradshaw, W. H., Roseborough, D. J., Pahwa, R., & Jordan, J. (2010). Evaluating psychodynamic psychotherapy in a community mental health clinic. *Journal of the American Academy of Psychoanalysis & Dynamic Psychiatry*, *37*, 663-678.

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

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, *40*, 211-223.

de Maat, S., De Jonghe, F., Schoevers, R., & Dekker, J. (2009). The effectiveness of long-term psychoanalytic therapy: A systematic review of empirical studies. *Harvard Review of Psychiatry*, *17*, 1-23.

Doidge, N. (1999). How to preserve psychoanalysis: Introduction to Gunderson & Gabbard. *Journal of the American Psychoanalytic Association*, *47*, 673-677.

Driessen, E., Cuijpers, P., de Maat, S. C. M., Abbass, A. A., de Jonghe, F., & Dekker, J. J. M. (2010). The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis. *Clinical Psychology Review*, *30*, 25-36.

Fraser, M. W., & Galinsky, M. J. (2010). Steps in intervention research: Designing and developing social programs. *Research on Social Work Practice*, *5*, 459-466.

Gabbard, G. O. (2008). *Textbook of psychotherapeutic treatments.* Washington, DC: American Psychiatric Publishing.

Harvard Mental Health Letter. (2010, September). *Merits of psychodynamic therapy*, *27*, 1-3.

James, I. A. (2008). Stuff and nonsense in the treatment of older people: Essential reading for the over-45's. *Behavioural and Cognitive Psychotherapy*, *36*, 735-747.

Kernberg, O. F. (1993). The current state of psychoanalysis. *The Journal of the American Psychoanalytic Association*, *41*, 45-62.

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

Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., & Burlingame, G. M. (2004). *Administration and scoring for the OQ 45.2.* Salt Lake City: American Professional Credentialing Services, LLC.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., & Vermeersch, D. A. (2003). Is it time to routinely track patient outcomes: A meta-analysis. *Clinical Psychology: Science and Practice*, *10*, 288-315.

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

Leichsenring, F., & Rabung, S. (2008). The effectiveness of long-term psychodynamic psychotherapy. *Journal of the American Medical Association*, *300*, 1551-1565.

Littell, J., & Shlonsky, A. (2010). Making sense of meta-analysis: A critique of the effectiveness of long-term psychodynamic psychotherapy [Special Issue]. *Clinical Social Work Journal*.

Maina, G., Rosso, G., & Bogetto, F. (2009). Brief dynamic therapy combined with pharmacotherapy in the treatment of major depressive disorder: Long-term results. *Journal of Affective Disorders*, *114*, 200-207.

Mauish, M. E. (2004). *The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults.* Mahwah, NJ: Lawrence Erlbaum.

Milrod, B., Leon, A. C., Busch, F., Rudden, M., Schwalberg, M., Clarkin, J., . . . Shear, M. K. (2007). A randomized controlled clinical trial of psychoanalytic psychotherapy for panic disorder. *American Journal of Psychiatry*, *164*, 265-272.

Pinsoff, W. M., & Wynne, L. C. (2000). Toward progress research: Closing the gap between family therapy practice and research. *Journal of Marital and Family Therapy*, *26*, 1-8.

Puschner, B., Kraft, S., Kachele, H., & Kordy, H. (2007). Course of improvement over 2 years in psychoanalytic and psychodynamic outpatient psychotherapy. *Psychology & Psychotherapy: Theory, Research, & Practice*, *80*, 51-68.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE.

Rose, C. (2007). *The legacy of Sigmund Freud. Charlie Rose.* New York, NY: Public Broadcasting System (PBS).

Schilling, R. (2010). Commentary: The challenge of nonexperimental intervention studies in social work. *Research on Social Work Practice*, *20*, 550-552.

Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologist*, *65*, 98-109.

Spybrook, J., Raudenbush, S., Congdon, R., & Martinez, A. (2009). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software.* Software Manual. Lincolnwood, IL: Scientific Software International.

State of Minnesota. (2001). Minnesota House of Representatives Information Brief. *Racial/ethnic minorities in Minnesota: Geographical distribution among the counties*. Retrieved on November 14, 2010, from http://www.house.leg.state.mn.us/hrd/pubs/mintminn.pdf

Summers, R. F., & Barber, J. P. (2009). *Psychodynamic psychotherapy: A guide to evidence-based practice.* New York, NY: Guilford.

Svanborg, C., Wistedt, A. A., & Svanborg, P. (2008). Long-term outcome of patients with dysthymia and panic disorder: A naturalistic 9 year follow-up study. *Nordic Journal of Psychiatry*, *62*, 17-24.

Umphress, V. J., Lambert, M. J., Smart, D. W., & Barlow, S. H. (1997). Concurrent and construct validity of the outcome questionnaire. *Journal of Psychoeducational Assessment*, *15*, 40.

Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000). Outcome questionnaire: Item sensitivity to change. *Journal of Personality Assessment*, *74*, 242-261.

Vinnars, B., Barber, J. P., Noren, K., Gallop, R., & Weinryb, R. M. (2005). Manualized supportive-expressive psychotherapy versus nonmanualized community-delivered psychodynamic therapy for patients with personality disorders: Bridging efficacy and effectiveness. *American Journal of Psychiatry*, *162*, 1933-1940.