

# Rates of Change in Naturalistic Psychotherapy: Contrasting Dose–Effect and Good-Enough Level Models of Change

Scott A. Baldwin and Arjan Berkeljon  
Brigham Young University

David C. Atkins  
University of Washington

Joseph A. Olsen and Stevan L. Nielsen  
Brigham Young University

Most research on the dose–effect model of change has combined data across patients who vary in their total dose of treatment and has implicitly assumed that the rate of change during therapy is constant across doses. In contrast, the good-enough level model predicts that rate of change will be related to total dose of therapy. In this study, the authors evaluated these competing predictions by examining the relationship between rate of change and total dose in 4,676 psychotherapy patients who received individual psychotherapy. Patients attended 6.46 sessions on average ( $SD = 4.14$ , range = 3–29,  $Mdn = 5$ ). The results indicated that although patients improved during treatment, patients' rate of change varied as a function of total dose of treatment. Small doses of treatment were related to relatively fast rates of change, whereas large doses of treatment were related to slower rates of change. Total dose had a nonlinear relationship with the likelihood of clinically significant change. Given the variability in rates of change, it appears that time limits for treatment uniform to all patients would not adequately serve patients' needs.

*Keywords:* dose–response models, good-enough level, growth curve models, psychotherapy outcome

Psychotherapy research has conclusively demonstrated that people improve because of treatment (Lambert & Ogles, 2004). Given the convincing evidence that people improve during therapy, researchers have studied the rate at which people improve. Two related questions drive this research: (a) How much therapy is needed to achieve significant improvement, and (b) how much do patients benefit from each session of therapy? (Hansen, Lambert, & Forman, 2002). To address these questions, researchers have focused on the effects of different “doses” of therapy, where dose is usually defined as the number of sessions. These studies are typically called dose–response studies because treatment response is modeled as a function of treatment dose. In this study we contrast two approaches to understanding dose–response relationships: (a) the dose–effect model (Howard, Kopta, Krause, & Orlinsky, 1986) and (b) the good-enough level model (Barkham et al., 2006).

The most common approach to understanding dose–response relationships in psychotherapy has been termed the dose–effect

model (Howard et al., 1986). The dose–effect model is based on a medical understanding of dose, whereby sessions in psychotherapy are compared to milligrams in pharmacotherapy (Kopta, Howard, Lowry, & Beutler, 1994). Just as increasing the dose of medications will expose medical patients to higher amounts of the active ingredients of medications, increasing the number of sessions will expose psychotherapy patients to higher amounts of the active ingredients of psychotherapy. The dose–effect model suggests that the relationship between dose and rate of change during therapy is negatively accelerating. That is, patients improve with increasing sessions, but the benefit of additional sessions appears to decrease at higher doses.

Much of the early research on the dose–effect model examined the relationship between probability of recovery and dose and demonstrated a negatively accelerating relationship between dose and recovery (e.g., Howard et al., 1986; Kopta et al., 1994). Although the results of these studies were often interpreted as indicating that the expected rate of change during therapy follows a negatively accelerating pattern (Lutz, Martinovich, & Howard, 1999), the early results did not provide information about the shape of session-to-session change. However, more recent research that used multilevel growth curve models to explicitly model session-to-session change during therapy has suggested that the session-to-session change follows a negatively accelerating curve (e.g., Lutz, Lowry, Kopta, Einstein, & Howard, 2001; Lutz et al., 1999).

Consequently, many researchers interpret the negatively accelerating curve as the expected pattern of change in psychotherapy. For example, Lutz et al. (1999) interpreted the negatively accelerating curve as “lawful” (p. 571) and concluded:

---

Scott A. Baldwin and Arjan Berkeljon, Department of Psychology, Brigham Young University; David C. Atkins, Department of Psychiatry and Behavioral Science, University of Washington; Joseph A. Olsen, College of Family, Home, and Social Sciences, Brigham Young University; Stevan L. Nielsen, Counseling and Career Center, Brigham Young University.

Portions of this research were presented at the annual meeting of the North American Society for Psychotherapy Research, New Haven, CT, in September 2008.

Correspondence concerning this article should be addressed to Scott A. Baldwin, 268 TLRB, Brigham Young University, Provo, UT 84602. E-mail: scott\_baldwin@byu.edu

All of this [i.e., dose–response research] indicated that the “true” course of recovery could be described by a log-normal curve [i.e., negatively accelerating] and that the actual progress of a patient could be compared with a hypothetical true course that would be expected for that patient. (p. 571; see also Grissom, Lyons, & Lutz, 2002; Howard et al., 1986; Kopta, 2003; Lueger et al., 2001; Lutz et al., 2001)

However, the dose–effect interpretation of the negatively accelerating curve makes a crucial assumption: The effect of additional sessions (or time in treatment) is on average equal across people.<sup>1</sup>

It is unclear whether this assumption is tenable, because much previous dose–response research has aggregated across people who attended different numbers of sessions. For example, Howard et al. (1986) used naturalistic psychotherapy data and thus did not fix the dose of treatment. Rather, the dose was determined by the patient and therapist, and it varied considerably (see also Howard, Lueger, Maling, & Martinovich, 1993; Kopta et al., 1994; Lutz et al., 1999). Patients may leave treatment because they have improved (or have not improved). Therefore, the dose of therapy that patients received was systematically related to treatment response (Stiles, Honos-Webb, & Surko, 1998) as opposed to being independent of treatment response, which would be the case if the researchers fixed the dose (i.e., patients randomly assigned to 10 or 20 sessions of treatment) or if patients dropped out of therapy for unsystematic reasons. Having the dose under the control of therapists and patients is not a problem if rate of change does not vary as a function of dose. For example, if patients who come for 5 sessions change at the same rate on average as patients who come for 10 sessions, then aggregating across patients with different doses will not bias the analysis. This situation is consistent with the dose–effect model. On the other hand, if patients who come for 5 sessions change at a different rate than patients who come for 10 sessions, aggregating will not accurately reflect the pattern of change among the different groups of patients.<sup>2</sup>

The latter situation is consistent with the *good-enough level* (GEL) model, which is another way of interpreting dose–response relationships. Specifically, the GEL model assumes that patients who come for different numbers of sessions change at different rates (Barkham et al., 2006; Barkham, Rees, Stiles, et al., 1996; Stiles, Barkham, Connell, & Mellor-Clark, 2008). The GEL model predicts that patients remain in therapy until they, in conjunction with their therapist, determine that they have sufficiently improved—to the good-enough level. Therefore, on average the dose of treatment reflects treatment response and indicates how malleable patients’ symptoms are, rather than being the driving force of treatment response as in the dose–effect model. Thus, the GEL model predicts that patients who receive low doses of treatment are those who change rapidly, whereas patients who receive high doses of treatment are those who change slowly.

A second prediction of the GEL model is that patients with high doses of therapy should be no more likely to have experienced clinically significant change than those with low doses. Consistent with this prediction, Barkham et al. (2006) and Stiles et al. (2008) found that the rate of reliable and clinically significant change (Jacobson & Truax, 1991) did not increase as the total number of sessions attended increased. Barkham et al. (2006) speculated that if an analysis were limited to one group of patients, such as those who attended five total sessions, the rate of change would be linear rather than negatively acceler-

ating. Furthermore, they suggested that the negatively accelerating curve might be an artifact of aggregating across groups of people with different treatment lengths. That is, it appears that therapy becomes less effective over time because at later sessions the rapid changers have terminated, leaving only slowly changing patients.

Both Barkham et al. (2006) and Stiles et al. (2008) used only baseline and termination data to calculate improvement. Consequently, they were not able to assess the shape of change (i.e., linear or nonlinear). It does not appear that the GEL model requires that change be linear—patients can reach their GEL via linear or nonlinear change. Rather, the key prediction is that the effect of additional sessions is not, on average, equal across people with varying doses of treatment.

In sum, the dose–effect and GEL models make two competing predictions. First, the dose–effect model predicts that rate of change during therapy will not vary as a function of total number of sessions, whereas the GEL model predicts that it will vary. Second, the dose–effect model predicts that the likelihood of achieving clinically significant change will be related to the total number of sessions (i.e., positively correlated); whereas the GEL model predicts that they will be unrelated (or even negatively related). The purpose of this article is to test these competing predictions using session-by-session therapy outcome data to determine which model best accounts for the patterns of change during treatment.

## Method

### *Participants and Procedures*

Participants included in this study were drawn from an archival dataset of therapy outcomes at a large university counseling center. Patients of this counseling center complete the Outcome Questionnaire-45 (OQ-45; Lambert et al., 2004), which is a measure of treatment outcome, at intake and prior to each session. This counseling center provides a variety of services, including individual, group, and couples therapy. We limited our analyses to individual therapy outcomes during patients’ first therapy episode. We considered an episode to have ended if the time interval between sessions exceeded 90 days. In addition, patients had to have attended at least three sessions and have been in therapy no longer than 40 weeks. Although a very small proportion of patients attended therapy longer than 40 weeks, models including these outliers were unstable. Finally, we excluded patients for whom demographic and diagnostic information was unavailable so that

<sup>1</sup> This assumption is true whether or not researchers include covariates (e.g., demographics, diagnoses, symptom type) in their analyses (e.g., Kopta et al., 1994; Lutz et al., 1999). That is, for people categorized by any given combination of values on the covariates, the models assume that the benefit (or harm) of additional sessions is equal across those people, regardless of total dose.

<sup>2</sup> Technically speaking, the model would be misspecified because it would not include an important covariate (i.e., total number of sessions).

we could rule out these potential confounds.<sup>3</sup> Descriptive data for total number of sessions and weeks in treatment are reported in the Results section. The data for both patients and therapists were anonymized by the counseling center.

Using these criteria, we identified 4,676 patients seen by 204 therapists. Most patients saw either 1 therapist (42.2%) or 2 therapists (54%), although some patients saw 3 (3.7%) or 4 therapists (0.1%). Most patients who switched therapists switched after their first session, which is consistent with the common practice in the counseling center of having an intake session with one therapist and starting with a new therapist at the second session. The patients were predominantly female (62%) and single (65%). Their ages ranged from 17 to 60 ( $M = 22.3$ ,  $Mdn = 21.8$ ,  $SD = 3.7$ ). The majority of patients were Caucasian (88%), followed by Hispanic (5%), Asian (2%), Pacific Islander (1.1%), and other ethnic groups (3.9%). Initial diagnostic impressions were recorded by therapists after the first session. Adjustment disorders were the most common (38%), followed by mood disorders (25%), anxiety disorders (12%), and eating disorders (5%). A mix of other diagnostic categories accounted for 20% of patients.

### Outcome Measure

Treatment outcome was assessed by the OQ-45 (Lambert et al., 2004). The OQ-45 is a self-report measure specifically designed to track symptom change during therapy. The 45 items assess three primary dimensions: (a) subjective discomfort (e.g., anxiety and depression—"I feel blue"), (b) interpersonal relationships (e.g., "I feel lonely"), and (c) social role performance (e.g., "I have too many disagreements at work/school"). Typically all 45 items are summed to create a total score, which was used in this study. Total scores can range from 0 to 180, with higher scores reflecting poorer psychological functioning. The OQ-45 has been shown to have good internal consistency ( $\alpha = .93$ ), 3-week test-retest reliability ( $r = .84$ ), and concurrent validity (Lambert et al., 2004; Snell, Mallinckrodt, Hill, & Lambert, 2001).

### Statistical Analyses

*Rate of change.* We used multilevel growth curve models (Raudenbush & Bryk, 2002; Singer & Willet, 2003) to determine whether rate of change varies as a function of the total number of sessions patients attended. All models were estimated with the lme4 library (Bates, 2007) in the R programming language (version 2.7.1; R Development Core Team, 2007), using full maximum likelihood estimation procedures. The lme4 library was explicitly developed to handle complex data structures like the data used in this article, where data are correlated but not necessarily nested (i.e., some patients saw two or more therapists). The present data are partially cross-classified as opposed to strictly nested (see, e.g., Raudenbush & Bryk, 2002, chapter 12). The basic idea of random effects that control for correlations within the data still holds for cross-classified data, but the computational burdens are far greater.

To test whether rate of change was a function of the total number of sessions a patient attended, we compared the results of two growth models. First, consistent with the dose-effect model, we estimated an aggregate model, which averaged the rate of change across all patients, ignoring the total number of sessions patients attended. Second, consistent with the GEL model, we

estimated a stratified model, which included an interaction between rate of change and the total number of sessions patients attended. A significant interaction between total number of sessions and rate of change would indicate that patients who attended different numbers of sessions changed at different rates.

In many dose-response studies, researchers model change during treatment as a log-linear function of time (e.g., Lutz et al., 2001, 1999). However, our initial inspection of the data suggested that change might have followed a cubic pattern. Thus, we compared the model fit of both log-linear and cubic models. The cubic model significantly improved model fit over the log-linear model (Bayesian information criterion = BIC;  $BIC_{cubic} = 244,425$ ,  $BIC_{log} = 244,986$ ,  $BIC_{\Delta} = 521$ , where smaller BICs indicate better fit). Consequently, in both the aggregate and stratified models, rate of change was modeled as a cubic function of time.

*Aggregate model.* The aggregate model was as follows:

$$Y_{ijk} = \beta_{00} + \beta_{10}(session)_{ij} + \beta_{20}(session)_{ij}^2 + \beta_{30}(session)_{ij}^3 + \left[ b_{00j} + b_{10j}(session)_{ij} + b_{20j}(session)_{ij}^2 + c_{10k}(session)_{ij} + e_{ijk} \right],$$

where  $Y_{ijk}$  is the OQ-45 score at time  $i$  for person  $j$  seeing therapist  $k$ ;  $\beta_{00}$  is the overall intercept (i.e., average OQ-45 score at the beginning of treatment); and  $\beta_{10}$ ,  $\beta_{20}$ , and  $\beta_{30}$  are the average linear, quadratic, and cubic rates of change, respectively (i.e., fixed effects). The parameters inside the brackets represent the random effects. The random effects accounted for patient variability around the overall intercept ( $b_{00j}$ ), linear rate of change ( $b_{10j}$ ), and quadratic rate of change ( $b_{20j}$ ). The model also included a random effect that accounted for therapist variability around the linear rate of change ( $c_{10k}$ ). We were not able to estimate a random effect for individual level variability around the cubic rate of change or a random effect for therapist variability around the quadratic or cubic rates of change. We did estimate a model with a random effect for therapist variability around the intercept and the linear rate of change. However, that model did not fit the data as well as the above model. Consequently, we did not include a random effect for therapist variability around the intercept in the final model.

*Stratified model.* The stratified model was as follows:

$$Y_{ijk} = \beta_{00} + \beta_{01}(\#sessions)_j + \beta_{10}(session)_{ij} + \beta_{20}(session)_{ij}^2 + \beta_{30}(session)_{ij}^3 + \beta_{11}(\#sessions)_j(session)_{ij} + \beta_{21}(\#sessions)_j(session)_{ij}^2 + \beta_{31}(\#sessions)_j(session)_{ij}^3 + \left[ b_{00j} + b_{10j}(session)_{ij} + b_{20j}(session)_{ij}^2 + c_{10k}(session)_{ij} + e_{ijk} \right].$$

<sup>3</sup> We replicated the analyses with patients who met all the inclusion criteria but did not necessarily have demographic and diagnostic information. The results were consistent with the reported analyses.

The random effects portions of the stratified model and aggregate models are identical. The fixed effects portion added a main effect for total number of sessions ( $\beta_{01}$ ), as well as interactions between the linear, quadratic, and cubic rates of change and total number of sessions ( $\beta_{11}$ ,  $\beta_{21}$ ,  $\beta_{31}$ , respectively). The main effect for total number of sessions tested whether those who attended different numbers of sessions had different baseline OQ-45 scores. The interactions provided a test of whether rate of change varied as a function of the total number of sessions attended. Because total number of sessions was notably positively skewed, we used the natural log transformation for sessions.<sup>4</sup>

To rule out the influence of potential confounds, we reestimated the models including age (grand mean centered), ethnicity (minority vs. nonminority), sex, marital status (single vs. married), and diagnosis. Diagnosis consisted of three dummy variables comparing adjustment disorders, mood disorders, and anxiety disorders with other disorders (i.e., the "other" category constituted the reference group). All variables were included as main effects as well as interactions with the linear, quadratic, and cubic forms of sessions. We also allowed the demographic and diagnostic variables to interact with total dose. However, these interactions were excluded from the final model because none were significant.

*Clinically significant improvement.* We examined whether the likelihood of clinically significant change was related to the number of sessions patients attended (Barkham et al., 2006; Stiles et al., 2008). To assess this relationship, we used Jacobson and Truax's (1991) criteria to establish whether patients achieved reliable and clinically significant improvement (RCSI) at the end of treatment. To achieve RCSI, patients have to begin above the clinical cutoff for the OQ-45 (total score > 63) and change at least 14 points during therapy (Lambert et al., 2004). Because patients cannot achieve RCSI if they do not begin treatment above the clinical cutoff, we limited these analyses to only those patients who began treatment above the clinical cutoff (cf. Stiles et al., 2008). We examined the relationship between RCSI status and total dose of treatment with logistic regression. Because we were interested in the relationship between dose and clinically significant improvement, we did not categorize patients into other improvement categories (e.g., reliable improvement, deterioration, or no reliable change).

## Results

### Descriptive Data

Patients attended an average of 6.46 sessions ( $SD = 4.14$ , range = 3–29,  $Mdn = 5$ ) and remained in therapy an average of 10.4 weeks ( $SD = 8.30$ , range = 1–40,  $Mdn = 7.43$ ). Pooling across all patients, the mean first session OQ-45 score was 71.41 ( $SD = 22.19$ , range = 10–149,  $Mdn = 72$ ), and the mean last session OQ-45 score was 56.83 ( $SD = 23.51$ , range = 0–164,  $Mdn = 56$ ). There was a small positive correlation between total sessions attended and first session OQ-45 scores ( $r = .09$ ,  $p < .001$ ). Thus, first session symptom level accounted for between 1% and 9% of the variance in total number of sessions attended (Ozer, 1985). The correlation between total sessions attended and final session OQ-45 scores approached significance ( $r = .02$ ,  $p = .09$ ).

### Rate of Change

The results of the aggregate model were generally consistent with previous dose–response analyses. Table 1 presents the coef-

Table 1  
*Multilevel Growth Curve Models Predicting Change in OQ-45 During Treatment*

Variable	Coefficient	
	Aggregate model	Stratified model
Fixed effects		
Intercept ( $\beta_{00}$ )	70.48**	63.38**
Session ( $\beta_{10}$ )	−4.13**	−9.67**
Session <sup>2</sup> ( $\beta_{20}$ )	0.30**	1.02**
Session <sup>3</sup> ( $\beta_{30}$ )	−0.007**	−0.06**
#Sessions <sup>a</sup> ( $\beta_{01}$ )		4.17**
Session × #Sessions <sup>a</sup> ( $\beta_{11}$ )		2.69**
Session <sup>2</sup> × #Sessions <sup>a</sup> ( $\beta_{21}$ )		−0.29**
Session <sup>3</sup> × #Sessions <sup>a</sup> ( $\beta_{31}$ )		0.02**
Random effects		
Variance estimates		
Between patient		
Initial status ( $\tau_{b00}$ )	554.54	550.83
Session ( $\tau_{b10}$ )	24.10	21.95
Session <sup>2</sup> ( $\tau_{b20}$ )	0.08	0.07
Between therapist		
Session ( $\tau_{c10}$ )	0.13	0.11
Residual	129.56	128.79

*Note.* #Sessions = total number of sessions a patient attended; OQ-45 = Outcome Questionnaire-45.

<sup>a</sup> Because total number of sessions was positively skewed, we used the natural log transformation for sessions.

\*\*  $p < .01$ .

ficients for the aggregate and stratified models. As noted above, change during treatment followed a cubic pattern. The linear ( $\beta_{10} = -4.13$ ), quadratic ( $\beta_{20} = 0.30$ ), and cubic ( $\beta_{30} = -0.007$ ) coefficients were all significant ( $p < .01$ ). The top panel of Figure 1 illustrates the cubic pattern over 20 sessions of treatment. In the aggregate model, the average rate of change during early therapy was relatively steep but began to flatten out between Session 8 and Session 10. Patients did not change much between Session 10 and Session 20, although there were slight upward and downward fluctuations between these time points. This model is consistent with previous dose–effect models that suggest that change is most rapid in early sessions and tapers off during later sessions (e.g., Howard et al., 1986; Kopta et al., 1994).

The stratified model suggested that patients who came for different numbers of sessions changed at different rates, indicating that the aggregate model does not accurately reflect the pattern of change for patients who received different doses of treatment. The stratified model fit the data significantly better than the aggregate model,  $\chi^2(4) = 428.49$ ,  $p < .01$ . Similar to the aggregate model, in the stratified model the linear ( $\beta_{10} = -9.67$ ), quadratic ( $\beta_{20} = 1.02$ ), and cubic ( $\beta_{30} = -0.06$ ) terms were significant ( $p < .01$ ; see Table 1). The main effect for the log of total sessions was

<sup>4</sup> In a supplementary analysis, we fit a model where we estimated rate of change separately for each dosage group rather than using an interaction. The results were essentially the same as the models we report, and fit indices suggested that the model we report was preferable.

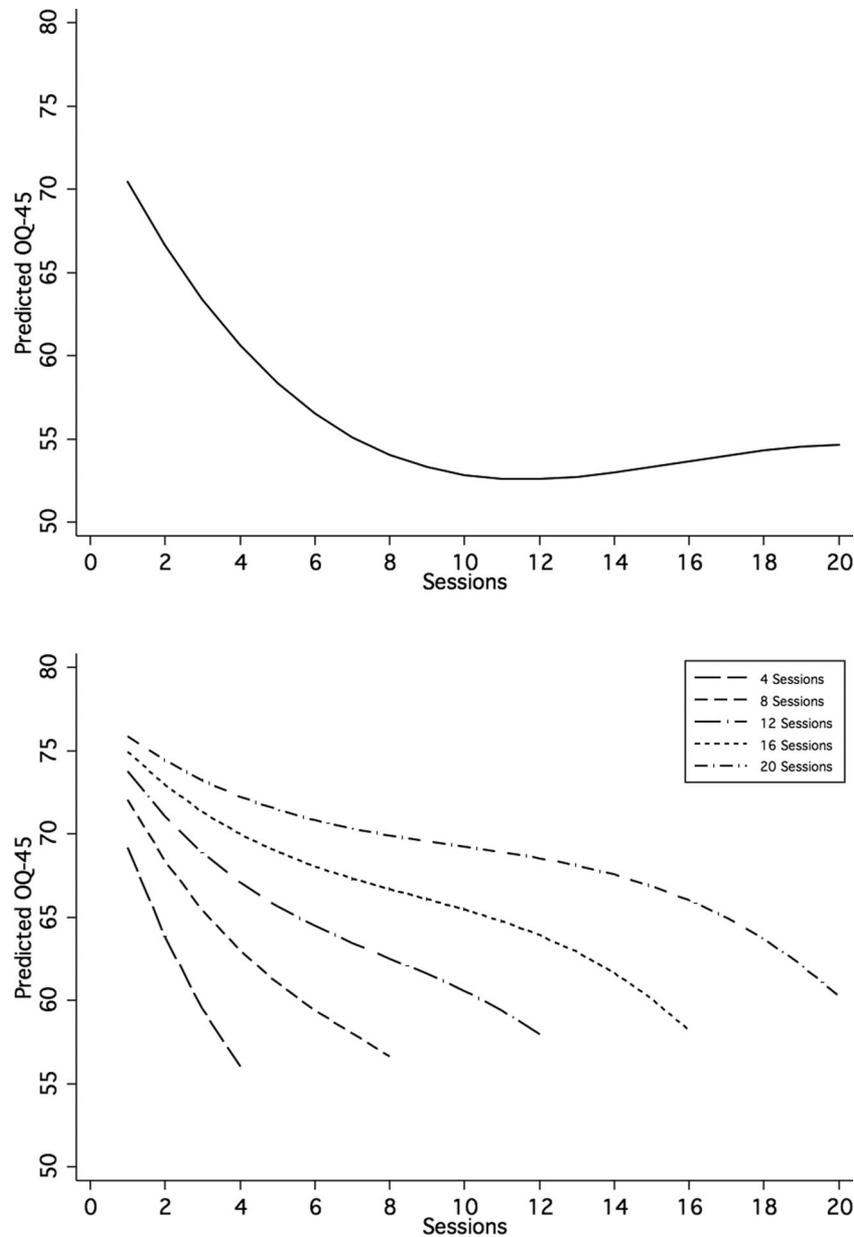


Figure 1. Predicted rate of change in Outcome Questionnaire-45 (OQ-45) scores across sessions in treatment. The top panel represents the aggregate model, which averaged the rate of change across all patients, ignoring the total number of sessions patients attended. The bottom panel represents the stratified model, which stratified rate of change across the total number of sessions patients attended.

significant ( $\beta_{01} = 4.17$ ,  $p < .01$ ), indicating that there were differences in Session 1 OQ-45 scores among people who attended different numbers of sessions. Specifically, patients with higher OQ-45 scores attended more total sessions. The interactions between the log of total sessions and the linear ( $\beta_{11} = 2.69$ ), quadratic ( $\beta_{21} = -0.29$ ), and cubic ( $\beta_{31} = 0.02$ ) forms of session (i.e., the time variable) were all significant ( $p < .01$ ). These significant interactions indicate that rate of change was related to the total number of sessions a person attended. As can be seen in the bottom panel of Figure 1, on average patients improved over

time regardless of total dose. However, the fewer sessions patients attended, the faster their rate of change. Thus, number of sessions attended appeared to reflect the speed at which people change—how long patients stayed in treatment depended upon how they responded to treatment.

Like the aggregate model, the stratified model suggested that within a stratum (e.g., 15 visits), the amount of change per session tended to decrease across time. However, the decrease was not as sharp in the stratified model as in the aggregate model. In fact, the sharp decrease in rate of change observed in the aggregate model

was a consequence of the fact that as time went on, those people who responded to treatment rapidly dropped out of therapy.

The results were not significantly affected by including age, ethnicity, sex, marital status, and diagnosis in the model. The main effects for the demographic variables and diagnosis as well as their interactions with the session variables were significant. The main effects for log of total sessions ( $\beta_{01} = 2.35$ ) and the linear ( $\beta_{11} = -9.88$ ), quadratic ( $\beta_{21} = 1.08$ ), and cubic ( $\beta_{31} = -0.06$ ) forms of session all remained significant ( $p < .01$ ) and were of similar magnitude to the model that excluded the demographic and diagnosis variables. Likewise, the interactions between the log of total sessions and the linear ( $\beta_{11} = 2.82$ ), quadratic ( $\beta_{21} = -0.32$ ), and cubic ( $\beta_{31} = 0.02$ ) forms of session all remained significant ( $p < .01$ ) and were of a similar magnitude to the model that excluded the demographic and diagnosis variables.

Given that the data were drawn from a university-based counseling center, it is possible that therapy termination was influenced by academic year timing (i.e., end of semester, term, or quarter). We explored this issue by identifying patients who terminated within 14 days of the end of a given semester. If patients terminated because of the end of the semester and not because they reached their GEL, we expected them to have higher termination scores than other patients because they had left treatment prematurely. However, patients who terminated within 14 days of the end of a semester did not differ significantly from other patients on their final session OQ-45 score. Furthermore, including timing of termination in our final models did not affect our results.

### *Clinically Significant Improvement*

Patients who started below the clinical cutoff for the OQ-45 could not have achieved clinically significant change as defined by Jacobson and Truax (1991). Consequently, for the clinically significant improvement analyses, we limited the data to those patients who started above the clinical cutoff ( $N = 2,985$ ). A total of 1,242 patients achieved clinically significant change, which represents 41.6% of all patients above the clinical cutoff.

Table 2 presents the rates of clinically significant change stratified by the total number of sessions attended. In Table 2 we combined patients who attended 18 or more sessions (95th percentile and above) into one category to conserve space. However, the analyses discussed below did not aggregate patients who attended 18 or more sessions. The relationship between sessions and RCSI status appeared nonlinear. We used logistic regression to predict RCSI status from log of sessions, log of sessions squared, demographics, and diagnosis. As can be seen in Table 3, the linear term for log of sessions was significant and the quadratic term was very close to significant. Both age, which was grand mean centered, and being single were related to RCSI status. Specifically, being above the mean age (22.3) and being single were both associated with decreased odds of achieving RCSI status. To aid in the interpretation of the logistic regression model, Figure 2 presents the predicted percentage of patients expected to have achieved RCSI by number of total sessions. As can be seen, the predicted percentages of patients achieving RCSI increased with increases in dose up to about Session 8. Although this relationship is statistically significant, it is relatively small. The difference between percentages at Session 3 and Session 8 is 9.8 percentage points. After Session 8, the predicted percentages fluctuate slightly

Table 2  
*Rates of Reliable and Clinically Significant Change Stratified by Total Number of Sessions*

Sessions attended	Total number of patients	Patients above cutoff	Patients achieving RCSI	
	<i>N</i>	<i>N</i>	<i>N</i>	%
3	1,195	706	253	35.84
4	843	520	210	40.38
5	597	381	154	40.42
6	418	270	114	42.22
7	311	208	90	43.27
8	257	172	80	46.51
9	229	153	73	47.71
10	152	100	50	50.00
11	128	92	43	46.74
12	110	76	36	47.37
13	93	60	25	41.67
14	82	63	31	49.21
15	43	32	17	53.12
16	41	34	16	47.06
17	32	23	9	39.13
18+	145	95	41	43.16

Note. RCSI = Reliable and clinically significant improvement.

but remain relatively constant, suggesting no relationship between dose and clinically significant improvement after Session 8.

### Discussion

Most dose–response research has combined data across patients who vary in their total dose of treatment. Consistent with the dose–effect model, many researchers have ignored the variability in total dose and have assumed that the average rate of change during therapy is constant across doses. Thus, the negatively accelerating pattern of change is often interpreted as the expected pattern of change in psychotherapy (e.g., Kopta, 2003; Kopta et al., 1994; Lueger et al., 2001; Lutz et al., 1999, 2001). In contrast, proponents of the GEL model suggest that rate of change varies as a function of total dose. The GEL model predicts that dose is a reflection of treatment response—people generally remain in treatment until they get better (Barkham et al., 2006; Stiles et al., 2008). Thus, patients who stay for only a few sessions change at a faster rate than those who stay many sessions. Moreover, patients who attend many sessions will be no more likely to experience clinically significant change than those who attend few sessions. Our results were most consistent with the GEL model. Specifically, rate of change was related to total dose of treatment—small doses were related to relatively fast rates of change, whereas large doses were related to slow rates of change. There was a nonlinear relationship between total number of sessions and likelihood of recovery, which provided mixed support for the GEL and dose–effect models. Early on in therapy there was a small increase in the likelihood of recovery with increasing doses of treatment, which is consistent with the dose–effect model. However, after about Session 8, there was no relationship, which is consistent with the GEL model.

These results provide little evidence of the dose–response relationship predicted by the dose–effect model. Instead, on average patients appear to remain in treatment until they have achieved

Table 3  
*Logistic Regression Models Predicting Reliable and Clinical Significant Improvement*

Variable	Odds ratio
#Sessions <sup>a</sup>	3.08*
#Sessions <sup>2a</sup>	0.81 <sup>†</sup>
Session 1 OQ-45	0.98**
Age <sup>b</sup>	0.96***
Minority	0.82
Female	1.02
Single	0.82*
Anxiety disorder <sup>c</sup>	0.97
Mood disorder <sup>c</sup>	1.14
Adjustment disorder <sup>c</sup>	1.18

Note. #Sessions = total number of sessions a patient attended; OQ-45 = Outcome Questionnaire-45.

<sup>a</sup> Because total number of sessions was positively skewed, we used the natural log transformation for sessions. <sup>b</sup> Age was grand mean centered. <sup>c</sup> The reference category for these dummy variables consisted of individuals diagnosed with a disorder other than an anxiety, mood, or adjustment disorder.

<sup>†</sup>  $p < .10$ . \*  $p < .05$ . \*\*  $p < .01$ .

sufficient change—their GEL—and then terminate treatment. Therefore, in naturalistic data, dose is not a predictor of treatment response but a marker of the speed of treatment response. This is consistent with Barkham et al.'s (2006) population interpretation of dose–response relationships. The population interpretation is based on an analogy to the effects of insecticides in agriculture, where low doses of insecticide are needed for weak insects and large doses for hardy insects. As Barkham et al. pointed out, “As applied to psychotherapy, this population interpretation does not suggest that increasing doses lose potency—for example, that the 10th session tends to be less powerful than the 2nd—but instead that the easy-to-treat clients have responded by the 10th session, so only the hard-to-treat or resistant remain” (p. 165).

It is interesting that Howard et al. (1986) stratified their sample by total dose and found that higher doses of treatment were associated with an increased likelihood of improvement, which is in contrast to our results as well as Barkham et al. (2006) and Stiles et al. (2008). The discrepant findings may be a consequence of the different methodologies used by Howard et al. and the more recent studies. For example, the present study as well as Barkham et al. and Stiles et al. used patients' ratings of improvement on well-established outcome measures and a replicable formula (Jacobson & Truax, 1991) to establish clinically significant change, whereas Howard et al. used either patients', therapists', or researchers' ratings of global improvement to establish clinically significant change. Future dose–response research should use replicable methodologies (e.g., valid measures of outcome, clinical significance formulae) to allow better comparisons across studies.

It is unclear whether the variability in treatment response seen in the stratified model actually indicates different populations of patients. Consequently, an important area for future research is to identify whether different populations exist and what variables predict population membership. Potential variables could be patient variables, therapist variables, treatment variables, or any combination of the three. For example, our analysis indicated that compared to patients with low distress levels at intake, patients

with high distress levels attended more sessions and thus changed at a slower rate. Others have shown that patients with high levels of baseline distress change slowly (Lutz et al., 1999) and that characterological symptoms change more slowly than distress symptoms (Kopta et al., 1994; Pilkonis & Frank, 1988), although most of this research has not accounted for varying doses of treatment.

Growth mixture models provide an excellent methodology for identifying the number of populations and for predicting population membership. Growth mixture models use longitudinal data to identify classes of people on the basis of their pattern of change over time. Mixture models also allow researchers to incorporate variables that predict class membership. For example, Stulz, Lutz, Leach, Lucock, and Barkham (2007) used growth mixture models to explore the shape of change during the first six sessions of therapy among 192 patients. They identified five classes of people. Three classes showed little change and were distinguished largely by their initial symptom severity. The other two classes showed change during the first six sessions, although for one class the change was rapid. Symptom level, as measured by the Beck Depression Inventory and Beck Anxiety Inventory and age, predicted class membership. Consistent with our results, classes with small amounts of change during the first six sessions consisted of patients who had higher doses of treatment compared to classes consisting of people who improved significantly in the first six sessions. However, as with most other research using naturalistic data, Stulz et al. did not specifically incorporate dose into their models.

It might be tempting to conclude from our results that there is no dose–response relationship in psychotherapy. After all, there was only a small relationship between total dose and amount of change or likelihood of recovery. However, if it is true that dose of therapy is an indicator of different populations of patients, then it is impossible to tell whether the lack of a clear dose–response relationship is an artifact of combining these different populations. As Feaster, Newman, and Rice (2003) have pointed out, the true dose–response question is, “Do otherwise equal clients show different results when given different levels of a particular type of therapy?” (p. 353). Although statistical controls, like those used in

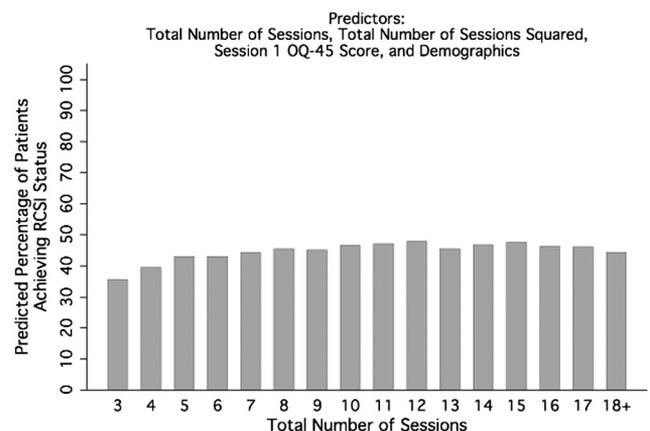


Figure 2. Predicted percentage of patients achieving reliable and clinically significant improvement (RCSI) on the Outcome Questionnaire-45 (OQ-45).

this study, can help address this question, random assignment of patients to different doses of treatment will provide the clearest answer. For example, it is possible that if we randomly assigned the group of patients who came for 8 total sessions to receive either 8 sessions or 16 sessions, the patients attending 16 sessions would change more (Barkham, Rees, Shapiro, et al., 1996; Shapiro et al., 1994).

### *Limitations*

A limitation of this study is that we used only a single outcome variable. It is unclear whether these results will generalize to other measures of outcomes, although similar research suggests that it is likely (Barkham et al., 2006; Beckstead et al., 2003; Stiles et al., 2008). A second limitation is that we did not have information on the specific treatment delivered to each patient—a common limitation of naturalistic data. Thus, we were not able to explore whether the relationship between dose and rate of change varies across treatment type. A third limitation is that the sample of participants was predominantly White and was drawn from a university counseling center. It is unclear whether these results would generalize to more ethnically diverse samples and other clinical populations. A fourth limitation is that we do not have specific information about why patients terminated therapy. Although our results are consistent with the predictions of the GEL model, we cannot verify each patient's reason for termination. This is an important area of future research on the GEL model. However, this limitation does not affect the central point of this article—session-to-session change is not, on average, equal across patients. A fifth limitation is that diagnostic information was drawn from therapists' clinical impressions. Thus, it is unclear whether the results would change if diagnoses had been obtained from a structured clinical interview. Finally, our results are limited to patients who attended at least three sessions. However, both Barkham et al. (2006) and Stiles et al. (2008) included patients who attended less than three sessions in their analyses, and the results in both studies were consistent with the GEL model.

### *Research Implications*

Our results have implications for researchers using naturalistic data. A strength of multilevel models (or hierarchical linear modeling) in analyzing longitudinal data is that they can accommodate situations where participants vary in their number of observations. Thus, these models seem ideal for naturalistic data because patients attend different numbers of sessions and thus have different numbers of observations. However, multilevel models assume that the participants' data are missing at random (see Singer & Willet, 2003, for a readable and more detailed discussion of the missingness assumptions of multilevel models). A patient who attends four sessions, achieves clinically significant change, and then terminates, left treatment for nonrandom reasons and is not missing data for sessions five and above. Measurement could not occur at those sessions because these sessions never occurred. Researchers could measure patients independent of therapy, but that changes the meaning of the parameters, and they would no longer be modeling "change during therapy." The upshot is that researchers using naturalistic psychotherapy data need to carefully consider the

structure and meaning of their data and evaluate whether their data meet the assumptions of the statistical procedures they use.

The difficulties in using multilevel models with naturalistic data have implications for patient-focused research (Lambert, 2007; Lutz, Martinovich, Howard, & Leon, 2002). A major emphasis of patient-focused research has been the identification of expected rates of change for various settings and patient groups. Once the expected rates of change have been established, an individual patient's progress can be tracked and compared to the expected rate of change. If a patient is not progressing as expected, feedback can be given to the patient and his or her therapist and corrective action can be taken. Researchers have largely used multilevel models and naturalistic data where patients' doses vary to establish the expected rates of change (e.g., Lutz et al., 2002). Given the relationship between rate of change and total dose, it is unclear whether the expected rates of change reported in the literature are accurate. Future research in this area should incorporate total dose into the expected rates of change. Given that total dose is not known until the end of treatment, researchers would need to track a patient's progress and, given the patient's progress at any given point, determine the probability that the patient falls into one of the total dose categories. Thus, the systematic differences in rate of change would be reflected in the expected pattern of change and feedback to therapists and patients.

Our results have implications for studying change not only in naturalistic studies but also clinical trials. In most clinical trials, dose is fixed for patients regardless of their response to treatment. This has the practical advantage of putting a time limit on treatment so that the study does not go on indefinitely. It also has the obvious design advantage of controlling for total dose of treatment. However, fixing dose does not reflect the clinical reality that patients change at different rates and thus need different doses of treatment. If research is going to speak to actual clinical phenomena, then more attention needs to be paid to how much treatment a given patient needs.

### *Clinical Implications*

In their seminal article on dose–response relationships, Howard et al. (1986) argued that the negative accelerating dose–response curve could be used to set time limits on psychotherapy (cf. Kopta, 2003). However, our results suggest that the negatively accelerating curve documented by Howard et al. (1986) and more recent dose–effect research is not an accurate representation of change during treatment for the individual patient. Rather, there is substantial variability in response to treatment, which suggests that uniform time limits would not adequately serve patients' needs (Barkham et al., 2006; Stiles et al., 2008). Consequently, a better understanding of what factors—patient, therapist, treatment, and contextual—influence treatment response is imperative. This is especially true for treatment nonresponse. Indeed, this understanding could provide strategies and direction to clinicians treating patients who are not making improvement or are even deteriorating (Lambert, 2007). That is, this research could help clinicians provide individualized, responsive treatment.

### References

- Barkham, M., Connell, J., Stiles, W., Miles, J. N., Margison, F., Evans, C., et al. (2006). Dose–effect relations and responsive regulation of treat-

- ment duration: The good enough level. *Journal of Consulting and Clinical Psychology*, 74, 160–167.
- Barkham, M., Rees, A., Shapiro, D. A., Stiles, W. B., Agnew, R. M., Halstead, J., et al. (1996). Outcomes of time-limited psychotherapy in applied settings: Replicating the Second Sheffield Psychotherapy Project. *Journal of Consulting and Clinical Psychology*, 64, 1079–1085.
- Barkham, M., Rees, A., Stiles, W. B., Shapiro, D. A., Hardy, G. E., & Reynolds, S. (1996). Dose–effect relations in time-limited psychotherapy for depression. *Journal of Consulting and Clinical Psychology*, 64, 927–935.
- Bates, D. (2007). *lme4: Linear mixed-effects models using Eigen and Eigen++* (R package Version 0.99875–9) [Computer software]. Retrieved from <http://www.r-project.org/>
- Beckstead, D. J., Hatch, A. L., Lambert, M. J., Eggett, D. L., Goates, M. K., & Vermeersch, D. A. (2003). Clinical significance of the outcome questionnaire (OQ-45.2). *Behavior Analyst Today*, 4, 79–90.
- Feaster, D., Newman, F., & Rice, C. (2003). Longitudinal analysis when the experimenter does not determine when treatment ends: What is dose–response? *Clinical Psychology and Psychotherapy*, 10, 352–360.
- Grissom, G. R., Lyons, J. S., & Lutz, W. (2002). Standing on the shoulders of a giant: Development of an outcome management system based on the dose model and phase model of psychotherapy. *Psychotherapy Research*, 12, 397–412.
- Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose–response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice*, 9, 329–343.
- Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose–effect relationship in psychotherapy. *American Psychologist*, 41, 159–164.
- Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology*, 61, 678–685.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.
- Kopta, S. M. (2003). The dose–effect relationship in psychotherapy: A defining achievement for Dr. Kenneth Howard. *Journal of Clinical Psychology*, 59, 727–733.
- Kopta, S. M., Howard, K. I., Lowry, J. L., & Beutler, L. E. (1994). Patterns of symptomatic recovery in psychotherapy. *Journal of Consulting and Clinical Psychology*, 62, 1009–1016.
- Lambert, M. J. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17, 1–14.
- Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., et al. (2004). *Administration and scoring manual for the OQ-45.2*. Orem, UT: American Professional Credentialing Services.
- Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 139–193). New York: Wiley.
- Lueger, R. J., Howard, K. I., Martinovich, Z., Lutz, W., Anderson, E. E., & Grissom, G. (2001). Assessing treatment progress of individual patients using expected treatment response models. *Journal of Consulting and Clinical Psychology*, 69, 150–158.
- Lutz, W., Lowry, J., Kopta, S. M., Einstein, D. A., & Howard, K. I. (2001). Prediction of dose–response relations based on patient characteristics. *Journal of Clinical Psychology*, 57, 889–900.
- Lutz, W., Martinovich, Z., & Howard, K. I. (1999). Patient profiling: An application of random coefficient regression models to depicting the response of a patient to outpatient psychotherapy. *Journal of Consulting and Clinical Psychology*, 67, 571–577.
- Lutz, W., Martinovich, Z., Howard, K. I., & Leon, S. C. (2002). Outcomes management, expected treatment response, and severity-adjusted provider profiling in outpatient psychotherapy. *Journal of Clinical Psychology*, 58, 1291–1304.
- Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin*, 97, 307–315.
- Pilkonis, P. A., & Frank, E. (1988). Personality pathology in recurrent depression: Nature, prevalence, and relationship to treatment response. *American Journal of Psychiatry*, 145, 435–441.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- R Development Core Team. (2007). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.r-project.org>
- Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology*, 62, 522–534.
- Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Snell, M. N., Mallinckrodt, B., Hill, R. D., & Lambert, M. J. (2001). Predicting counseling center clients' response to counseling: A 1-year follow-up. *Journal of Counseling Psychology*, 48, 463–473.
- Stiles, W., Barkham, M., Connell, J., & Mellor-Clark, J. (2008). Responsive regulation of treatment duration in routine practice in United Kingdom primary care settings: Replication in a larger sample. *Journal of Consulting and Clinical Psychology*, 76, 298–305.
- Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice*, 5, 439–458.
- Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, 75, 864–874.

Received May 13, 2008

Revision received January 12, 2009

Accepted January 13, 2009 ■