Validating the Rapid Responder Construct Within a Practice Research Network

Samuel S. Nordberg,[1] Louis G. Castonguay,[2] Aaron J. Fisher,[3] James F. Boswell,[4] and David Kraus[5]

[1] The Center for Healthcare Organization and Implementation Research
[2] The Pennsylvania State University
[3] University of California, Berkeley
[4] State University of New York
[5] Outcome Referrals

Objective The present study was a replication and extension of prior work (Stulz, Lutz, Leach, Lucock, & Barkham, 2007) that identified multiple groups of clients in treatment with high-symptom severity and markedly different recovery trajectories (rapid/early response vs. little or no response). **Method** Using data collected through repeated administrations of the Depression subscale of the Treatment Outcome Package (n = 147), growth mixture modeling was employed to determine whether clients fell into discrete groups of response trajectories during 15 sessions of psychotherapy. Additionally, logistic regressions were conducted to assess possible predictors of group membership. **Results** Three separate groups of treatment responders were identified: 2 high-symptom groups–rapid responders and nonresponders–and 1 low-symptom group of nonresponders. Elevated social conflict and suicidality predicted increased likelihood of membership in the high-symptom nonresponder group. Increased feelings of interpersonal hostility and better sexual functioning predicted increased likelihood of membership in the rapid responder group. **Conclusion** Replication of earlier results provides further evidence for the usefulness of modeling change during psychotherapy using multiple trajectories. Predictors of group membership indicate the influence of functional impairment on recovery, and support the importance of multidimensional measurement of client problems. © 2014 Wiley Periodicals, Inc. J. Clin. Psychol. 70:886–903, 2014.

Keywords: patient-focused research; shapes of change; growth mixture modeling; prediction of change in psychotherapy

Over the past few decades, conventional randomized controlled trials (RCTs) and–to a lesser extent–naturalistic studies of psychotherapy have demonstrated the effectiveness of many different treatments. Research has established that psychotherapy works (Lambert & Ogles, 2004); however, significantly less is known about how, and for whom, psychotherapy works (Paul, 1967). Thus, in addition to establishing the efficacy and effectiveness of psychotherapy, complementary questions and methods might provide us with more information regarding the way in which clients change in treatment.

In recognition of this, Division 12 (Society of Clinical Psychology) of the American Psychological Association (APA) has called for increased attention to aspects of therapeutic change other than the aggregate effect of particular forms of therapy at the conclusion of treatment and, in the interest of elucidating such information, has endorsed using data from already completed psychotherapy cases to pursue this (Weisz, Hawley, Pilkonis, Woody, & Follette, 2000). Additionally, some researchers have called for methodological modifications to the traditional pre-post design often used in RCTs, such as the use of repeated measurements of important

*Dropout and missing data made including more than four time points a poor choice, as the sample size fell after session 15 (to 118 of the original 147).

Please address correspondence to: Samuel S. Nordberg, Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System (152M), 150 South Huntington Ave., Boston, MA 02130. E-mail: Samuel.nordberg@va.gov

Published online in Wiley Online Library (wileyonlinelibrary.com/journal/jclp). DOI: 10.1002/jclp.22077

outcomes (Fisher, Newman, & Molenaar, 2011; Hayes, Laurenceau, & Cardaciotto, 2007) to further explore within-treatment change.

Research that utilizes repeated measurements of individual symptom profiles and attempts to identify what works for whom has been termed "patient-focused research" (Howard, Moras, Brill, Martinovitch, & Lutz, 1996). A core component of patient-focused research is the study of client variables that influence or predict the client's progress through therapy. Recently, some studies have sought to elucidate relationships between client characteristics and expected treatment course to predict individual treatment response (Lambert, Harmon, Slade, Whipple, & Hawkins, 2001; Lutz, Lowry, Kopta, Einstein, & Howard, 2001). Based on the premise that clinical practice can be improved through research that focuses on understanding what works for particular clients, such models have shown promise. Some of this research has indeed led to improvements in therapy by providing clinically relevant feedback to therapists regarding clients who are not responding to treatment as would be expected, based on self-reported distress (Lambert, Hansen, & Finch, 2001; Lutz et al., 2006).

Prior work has indicated that a small but meaningful number of clients do not appear to improve significantly over the course of psychotherapy. Indeed, a portion of clients in any given sample–roughly 5%–10%–appears to worsen during treatment (Lambert & Ogles, 2004). Additionally, for those who do change meaningfully for the better, there appear to be multiple trajectories of change, with each trajectory following a different course (e.g., Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009; Stulz et al., 2007). Such findings may offer clues to help clinicians and researchers anticipate how a particular type of client can be expected to respond to treatment. In particular, some have pointed to clients with initially high-symptom severity as warranting further study, both because of the severity of their symptoms and the emerging evidence that there is meaningfully significant variation in response to treatment for individuals with high levels of distress (Stulz & Lutz, 2007).

There is good reason to focus on highly symptomatic clients. With regard to treatment outcome, pretreatment, or baseline, symptom severity is well supported as a predictor of poorer response (Beutler, Blatt, Alimohamed, Levy, & Angtuaco, 2006; Clarkin & Levy, 2004; Newman, Crits-Christoph, Gibbons, & Erickson, 2006). On average, clients with more severe symptomatology at the beginning of psychotherapy continue to exhibit more severe symptomatology after an equivalent dose of psychotherapy (Wampold & Brown, 2005). Essentially, the hierarchy of severity is largely preserved during treatment, yet less is known about how this might relate to any given individual's change in psychotherapy. Do all clients who report similarly high-symptom severity change in the same way?

Recently, Stulz et al. (2007), in a study involving 192 clients in managed care, used growth mixture models (GMM) to identify multiple trajectories of change in the early stages of routine outpatient psychotherapy for a variety of disorders. The two highest-symptom groups exhibited markedly different change trajectories–one rapidly improving and the other slowly worsening–essentially indicating that different groups of high-symptom clients can exist in the same sample. Additionally, in a reanalysis of the Treatment for Depression Collaborative Research Program (TDCRP; Elkin et al., 1989), Lutz, Stulz, and Köck (2009) employed GMMs and found two high-symptom groups of treatment responders, one with a steep trajectory of recovery, the other with a flatter, but still significant, change trajectory. The notion of different trajectories for high-symptom clients is neither novel nor controversial.

Researchers have argued that flatter trajectories for those with high-symptom severity may be due to the correlation between increased symptom severity and reduced client efficacy, wherein these clients have fewer recruitable resources than those with less impairment and/or distress (Hoberman, Lewinsohn, & Tilson, 1988; McLellan et al., 1994). Conversely, regression to the mean (Bland & Altman, 1994; Davis, 1976; Mortin & Torgerson, 2003) has been proposed as an explanation for the group of clients who begin at high levels of severity and quickly recover.

If groups of clients with markedly different recovery trajectories exist, it may be helpful to reexamine our methods for examining change over time. Aggregating across an entire sample may erroneously lead to the conclusion that clients are steadily changing over time, when, in fact, some are changing quite rapidly, while others are not changing at all (or getting worse). Replicating the results of the few studies exploring this phenomenon has implications that may indicate a departure from a single mean as the best measure of client change. It may be that, to capture the heterogeneity in client populations, disaggregation (by trajectory or pretreatment characteristics, for example) may be more appropriate.

While the findings of the Stulz et al. (2007) and Lutz et al. (2009) may indicate the presence of, and an ability to identify, two discrete groups of clients, such results should be interpreted with caution. One challenge in interpreting findings based on GMM techniques is that GMMs are *designed* to parse data into groups, such that a GMM will model multiple trajectories of change (multiple groups) even for samples well-described by a single trajectory (Nagin & Tremblay, 2005) and individual variation around that central tendency (for a more complete discussion of the limitations of GMMs, see Bauer & Curran, 2003a, 2003b). Thus, caution is warranted in interpreting the emergent groups from a GMM as qualitatively different from one another, rather than as clusters of clients distributed around a central tendency.

As Bauer and Reyes (2010) note, with models designed to find groups, there can be difficulty in determining differences in "kind" from differences in "degree." To increase confidence in the "rapid responder" construct, replication (the focus of the current study) and validation studies are needed. Through replication, we may repeatedly indicate the presence of this group, while through validation and theory testing, it may be possible to better describe and predict meaningful differences between this construct and others.

Predictors of Change

As noted above, if the rapid responder phenomenon can be reliably established, then it may be helpful to begin identifying meaningful predictors of individual differences related to group composition (and perhaps, even more importantly, to find predictors of nonresponse).

In a review of client variables that affect treatment outcome generally, Clarkin and Levy (2004) noted that there is inconsistent evidence on predictors of treatment outcome, with a few emerging areas of interest. With regard to diagnosis in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR; American Psychiatric Association, 2000), these authors suggested that comorbid Axis II diagnoses appear to negatively affect treatment outcome. However, they also added that, in isolation, diagnosis fails to capture the complexity of a given client. Indeed, these authors stressed the importance of examining the interaction between diagnosis and other salient client variables. They identified both high initial symptom severity and functional impairment as being related to poorer outcomes in psychotherapy, with functional impairment defined as an individual's overall level of functioning in nonsymptomatic domains (i.e., work history, social relationships, quality of life, and self-care). In contrast, sociodemographic variables–age, gender, socioeconomic status, and race–do not appear to have significant support as predictors of change for clients who remain in psychotherapy.

Beutler and colleagues (2006) largely supported these conclusions. They found additional support for the presence of Axis II as a negative predictor of outcome. Similarly, these authors demonstrated that higher levels of functional impairment at the start of treatment, as well as higher initial symptom severity, predict poorer response to treatment. Beutler et al. (2006) additionally noted that there is evidence to show that short-term treatments are not nearly as effective as long-term treatments for highly dysphoric patients, regardless of the type of treatment offered.

Some of the variables identified above have been further explored with mixed results. Stulz and colleagues (2007), using a global measure of psychological distress, found that increased anxiety reported at intake predicted an increased probability of being in a high-symptom group of clients with steep trajectories of change. They additionally tested age, depression, and interpersonal functioning at intake, but did not find that any of these variables discriminated between rapidly recovering versus slowly worsening high-symptom groups. These authors added that the presence of Axis II pathology was a promising candidate. Cuijpers, van Lier, van Straten, and Donker (2005), using a global measure of distress with a population of clients being treated for major depressive disorder (MDD), found that pretreatment symptoms of anxiety as well as severity of MDD symptoms predicted membership in higher symptom groups.

The current study proposes to further examine patterns of change using a sample of clients from a practice research network in the Northeast United States and measure of general psychological distress. To this end, the aim of this work is to replicate and extend the results of the Stulz et al. (2007) and Lutz et al. (2009) studies. Specifically, we will focus on the presence of the two putatively different, high-symptom groups–exploring whether they replicate under different conditions and with a multidimensional self-report measure, and attempting to extend the body of work by examining the role of meaningful predictors of group membership. The establishment of class predictors may facilitate the understanding of the qualitative differences in group composition and related treatment response. Thus, the current study will test several pretreatment variables that have been identified in the literature as potentially meaningful.

Given these conditions, we had two hypotheses. First, we hypothesized that we would find, among multiple groups of treatment responders, two groups of clients who exhibited high symptoms at the first session of psychotherapy: rapid responders and nonresponders; thus, we expected to replicate the results of Stulz et al. (2007). Second, we hypothesized that higher functional impairment, higher symptom severity, and presence of Axis II disorder at intake (approximately 4 weeks prior to the first session) would predict membership in the nonresponder group.

Method

Participants

Data were collected as part of a routine practice from clients attending weekly individual psychotherapy sessions at a large university-based psychotherapy training clinic in the Northeast United States between July 2002 and September 2008. This clinic functions as a community mental health center (CMHC) and provides outpatient psychological services to county residents and students with severe mental health problems. Additionally, the clinic serves as a practice research network (see Castonguay, 2011), integrating routine clinical work and research under the same roof. During the data collection period, a total of 70 therapists-in-training treated a total of 668 clients. The average number of clients per therapist was nine, with a range from 1 to 39. Of these 668 clients, 147 met inclusion criteria for the current study by reporting initial distress on the Depression subscale of the Treatment Outcome Package (TOP; see the Measures section) of one or more standard deviations above the normative, nonclinical mean, and attending a minimum of 15 psychotherapy sessions, with four time points–intake, session 1, session 7, and session 15.

There were some missing demographic data, likely due to clients skipping certain questions. Gender data were the most compromised, perhaps because gender is the first question on the TOP registration form and could have been missed. Study clients had an average age of 39 years (standard deviation [SD] = 12, $n = 147$). They were mostly female (71.2%, $n = 104$), and averaged 13.4 years of education ($n = 109$). Sixty-nine percent were single ($n = 144$). The average income was between \$10,000 and \$20,000 annually ($n = 135$), and 86.4% of clients identified themselves as Caucasian ($n = 147$), which is typical of the population served by the clinic as a CMHC. The average participant Depression subscale score was three standard deviations above the nonclinical mean. There were 45 therapists included in the final sample, each treating between 1 and 13 clients.

Measures

TOP. The TOP (Kraus, Seligman, & Jordan, 2005) is a behavioral health assessment and outcome battery designed for clinical and research purposes in naturalistic settings. Developed to meet all of the criteria established by the Society for Psychotherapy Research (SPR) and APA-sponsored Core Battery Conference (Horowitz, Strupp, Lambert, & Elkin, 1997), it assesses a wide array of behavioral health symptoms and functioning, demographics, and case-mix variables. The clinical scales comprise 58 items that assess 12 symptom and functional domains: work functioning, sexual functioning, social conflict, depression, panic (somatic anxiety), psychosis, suicidal ideation, violence, mania, sleep, substance abuse, and quality of life. The TOP reports symptom severity for each of the 12 subscales in terms of standard deviations above or below the normative, nonclinical mean.

Additionally, the TOP assesses general health, stressful life events, treatment goals, and satisfaction with treatment. It has good test-retest reliability (.76 to .94 for the 12 subscales), sensitivity to change, and high levels of convergent validity with scales such as the Beck Depression Inventory (BDI; Beck, Steer, & Ranieri, 1988), the Brief Symptom Inventory (BSI; Derogatis, 1975), and the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Graham, 1993). The TOP requires approximately 5 minutes to complete.

For this sample, the TOP total score was unavailable, because a short version of the TOP was used for repeated measures, after intake, which did not generate a total score but did maintain the psychometric properties of the scales described above. Thus, we chose to use the Depression subscale[1] to track symptom change over time, as it has a high correlation with the Total Score ($r = 0.85$, $n = 255$; Boswell, Kraus, Nordberg, & Castonguay, 2009). In addition, the Depression subscale correlates highly with the SF-36 (Brazier et al., 1992) Mental Health subscale ($r = .82$, $p < .01$), Social Functioning subscale ($r = .75$, $p < .01$), and Vitality subscale ($r = .68$, $p < .01$), the Basis 32 Daily Role Functioning subscale ($r = .84$, $p < .01$), the BSI Anxiety subscale ($r = .70$, $p < .01$), and Psychoticism subscale ($r = .63$, $p < .01$; Kraus et al., 2005). Thus, this subscale appears to be an acceptable proxy for general distress.

The Depression subscale of the TOP comprises 10 items, each rated on a 6-point Likert scale ranging from 1 (All) to 6 (None). The items are as follows: Felt down or depressed, Felt little or no interest in most things, Felt guilty, Felt restless, Felt worthless, Felt tired, slowed down, or had little energy, Worried about things, Had trouble concentrating or making decisions, Noticed your thoughts racing ahead, and Thought about killing yourself or wished yourself dead.

Procedure

Clients were referred to the clinic for psychological services from multiple sources (e.g., physician, county base service unit, and self). Upon initial acceptance, clients were scheduled for an intake interview with a graduate student psychotherapist-in-training. The intake interview comprised a general information-gathering clinical overview with questions related to what was bringing the client to treatment at this time, his or her understanding of the problem, work/school history, and significant interpersonal relationships. This was followed by an oral administration of the Structured Clinical Interview for the DSM-IV (SCID-I; First, Spitzer, Gibbon, & Williams, 2002.). The training clinic calculates reliability on an ongoing basis. For the period in question, overall interrater reliability with the SCID-I was fair (*kappa* = .502). Treatment was provided from two orientations–cognitive behavioral and psychodynamic. All staff in the clinic supervise trainees from one of these two perspectives and practice using the orientation they train.

Before this intake information was collected, as part of routine clinic practice, clients were asked to complete the full length TOP in a waiting area. After intake, clients were assigned to therapists based on the standard practices of the clinic, which account for therapist availability and the appropriateness of the client as an outpatient training case. Typically, clients began their first session roughly 4 weeks after intake. Clients were then asked to fill out the TOP[2] before their 1st, 7th, 15th, and every 15 sessions thereafter (e.g., 30 and 45), again, as part of routine practice in the clinic. For this study, as mentioned above, clients needed to be seen in the clinic for a minimum of 15 sessions.

[1]We, additionally, evaluated change on six other TOP subscales. See Data Analysis.

[2]Over the course of the sampling period, a change in standard procedure resulted in two different versions of the TOP being used as a repeated measure. For some subjects, a short version was used which contains fewer subscales, but retains the structure of those subscales and associated questions completely from the full version.

Data Analysis

Modeling and Predicting Shapes of Change

Repeated measurement facilitates modeling shapes of client change trajectories. From within the field of patient-focused research, GMMs have been employed to examine clients' trajectories through treatment, as well as the effect of predictor variables on those trajectories. Consistent with the recommendations of Laurenceau, Hayes, and Feldman (2007), latent class models such as GMMs help to explore the possibility that repeated measurements of treatment outcome and process are better modeled by multiple trajectories of recovery (Cuijpers et al., 2005; Laurenceau et al., 2007; Stulz & Lutz, 2007; Stulz et al., 2007).

The advantage of GMMs is that, in contrast to single-class or a priori class assumption, the data are allowed to drive class formation and explore the possibility that treatment response is better described by multiple slopes and intercepts, each representing a unique population. The use of such models allows researchers to distance themselves from what Kiesler (1966) eloquently called the "patient uniformity myth." If the sample data are best represented by distinct client populations (or classes), then these can be identified and researchers can ask questions about membership in those classes, such as investigating whether or not particular patient variables differentially predict class membership.

In the present study, the slopes, intercepts and class memberships were calculated using maximum likelihood estimation. Our dependent variable was symptom severity as operationalized by self-report on the Depression subscale of the TOP at the 1st, 7th, and 15th sessions. Of 147 clients, there were 147 observations at the 1st, 7th, and 15th session. Time, coded as months in treatment (i.e., 0, 1.75, and 3.75), was modeled as a linear effect, given that we modeled three time points. Consistent with the methods used by Stulz et al. (2007), we selected a model that allowed for growth parameter (slope and intercept) (co)variances to be freely estimated, but equivalent across classes.[3]

GMMs were generated using Mplus version 6.11 (Muthen & Muthen, 2011). The best fitting model was determined by starting with a one-class model, adding an additional class and then testing for significantly improved fit using the bootstrap likelihood ratio test (BLRT). Recommended most highly for GMMs in a recent Monte Carlo study (Nylund, Asparouhov, & Muthen, 2007), the BLRT assesses the null hypothesis that the data are equivalently explained by a model with one fewer class than the current model. Significant p values indicate that the current model improves fit over a model with one fewer class.

Once the best fit model was found, we employed logistic regression to test the predictive utility of baseline variables for predicting class membership. Because we were interested in predicting membership in the classes that emerged from our initial analyses, we fixed class membership for clients to the classes from the unconditional GMM, and freely estimated growth and logistic regression parameters. We did this to maintain the structure of the initial unconditional GMM.

We selected intake variables (collected approximately 4 weeks before the first session of treatment) from the TOP that would best correspond to predictors of interest. To assess for functional impairment, we included income level, the Work Functioning subscale, the Suicide subscale, the Social Conflict subscale, Life Quality subscale, Sexual Functioning subscale, Mania subscale, Psychosis subscale, and the Violence subscale (which assesses feelings of violence, not acts, and thus will be referred to as the "hostility" subscale[4]). Utilization of the Substance Abuse subscale was precluded by an overabundance of missing data. Because prior work had found increased symptoms of both initial depression and initial anxiety to be meaningful predictors, we included the Depression subscale and the Panic subscale (which assesses avoidance and

[3]We, additionally, evaluated models with growth parameter (co)variances fixed at zero, so called Latent Class Growth Analysis, and models with growth parameters freely estimated and allowed to vary across class. These models are not included for space reasons but replicated the structure of the model reported in the current study (see Results), except that class membership was slightly altered between the various methods. Complete modeling information is available from the first author.

[4]Additionally, in an assessment of concurrent validity, the Violence subscale correlated most highly with the Hostility subscale on the BSI ($r = -.77$, $p < .01$; Kraus, Seligman & Jordan, 2005).

Table 1
BIC, VLMR LRT, BLRT, and Entropy for Models With 1–4 Classes

No. Groups	BIC	VLMR LRT	BLRT	Entropy
1	1605	n/a	n/a	n/a
2	1491	<0.001	<0.001	0.777
3	1490	0.015	0.013	0.710
4	1498	0.260	0.167	0.754

Note. BIC = Bayesian information criteria; VLMR LRT = Vuong Lo Mendal Rubin likelihood ratio test; BLRT = boostrapped likelihood ratio test.

physiological symptoms of anxiety) from the TOP. We additionally included Axis II diagnosis. Due to restrictions in cell size, we tested only presence/absence of personality disorder in our logistic regression.

Following the recommendations of Lambert and Ogles (2009), we also assessed whether the pre-post change demonstrated by each of the groups that emerged from our GMM was clinically significant. We calculated clinically significant change using both cutoff points and a Reliable Change Index (RCI), as suggested by Jacobson and Truax (1991). To meet this gold standard for clinically significant change, clients must move from being above the cutoff to below the cutoff, and clients must change enough such that the data cannot be an artifact of measurement error. Cutoff points are markers that distinguish healthy populations from patient populations. These cutoffs are based on the normative data on both populations and, in our case, were generated using method C (Jacobson & Truax, 1991), which was the most conservative estimate for our data, given significant overlap between the distributions of the clinical and nonclinical samples. Scores above the cutoff point are considered representative of a patient population, while scores below are considered healthy. In the case of the TOP Depression subscale, the cutoff is at .78 standard deviations above 0 (the healthy population mean).

Additionally, as a means for further describing the emergent classes, we evaluated clinically significant change on six clinical subscales—the Panic, Mania, Psychosis, Suicide, Hostility, and Sleep subscales. We selected these subscales for purely practical reasons: They had very little missing data across time points. We excluded the Life Quality, Substance Use, Social Conflict, Work Functioning, and Sexual Functioning subscales due to substantial missing data.

Results

Consistent with Stulz et al. (2007), we built an unconditional GMM by starting with a two-class model and comparing it to a model with one class. We then added a third and fourth class, and tested each new model versus a model with one less class. A three-class model best fit the data, based on the BLRT,[5] as well as BIC (see Table 1). Our model had fewer classes than the model in Stulz et al. (2007), which found five emergent classes. This was likely due to our cutoff criterion of 1 standard deviation above the clinical mean on the depression subscale, which removed from analysis the less symptomatic clients and thus the possibility of finding any low-symptom classes.

The three distinct groups that emerged from the analysis and their linear recovery trajectories are depicted in Figure 1, along with group and individual trajectories for the two high-symptom classes. Primary diagnoses at session 1 for each group are reported in Table 2. Mean subscale scores for session 1 and session 15 are reported in Table 3. One group, which comprised the majority (n = 86, or 58%) of the sample, had an intercept of 1.34 ($p < .001$) and a non-significant slope of −0.057 ($p = 0.15$), and was termed the *low-symptom nonresponders*. The second group (n = 41, or 28%) had an intercept of 3.68 ($p < .001$) and a slope of −0.18 ($p < .001$). We henceforth refer to this group as the *high-symptom nonresponders.* In our sample, the

[5]For the BLRT, we used 500 initial stage starts and 100 final stage starts.

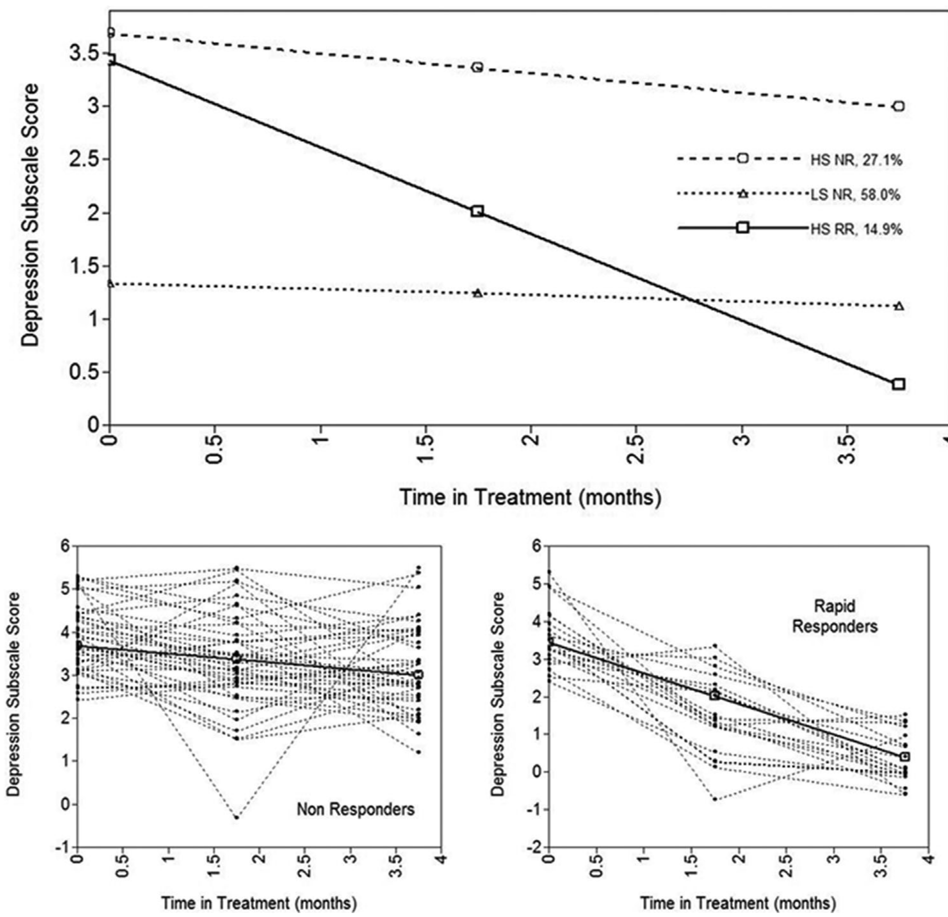

Figure 1. Plotted intercepts and slopes for each responder group, and individual trajectories of high-symptom groups.
Note. Depression severity (y axis) is given in standard deviations above the nonclinical mean, as reported throughout this article. HS NR = high-symptom nonresponder group; HS RR = high-symptom rapid responder group; LS NR = low-symptom nonresponder group.

high-symptom nonresponders appear to represent a significant subpopulation of severely impaired clients who did not exhibit clinically significant change during treatment.

The third group ($n = 20$, or 14%) had an intercept of 3.43 ($p < .001$) and a slope of -0.81 ($p < .001$). This group had a steep trajectory of change such that clients in this group achieved an average pre-post change of over three standard deviations, ending close to the healthy population mean of 0. We termed this group high-symptom rapid responders. The presence of the high-symptom nonresponders and the high-symptom rapid responders in the best fitting model met our first study aim, replication of Stulz et al. (2007), and identification of two groups of clients with similar high levels of severity on the Depression subscale, one of which recovered quickly, while the other did not. Consistent with their results, our groups had similar proportional sizes, with roughly twice as many clients in the high-symptom nonresponders class, relative to the high-symptom rapid responders class.

Clinically Significant Change

To assess the validity of the modeled groups, we calculated clinically significant change statistics (from Jacobson & Truax, 1991; see the Method section for further description) for each client

Table 2
Primary Diagnosis at Session 1 by Group

	High-symptom nonresponders ($n = 41$)		High-symptom rapid responders ($n = 20$)		Low-symptom nonresponders ($n = 86$)	
	Frequency	%	Frequency	%	Frequency	%
Schizoaffective	2	5.00%	-	-	2	2.60%
MDD single episode	3	7.50%	2	10.00%	5	6.50%
MDD recurrent	14	35.00%	8	40.00%	28	36.40%
Bipolar disorder	3	7.50%	1	5.00%	11	14.30%
GAD	1	2.50%	2	10.00%	3	3.90%
Social phobia	2	5.00%	-	-	3	3.90%
OCD	1	2.50%	1	5.00%	1	1.30%
Dysthymic disorder	1	2.50%	1	5.00%	2	2.60%
Borderline PD	1	2.50%	1	5.00%	2	2.60%
PD NOS	3	7.50%	-	-	-	-
PTSD	3	7.50%	-	-	2	2.60%
Adjustment disorder	-	-	-	-	6	7.80%
Other	6	15.00%	4	20.00%	12	15.60%
Total	41 (1)*	100.00%	20	100.00%	86 (9)*	100.00%

Note. The "other" category accounts for the total number of clients who had uncommon, single instance diagnoses or V-codes. MDD = major depressive disorder; GAD = general anxiety disorder; PD NOS = personality disorder not otherwise specified; PTSD = posttraumatic stress disorder; * = missing data for session 1 primary diagnosis due to nonentry by clinician.

and aggregated them by group. The high-symptom rapid responders group most clearly met the criteria for clinically significant change, with 75% of group members both crossing the cutoff point and achieving change in excess of that which would be expected from measurement error and natural fluctuation. The high-symptom nonresponders group appears well characterized, as none of that group met the same criteria. The low-symptom nonresponder group is more mixed, with 18.60% meeting the criteria for clinically significant change.

Of the participants with high initial symptom severity, three nonresponders exhibited greater change than four rapid responders–indicating some degree of misspecification among classes. Indeed, we would expect any means of binning clients to be inappropriate for some not well captured by either class (Meehl, 1992). To further examine whether the emergent classes were appropriately defined (as rapid/nonresponders), we examined the selected additional clinical subscales of the TOP for change over time. These were the Sleep Functioning, Suicidality, Hostility, Mania, Panic, and Psychosis subscales.

The percentage of each class that met the criteria for clinically significant change on each of the seven clinical subscales is shown in Table 3. These descriptors reinforce the class labels, as the high-symptom rapid responders group exhibited at least twice as much clinically significant change as the other two groups, while the nonresponder groups showed little change.

Logistic Regression Results

We tested potential predictors of class membership by fixing classes established in the unconditional GMM, and modeling predictors and trajectories simultaneously. We built our logistic regression model by first assessing each identified predictor in a univariate model.[6] Results of univariate analyses indicated that the Depression, Panic, Social Conflict, Hostility, Sleep Functioning, Sexual Functioning, and Suicidal Ideation subscales were significant predictors of class

[6]These preliminary analyses are not included due to space restrictions, but are available on request from the first author.

Table 3
Subscale Values at Session 1 and 15 and Clinically Significant Change by Group

	High-symptom nonresponders					High-symptom rapid responders					Low-symptom nonresponders				
	Session 1		Session 15			Session 1		Session 15			Session 1		Session 15		
	N	Mean	*N*	Mean	Sig. change	*N*	Mean	*N*	Mean	Sig. change	*N*	Mean	*N*	Mean	Sig. change
DEPRS	41	3.85	41	3.14	0.0%	20	3.56	20	0.31	75.0%	86	1.29	86	1.10	18.6%
LIFEQ	13	2.75	20	2.42	n/a	9	2.87	11	0.50	n/a	22	1.98	34	1.58	n/a
MANIC	41	0.05	41	−0.19	4.9%	20	0.29	20	−0.36	15.0%	86	−0.11	86	−0.09	5.8%
PANIC	40	3.64	41	2.62	7.5%	20	2.34	20	0.94	15.0%	86	1.16	86	1.06	7.1%
PSYCS	41	2.60	41	2.12	2.4%	20	1.55	20	0.14	15.0%	86	0.99	85	0.68	2.4%
SCONF	13	1.83	20	1.73	n/a	9	1.18	11	−0.24	n/a	22	0.41	34	0.30	n/a
SEXFN	12	2.12	20	1.39	n/a	9	0.88	11	0.34	n/a	22	0.46	34	0.07	n/a
SLEEP	40	2.12	41	1.67	20.0%	20	1.50	20	0.24	45.0%	86	0.74	86	0.52	18.6%
SUICD	41	3.09	41	2.41	4.9%	20	1.93	20	0.02	25.0%	86	1.25	86	1.13	5.8%
VIOLN	41	1.01	41	1.17	9.8%	20	1.41	20	0.36	10.0%	86	0.84	86	0.77	4.7%
WORKF	13	0.39	19	0.43	n/a	8	0.92	11	−0.50	n/a	22	−0.20	34	0.05	n/a

Note. DEPRS = Depression; LIFEQ = Life Quality; MANIC = Mania Symptoms; PANIC = Anxiety; PSYCS = Psychotic Symptoms; SCONF = Social Conflict; SEXFN = Sexual Functioning; SLEEP = Sleep Functioning; SUICD = Suicidality; VIOLN = Violence (here referred to as Hostility); WORKF = Work Functioning. Sig Change = the percentage of the class meeting the criteria for Clinically Significant Change on the subscale by session 15. A short version of the TOP was used after intake, for part of the duration of the current study. Therefore, Substance Use does not appear on this table, and the Life Quality, Social Conflict, Sexual Functioning, and Work Functioning subscales are underrepresented, as these scales were not included in the short version.

Table 4
Prediction of Class Membership Based on Intake Characteristics ($N = 138$)[a]

	OR	t	Sig.	95% CI Lower	95% CI Upper
Low-symptom nonresponders					
Depression subscale	**0.409**	**−3.285**	**0.001**	**0.240**	**0.698**
Panic subscale	0.841	−0.856	0.392	0.566	1.250
Social Conflict subscale	1.252	1.348	0.178	0.903	1.737
Sexual Functioning subscale	1.042	0.179	0.858	0.663	1.639
Sleep Functioning subscale	0.765	−1.39	0.164	0.524	1.116
Suicide subscale	**1.351**	**2.311**	**0.021**	**1.047**	**1.744**
Hostility subscale	0.873	−1.616	0.106	0.741	1.029
High-symptom nonresponders					
Depression subscale	1.213	0.579	0.563	0.630	2.337
Panic subscale	0.908	−0.51	0.61	0.624	1.318
Social Conflict subscale	**1.610**	**2.194**	**0.028**	**1.052**	**2.465**
Sexual Functioning subscale	**1.495**	**1.744**	**0.081**	**0.951**	**2.349**
Sleep Functioning subscale	1.223	0.81	0.418	0.752	1.988
Suicide subscale	**1.347**	**2.065**	**0.039**	**1.015**	**1.788**
Hostility subscale	**0.751**	**−2.284**	**0.022**	**0.587**	**0.960**

Note. OR = odds ratio; CI = confidence interval; sig. = p value of the t statistic.
All predictors were recorded at intake, before the first session of treatment. Items with p values less than .10 are in boldface. All results were generated with the high-symptom rapid responder group as the reference class.
[a]Due to listwise deletion, nine clients were removed from the final analysis. Most clients were lost from the high-symptom nonresponders group ($N = 34$). Three were removed from the low-symptom nonresponders group ($N = 83$).

membership. Income level, presence of personality disorder, and the Life Quality, Mania, Work Function, and Psychotic subscales failed to significantly predict class membership.

Significant univatriate predictors were assessed concurrently via multivariate multinomial logistic regression, with class membership as the dependent variable. These results can be found in Table 4. We evaluated predictors for shared covariation and potential multicollinearity. While some predictors had significant correlations, none were higher than $r = 0.45$. In building a final model, we tested whether parameters varied significantly if one correlated variable was removed and another retained. The model results did not change meaningfully, and so we report the model with all significant *univariate* predictors. Some of the significant predictors were missing data for some clients and reduced the available sample size (due to listwise deletion). These were Panic ($n = 146$), Social Conflict ($n = 143$), Sexual Functioning ($n = 139$), Sleep Functioning ($n = 146$), Suicide ($n = 146$), and Hostility ($n = 145$). This resulted in a listwise deletion of 9 clients and a final sample size of $n = 138$. In all logistic regression analyses, the high-symptom rapid responder group was used as the reference class.

When assessed together in the multivariate model, the Depression, Social Conflict, Suicide, and Hostility subscales were significant predictors of class membership, and the Sexual Functioning subscale trended towards significance (see Table 4). Regarding differences between the two high-symptom groups: Increased hostility toward others increased the likelihood of being in the rapid responders group, while increased social conflict, poorer sexual functioning, and increased suicidality were associated with increased likelihood of belonging to the high-symptom nonresponder group. It is unsurprising that the low-symptom group was differentiated from the reference group by the Depression subscale at intake, as the intercepts of this scale at session 1 clearly distinguished the low-symptom group from the high-symptom groups.

In addition to the logistic regression analyses, we examined effect size differences between classes at intake by dividing the mean difference between classes by their pooled standard

Table 5
Effect Size Differences for High-Symptom Classes at Intake

	Nonresponders (*M*)	Rapid responders (*M*)	Pooled *SD*	Effect size
DEPRS	3.92	3.39	1.01	0.52
LIFEQ	3.07	2.66	0.85	0.47
MANIC	0.01	0.26	0.80	−0.31
PANIC	3.13	2.63	1.89	0.26
PSYCS	2.48	1.30	2.54	0.46
SCONF	2.52	1.15	1.59	0.86
SEXFN	1.38	0.56	1.47	0.56
SLEEP	2.55	1.81	1.61	0.46
SUICD	3.68	2.38	3.00	0.44
VIOLN	1.14	2.69	3.71	-0.42
WORKF	1.00	0.56	1.83	0.24

Note. M = mean; SD = standard deviation; DEPRS = Depression; LIFEQ = Life Quality; MANIC = Mania Symptoms; PANIC = Anxiety; PSYCS = Psychotic Symptoms; SCONF = Social Conflict; SEXFN = Sexual Functioning; SLEEP = Sleep Functioning; SUICD = Suicidality; VIOLN = Violence (here referred to as Hostility); WORKF = Work Functioning.

deviation. Effect sizes can be found in Table 5. Broadly, these support the logistic regression findings and indicate moderate-sized differences between the two high-symptom groups on most of the subscales, with the rapid responder group higher on the Mania and Hostility subscales and the nonresponders higher on all other subscales. While these differences may be meaningful, they should be taken as initial indications of variables worthy of future study.

Discussion

The present study replicated potentially important results from previous work: identifying two groups of clients who reported similar high levels of distress at the first session of psychotherapy, but followed markedly different recovery trajectories. As an extension of this earlier work, the present study also explored the relationship between pretreatment variables and membership in these two high-symptom groups.

Are There Multiple Groups of Responders?

At the latent group level, and with particular regard to the high-symptom groups, our results closely mirror those found by Stulz et al.'s (2007) examination of change in the first six sessions of psychotherapy. Both studies found two groups of high-symptom severity clients, one that improved rapidly and one that did not improve substantially. Although Stulz and colleagues found a high-symptom group with a nonsignificant slope, the current study found a high-symptom group with a significant slope, but one so flat that no member of that group met the full criteria for clinically significant change after 15 sessions in treatment, which is consistent with the results of Lutz et al.'s (2009) reexamination of the TDCRP data. The percentage of clients in our rapid responders group (14%) was similar to the group identified by Stulz and colleagues (2007).

Such similarity in findings across settings and measures increases our confidence that similar groups of rapid responders and nonresponders would be found in treatment as usual (TAU) samples. In addition to being consistent with earlier work using GMMs, these results appear to support findings in the field more broadly: (a) highly distressed clients respond poorly to treatment and (b) highly distressed clients respond more rapidly to treatment. The current results suggest that previous studies reporting contradictory findings with regard to the relationship between symptom severity and outcome may have included different proportions of the subgroups found herein. Assuming that both groups of treatment responders exist, the relative size of one versus the other would influence the results of any aggregate analysis. It may be useful to incorporate these findings into future methodology by assessing for noncontinuous heterogeneity in trajectories of change.

The RCI and cutoff results reinforce the usefulness of identifying and modeling rapid responders. These data indicate that the majority of the high-symptom rapid responders group (75%) demonstrated a level of improvement that met the full criteria for clinical significance, both crossing below the cutoff into the healthy range and exceeding requirements for the RCI. Of the high-symptom nonresponders, none achieved the same criteria. These results appear borne out by the other TOP subscales we were able to evaluate. Clients in the rapid responder group substantially exceeded their counterparts in changing on every subscale measured. Being able to identify and differentiate between these groups has potential clinical importance; for example, imagine knowing early in a treatment whether a particular client is likely to be slow to change over the next 15 sessions. This could allow for adjusted expectations about treatment course, useful to clinicians and administrators in planning caseload and upcoming availability and in setting expectations for change. Systems designed to predict client change and offer feedback are already in use or development, some of which attempt to differentiate between expected trajectories of change (see Barkham, Stiles, Lambert, & Mellor-Clark, 2010).

With regard to our second hypothesis, predicting that the high-symptom nonresponder group would be differentiated from the rapid responder group by higher functional impairment at intake, we found indications that reported higher social conflict, poorer sexual functioning, and increased suicidality predicted membership in the high-symptom nonresponders group, while, contrary to our expectations, higher hostility predicted membership in the high-symptom rapid responders group. With regard to our prediction that higher severity of symptoms at intake would predict membership in the high-symptom nonresponder group, there was no evidence that increased depression or anxiety at intake predicted membership in the nonresponder group relative to the rapid responder group. Last, contrary to our expectations, there was no evidence that the presence of Axis II disorder increased the likelihood of membership in the high-symptom nonresponders group.

One predictor of group membership—hostility—contradicted our hypothesis; higher values on this subscale at intake predicted increased odds of membership in the rapid responders group. It may be that distressed clients with high levels of hostility are particularly likely to benefit from psychotherapy (even when conducted by therapists in training) because this form of treatment generally allows clients to express or vent their negative emotions. Talking about and/or expressing anger toward others with a sympathetic therapist may provide these clients with a cathartic and/or corrective experience. Supporting this (albeit tentative) interpretation is a study by Pascual-Leone and Greenberg (2007), which showed that anger ("feelings of protest, repulsion, hate, disgust and so forth," p. 878) was associated with good in-session outcomes. These authors argued that anger, even in the form of hatred or hostility, can lead to more assertive, agentic states, which can, in turn, lead to better psychotherapy outcomes.

Broadly, however, our results indicate that clients who are highly impaired across multiple domains of functioning are less likely to respond to treatment. These results support the potential importance of assessing across multiple domains of distress and dysfunction, rather than aggregating across domains (see McAleavey, Nordberg, Kraus, & Castonguay, 2012). For example, in some measures (e.g., Administration and Scoring Manual for the Outcome Questionnaire; Lambert, Lunen, Umphress, Hansen, & Burlingame, 1994) that use a total score, a wide variety of symptoms and impairments contribute in the same manner to the overall reported score; however, the current results appear to indicate the importance of breaking out different symptom domains, as they predict different trajectories of change.

It is possible that increased suicidality is a proxy for more severe depression, and, as such, the suicide subscale indicates more severely depressed clients. However, this appears unlikely, as we directly tested symptoms of depression at intake as a predictor of group membership and found nonsignificant results. There is evidence suggesting that thoughts about death and suicide are not good indicators of major depressive disorder (Buchwald & Rudic-Davis, 1993). Increased suicidality may instead be important in and of itself. For example, it is possible that in the face of suicidal thoughts and feelings from clients, trainee therapists were intimidated and oriented their focus on the suicide, rather than on underlying causal and maintaining factors of the disorder, to which suicidality was a reaction. Thus, by ignoring underlying mechanisms in favor of a resultant condition, therapists made little progress in promoting change.

Beutler and colleagues (2006) point to a principle of change for the treatment of dysphoric disorders whereby clients with high levels of initial impairment benefit less from TAU, and more from well-structured, long-term treatments. We might add to this that our logistic regression results indicated impairment at session 1 should be examined in the context of impairment at an earlier baseline. If a client has lower functional impairment at intake, we might adopt a prediction of treatment response that does not call for a departure from TAU. Whereas a client with a high degree of pretreatment functional impairment might merit alternative treatment plans, as recovery may be slower, or nonexistent, over 15 sessions. At the very least, given the present data, an assessment of functional impairment early in treatment seems to meaningfully improve the prediction of outcome.

Is the differentiation between the high-symptom rapid responders and nonresponders spurious? The present results indicate that these groups are meaningful, but are also fuzzy at their boundaries. It would not be accurate to state that class membership perfectly describes every client in our sample, nor that the classes should be reified as constructs. However, there are clients who appear well described as high-symptom nonresponders and rapid responders. These do not appear to be artifacts of a statistical analysis, but replicated clinical phenomena with meaningful usefulness in facilitating clinical prediction of treatment course.

As discussed earlier, regression to the mean may have played a part in the trajectory of change for the high-symptom rapid responder group. Clients in this group were highly distressed at session 1; however, such distress may have been a brief deviation from their normal level of functioning. As such, some portion of their recovery might be attributable to a natural return to a less acute level of distress, rather than an effect of treatment. Nevertheless, the presence of the high-symptom nonresponder group supports the notion that regression to the mean, although a likely source of some change, cannot explain the trajectories of all, or even most, of the high-symptom clients in this study, unless we consider the possibility that different groups of clients have a different mean to regress toward. It is possible that clients with less functional impairment have a lower mean than their more impaired counterparts. They may, for example, have begun treatment during a spike in symptoms, while high-symptom nonresponders may have begun closer to a higher and more chronic level of severity.

It is also possible, in this case, that the high-symptom rapid responders have more treatment-facilitating factors in their lives than the high-symptom nonresponders. That these clients appear to report less social conflict and better sexual functioning, for example, may be an indication that they are generally more interpersonally connected than their nonresponding counterparts. Such clients may be better equipped to take advantage of treatment, and may have more support for change. One direction for future longitudinal research may be to establish a baseline of functioning before individuals become so distressed that they seek treatment; surely a daunting challenge, but one currently being pursued by researchers at Penn State with first-year college students.

Limitations

As in most studies that use data collected in naturalistic settings, this study offers high-external validity at the cost of relatively lower internal validity. Unmeasured and uncontrolled variables could have accounted for significant differences in our sample. Our ability to determine treatment orientation was hampered by the fluid nature of supervision in the training clinic and the lack of detailed attention regarding this variable in the clinical records. While students participate in practicum teams that operate under a specific orientation, many take on additional cases with independently assigned supervisors. In addition, practicum (and therefore orientation) changes with every new academic year. Records going back over a decade are poor or nonexistent, and thus it was impossible to reliably determine the orientation under which a particular therapist treated a particular client.

Similarly, it was not possible to determine the focus of any individual treatment. Although we used the Depression subscale from the TOP in our analyses, not all of the clients met DSM-IV diagnostic criteria for clinical depression, and not all of them may have been primarily treated for depression. As such, it is not appropriate to assume homogeneity of treatment across clients. In addition, while there was a strict protocol regarding the timing of intake and treatment sessions, we cannot confirm that all clients completed their intake paperwork exactly 4 weeks before beginning their first session of psychotherapy.

Moreover, there were significant missing data in this study, as a result of the fluid nature of sampling in a naturalistic setting. Clients, through choice or absent-mindedness, omitted some questions periodically, and the use of a shorter version of the TOP resulted in significant missing data for the majority of our clients on several subscales. It is possible that with more data on certain subscales, additional predictors could have been found in our exploratory analyses. However, we would note that the data for the primary measure, the Depression subscale, were fully intact as a requirement for this study.

We deliberately limited our sample to clients who had completed at least 15 sessions of psychotherapy. This eliminated an important group of clients, namely, those who drop out early in treatment. One limitation of growth mixture modeling is that it is difficult to perform analyses on groups with limited repeated-measures data, such as those who drop out of treatment. Accordingly, it is important to note that there is at least one additional group we did not model—those who did not engage with treatment for 15 sessions. Moreover, while our initial sample was acceptable, once clients had been divided into classes, the small size of these groups may have limited our ability to detect predictors of class membership.

Last, it is possible that the low-symptom group results were largely driven by floor effects. The TOP Depression subscale has a floor of −1.44 standard deviations. Put simply, this means that clients with lower initial symptom severity have less room for change. Therefore, a client who starts at one standard deviation above the normative, nonclinical mean and bottoms out at −1.44 standard deviations will have a limited impact on the slope of change (−2.44 total change), whereas a client who begins at four standard deviations above the population mean can have a much larger effect (−5.44 total change). It was largely for this reason that we focused on the two high-symptom groups in our discussion of findings, as these groups were equivalently affected by potential floor effects.

Future Directions

Despite these limitations, the current study offers a replication of previous findings, and supports the notion that clients may cluster into predictable groups of responders and nonresponders across settings and outcome measures. Such a finding supports the usefulness of research using latent classes to model change, and provides greater confidence in our ability to understand and predict change in psychotherapy. This study has also extended past findings by examining change across several domains, rather than a single total score.

Future research may benefit from the strengths of latent class models to test the predictive validity of other potentially important variables (participant, relationship, and technical), using the existing literature on variables predictive of outcome as a guide (e.g., Castonguay & Beutler, 2006). In addition, the current body of research using latent class models has almost entirely been conducted using global assessments of clients' well-being. Continuing to examine trajectories of change for different symptom groups (e.g., anxiety, psychosis, and substance abuse) should offer increased clarity into ways in which we can expect to see different clients with different symptoms change over time. Last, rather than treating heterogeneity in client response as noise or deviation, we would encourage researchers to seek out methods to model these differences, and then continue the work in predicting why some clients appear not to respond well to treatment, while others do.

References

American Psychiatric Association. (2000). Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. Washington, DC: Author.

Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change. Journal of Consulting and Clinical Psychology, 77, 203–211.

Barkham, M., Stiles, W. B., Lambert, M. J., & Mellor-Clark, J. (2010). Building a rigorous and relevant knowledge base for the psychological therapies. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), Developing and delivering practice-based evidence (pp. 21–61). West Sussex: John Wiley & Sons, Ltd.

Bauer, D. J., & Curran, P. J. (2003a). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. Psychological Methods, 8, 338–363

Bauer, D. J., & Curran, P. J. (2003b). Overextraction of latent trajectory classes: Much ado about nothing? Psychological Methods, 8, 384–393.

Bauer, D. J., & Reyes, H. L. M. (2010). Modeling variability in individual development: Differences of degree or kind? Child Development Perspectives, 4, 114–122

Beck, A. T., Steer, R. A., & Ranieri, W. F. (1988). Scale for suicide ideation: Psychometric properties of a self-report version. Journal of Clinical Psychology, 44, 499–505.

Beutler, L. E., Blatt, S. J., Alimohamed, S., Levy, K. N., & Angtuaco, L. (2006). Participant factors in treating dysphoric disorders. In L. G. Castonguay & L. E. Beutler (Eds.), Principles of therapeutic change that work (pp. 13–63). Oxford: Oxford University Press.

Bland, J. M., & Altman, D. G. (1994). Statistic notes: Regression toward the mean. British Medical Journal, 308, 1499.

Brazier, J. E., Harper, R., Jones, N. M., O'Cathain, A., Thomas, K. J., Usherwood, T., & Westlake, L. (1992). Validating the SF-36 health survey questionnaire: New outcome measure for primary care. British Medical Journal, 305, 160–164.

Boswell, F., Kraus, D. R., Nordberg, S., & Castonguay, L. G., (2009, June). The Treatment Outcome Package (TOP): An investigation of its validity. Poster presented at the meeting of the Society for Psychotherapy Research, Santiago, Chile.

Buchwald, A. M, & Rudic-Davis, D. (1993). The symptoms of major depression. Journal of Abnormal Psychology, 102, 197–205.

Castonguay, L. G. (2011). Psychotherapy, psychopathology, research and practice: Pathways of connections and integration. Psychotherapy Research, 21, 125–140.

Castonguay, L. G., & Beutler, L. E. (2006). Principles of therapeutic change: A task force on participants, relationships and techniques factors. Journal of Clinical Psychology, 62, 631–638.

Clarkin, J. F., & Levy, K. N. (2004). The influence of client variables on psychotherapy. In M. J. Lambert (Ed.), Bergin and Garfield's handbook of psychotherapy and behavior change (5^{th} ed., pp. 194–226). New York: John Wiley & Sons.

Cuijpers, P., van Lier, P. A. C., van Straten, A., & Donker, M. (2005). Examining differential effects of psychological treatment of depressive disorder: An application of trajectory analyses. Journal of Affective Disorders, 89, 137–146.

Davis, C. E. (1976). The effect of regression to the mean in epidemiologic and clinical studies. American Journal of Epidemiology, 104, 493–498.

Derogatis, L. R. (1975). Brief symptom inventory. Baltimore, MD: Clinical Psychometric Research.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., ... & Parloff, M. B. (1989). National Institute of Mental Health treatment of depression collaborative research program: General effectiveness of treatments. Archives of general psychiatry, 46(11), 971–982.

First, M. B., Spitzer, R. L, Gibbon M., & Williams, J. B.W. (2002). Structured Clinical Interview for DSM-IV-TR Axis I Disorders (research version, patient ed.) (SCID-I/P). New York: Biometrics Research, New York State Psychiatric Institute.

Fisher, A. J., Newman, M. G., & Molenaar, P. C. M. (2011). A quantitative method for the analysis of nomothetic relationships between idiographic structures: Dynamic patterns create attractor states for sustained posttreatment change. Journal of Consulting and Clinical Psychology, 79, 552–563.

Graham, J. R. (1993). MMPI-2: Assessing personality and psychopathology. New York: Oxford University Press.

Hayes, A. M., Laurenceau, J., & Cardaciotto, L. (2007). Methods for capturing the process of change. In A. M. Nezu & C. M. Nezu (Eds.), Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions (pp. 335–358). New York: Oxford University Press.

Hoberman, H. M., Lewinsohn, P. M., & Tilson, M. (1988). Group treatment of depression: Individual predictors of outcome. Journal of Consulting and Clinical Psychology, 56, 393–398.

Horowitz, L. M., Strupp, H. H., Lambert, M. J., & Elkin, I. (1997). Overview and summary of the Core Battery Conference. In H. H. Strupp, L. M. Horowitz, & M. J. Lambert (Eds.), Measuring patient changes in mood, anxiety, and personality disorders: Towards a core battery (pp. 11–54). Washington, DC: American Psychological Association.

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy. American Psychologist, 51, 1059–1064.

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.

Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. Psychological Bulletin, 65, 110–136.

Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. Journal of Clinical Psychology, 61, 285–314.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. Journal of Counseling and Clinical Psychology, 69, 159–172.

Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L., & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions. Journal of Clinical Psychology, 61, 165–174.

Lambert, M. J., Lunnen, K., Umphress, V., Hansen, N., & Burlingame, G. M. (1994). Administration and Scoring Manual for the Outcome Questionnaire (OQ-45.1). Salt Lake City: IHC Center for Behavioral Healthcare Efficacy.

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.194–226). New York: John Wiley & Sons.

Lambert, M. J., & Ogles, B. M. (2009). Using clinical significance in psychotherapy outcome research: The need for a common procedure and validity data. Psychotherapy Research, 19, 493–501.

Laurenceau, J. P., Hayes, A. M., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. Clinical Psychology Review, 27, 682–695.

Lutz, W., Lambert, M. J., Harmon, S. C., Tschitsaz, A., Schurch, E., & Stulz, N. (2006). The probability of treatment success, failure and duration—What can be learned from empirical data to support decision making in clinical practice? Clinical Psychology and Psychotherapy, 13, 223–232.

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(7), 889–900.

Lutz, W., Stulz, N., & Kock, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. Journal of affective disorders, 118, 60–68.

McAleavey, A. A., Nordberg, S. S., Kraus, D., & Castonguay, L. G. (2012). Errors in treatment outcome monitoring: Implications for real-world psychotherapy. Canadian Psychology, 53, 105–114.

McLellan, A. T., Alterman, A. I., Metzger, D. S., Grissom, G. R., Woody, G. E., Luborsky, L., & O'Brien, C. P. (1994). Similarity of outcome predictors across opiate, cocaine, and alcohol treatments: Role of treatment services. Journal of Consulting and Clinical Psychology, 62, 1141–1158.

Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. Journal of Personality, 60, 117–174.

Morton, V., & Torgerson, D. J. (2003). Effect of regression to the mean on decision making in health care. BMJ: British Medical Journal, 326(7398), 1083.

Muthén, B., & Muthén, L. K. (1998–2011). Mplus (Version 6.11) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Nagin, D. S., & Tremblay, R. E. (2005) Developmental trajectory groups: Fact or a useful statistical fiction? Criminology, 43, 873–904.

Newman, M. G., Crits-Christoph, P., Gibbons, M. B., & Erickson, T. M. (2006). Participant factors in treating anxiety disorders. In L. G. Castonguay, & L. E. Beutler (Eds.), Principles of therapeutic change that work (pp. 121–153). Oxford: Oxford University Press.

Nylund, K. L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. Structural Equation Modeling, 14, 535–569.

Pascual-Leone, A., & Greenberg, L. S. (2007). Emotional processing in experiential therapy: Why "the only way out is through." Journal of Consulting and Clinical Psychology, 75, 875–887.

Paul, G. L. (1967). Outcome research in psychotherapy. Journal of Consulting Psychology, 31, 109–118.

Stulz, N., & Lutz, W. (2007). Multidimensional patterns of change in outpatient psychotherapy: The phase model revisited. Journal of Clinical Psychology, 63, 817–833.

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.

Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. Journal of Consulting and Clinical Psychology, 73, 914–923.

Weisz, J. R., Hawley, K. M., Pilkonis, P. A., Woody, S. R., & Follette, W. C. (2000). Stressing the (other) three rs in the search for empirically supported treatments: Review procedures, research quality, relevance to practice and the public interest. Clinical Psychology: Science and Practice, 7, 243–258.