J Child Fam Stud (2009) 18:690–701
DOI 10.1007/s10826-009-9275-9

ORIGINAL PAPER

Identifying Youth at Risk for Treatment Failure in Outpatient Community Mental Health Services

Jared S. Warren · Philip L. Nelson · Gary M. Burlingame

Published online: 10 June 2009

Abstract We developed predicted change trajectories and a warning system designed to identify psychotherapy cases at risk for treatment failure as observed in archival Youth Outcome Questionnaire data (parent/guardian-report) from 363 children and adolescents (ages 4–17) served in an outpatient community mental health system. We used multilevel modeling procedures to develop models of predicted change based on demographic information. Controlling for the effects of age on intercept, no other variables were significant in the model. The warning system we created from half of the sample ($n = 181$) correctly identified 71% of treatment failures in the other half of the sample ($n = 182$), defined as cases whose symptoms were significantly higher at the end of treatment compared to symptoms at intake. As over half of youth cases in this usual care setting did not demonstrate reliable improvement in symptoms, these results further emphasize the value of patient-focused research in monitoring patient progress and prompting changes in the treatment approach if suitable progress is not observed.

Keywords Treatment failure · Change trajectories · Usual care · Patient-focused research · Child psychotherapy

J. S. Warren (✉) · G. M. Burlingame
Department of Psychology and Clinical Psychology, Brigham Young University, 291 John Taylor Building, Provo, UT 84602-8626, USA
e-mail: Jared_Warren@byu.edu

P. L. Nelson
Department of Counseling Psychology & Special Education, Brigham Young University, 340 MCKB, Provo, UT 84602, USA

Introduction

Evidence-based practice in psychology (EBPP) has been defined as the "integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (APA 2006, p. 273). Evidence-based practice includes the regular monitoring of patient outcomes such that treatment can be adjusted if suitable progress is not observed (APA 2006; Institute of Medicine 2006). Within this context, researchers have developed methods to enhance clinical decision-making and improve mental health outcomes in adults using a "patient-focused" research paradigm (Howard et al. 1996). Common threads in the patient-focused paradigm include regular and reliable monitoring of patient progress, providing feedback on progress to clinicians, and using rationally-or empirically-derived algorithms to identify patients who may be at risk for negative outcomes. With regard to these prospects in child and adolescent psychotherapy, Kazdin (2005) noted that "such information would be enormously helpful if used to monitor and evaluate treatment in clinical practice" (p. 555); however, very little research has evaluated the feasibility and utility of using a patient-focused paradigm for monitoring child and adolescent treatment progress and identifying cases that may be at risk for negative outcomes. Our purpose with this study was twofold: (1) to develop predicted change trajectories for children and adolescents based on archival outpatient data from a community mental health organization, and (2) to evaluate the accuracy of an empirically-derived system for identifying cases that may be at risk for treatment failure.

The patient-focused research paradigm is distinguished from, but complementary to, paradigms of treatment efficacy and effectiveness research. The focus of efficacy and effectiveness research is on the average *group* response to specific interventions; in patient-focused research, the focus is on monitoring an *individual's* progress over the course of treatment (Howard et al. 1996). Patient-focused research seeks to provide clinicians with valid methods for systematically evaluating individual patient response *during* the course of treatment. As such, the patient-focused paradigm asks the question "Is this treatment currently working for this particular individual?"

Howard et al. (1996) were among the first to document the use of a patient-focused approach in their efforts to identify adults who were making suitable progress in treatment versus those believed to be at risk for negative outcomes. Utilizing a model that included 18 pre-treatment patient variables (e.g., symptom severity, chronicity of problems, and attitude toward treatment), a predicted change trajectory was created for each patient. As outcome measures were administered periodically during treatment, actual change was compared to predicted change for each patient, allowing clinicians to judge whether the patient was progressing at a suitable pace or was at risk for a negative outcome. Subsequent variations and revisions have sought to improve the predictive accuracy or clinical utility of these procedures (Lueger et al. 2001; Lutz et al. 1999).

Similarly, Lambert and colleagues (e.g., Finch et al. 2001; Lambert et al. 2001) have developed a system for monitoring patient progress using the Outcome Questionnaire-45 (OQ-45; Lambert et al. 2004). Patient symptoms are measured on a session-by-session basis, and an early warning system notifies therapists as early as the second session if patients are judged to be at risk for a negative outcome. "Clinical support tools" have been developed in conjunction with this system to aid clinicians in examining and adjusting their approach to treatment, thus reducing the likelihood of a negative treatment outcome. The combination of early identification of at risk cases, feedback on patient progress to clinicians, and clinical support tools for adjusting treatment when necessary has resulted in improved outcomes and fewer numbers of patients who deteriorate (Harmon et al. 2007; Hawkins et al. 2004; Lambert et al. 2001, 2002b; Whipple et al. 2003).

With its focus on individual outcomes, patient-focused research offers new opportunities to study adverse effects in psychotherapy. Such study has received relatively little attention in the literature; however, this area has begun to receive increased interest in the contexts of managed care and evidence-based practice (Lilienfeld 2007). Psychotherapy research suggests that 5–10% of adult psychotherapy clients can be classified as experiencing *deterioration* or *treatment failure*—leaving treatment significantly worse off than when they entered (Lambert and Bergin 1994; Mohr 1995). Similar estimates of deterioration have been found for child and adolescent populations in managed care settings (Bishop et al. 2005; Bybee et al. 2007), and rates may be even higher for children and adolescents in traditional community mental health settings (Weisz et al. 1995). In a related vein, Lilienfeld (2007) asserted that greater emphasis should be placed on research identifying potentially harmful treatments than on identifying empiricallysupported therapies. He also cited work by Lambert and colleagues on routine patient outcome monitoring and providing feedback to clinicians as a potential antidote against potentially harmful treatments. Furthermore, increased attention to deterioration in treatment may be warranted given the high rates of treatment dropout observed in clinical practice. It is estimated that 40–60% of children and adolescents discontinue treatment prematurely (Kazdin 1996; Wierzbicki and Pekarik 1993); many of these dropouts are likely due to perceived lack of benefit from treatment.

The need for systematic methods for monitoring patient progress and identifying cases at risk for treatment failure is underscored by the fact that therapists are not adept at predicting such cases (Breslin et al. 1997; Grove and Meehl 1996). For example, Hannan et al. (2005) compared the accuracy of therapists' predictions of patient outcome (e.g., positive outcome, no reliable change, deterioration) to predictions based on empirically-derived recovery curves and algorithms developed from large archival databases of patient outcomes. Therapists ($N = 48$) were informed that the base rate of deterioration at their clinic had remained relatively stable at 8% over the preceding years; however, the therapists predicted that only 3 out of the 550 patients in the study (.01%) would end treatment with a negative outcome. Only one of those cases predicted by therapists to deteriorate actually finished treatment with a negative outcome, yet outcome data revealed that a total of 40 patients (7.3%) deteriorated. In contrast, the empirically-derived algorithms developed by the authors accurately identified—by the third session—86% of cases that ultimately ended treatment with a negative outcome. These results suggest that therapists tend to be optimistic about expected patient outcomes, that therapists have difficulty identifying patients that are likely to deteriorate in therapy, and that empirically-derived methods for early identification of deteriorating cases can be quite accurate.

Patient-focused research to prevent negative outcomes has been applied almost exclusively with adults. However, Kazdin (2005) has emphasized the potential value of patient-focused practices in child and adolescent psychotherapy. Two studies with child and adolescent samples suggest that promising results may also be expected with younger patients. Bishop et al. (2005) tested the accuracy of rationally-derived algorithms—those based on expert opinion and outcome measure characteristics—for identifying potential treatment failures in a sample of 300 residential and outpatient clients ages 3–18. Overall, this rationally-derived method was successful in identifying 77% of child/adolescent patients who had deteriorated by the end of treatment. However, prediction accuracy was significantly higher for residential than for outpatient clients. In addition, adult research suggests that empirically-derived algorithms for predicting treatment failure tend to be more accurate than those that use rationally-derived methods (Lambert et al. 2002a; Spielmans et al. 2006). Both approaches use outcome measures in the same way to identify individuals at risk for negative outcomes, but differ in the methods used for establishing criteria for identifying such individuals. More specifically, the rationally-derived algorithms used by Bishop et al. (2005) were established through consensus opinion of several experienced clinicians and researchers regarding the progress expected of most clients at a given point in therapy. Empirically-derived methods use actual data on average client symptom change in establishing the cutoffs for at risk clients.

Utilizing empirically-derived change trajectories based on multilevel modeling (MLM), Bybee et al. (2007) tested the accuracy of a similar warning system using a large archival database of children and adolescents served in a managed care setting. In this study, the warning system accurately identified 72% of youth patients who ultimately ended treatment with a negative outcome. However, a significant limitation of the study was that youth self-report and parent/guardian-report outcome measures were combined in the analyses. In addition, the limited data available did not allow for testing potentially important variables in the change trajectory models such as age, diagnosis, and other patient and treatment characteristics.

Although the Bishop et al. (2005) and Bybee et al. (2007) studies represent a significant step forward in applying patient-focused research to children and adolescents, progress lags far behind that observed in adult treatment settings. In addition to the need for empirically-derived change trajectories and algorithms that consider potentially important patient and treatment variables, the patient-focused research paradigm could be particularly useful if applied to public community mental systems in which millions of youth are served each year (National Advisory Mental Health Council 2001; Ringel and Sturm 2001). Such applications may help reduce high dropout rates and improve the modest outcomes often observed in "real-world" settings (Weisz et al. 1995). These efforts may also help bridge the oft-lamented gap between youth psychotherapy research and actual clinical practice by facilitating the use of evidence-based, patient-focused procedures that are both empirically supported and clinically practical.

In response to these issues, our purpose with the present study was to develop a system to aid clinicians in identifying cases that may benefit from modified treatment to avoid premature termination and/or treatment failure. In two phases, we examined scores on the Youth Outcome Questionnaire obtained from the archives of an outpatient community mental health system. In phase 1, we attempted to create a model that would predict scores over time and identify related predictor variables. In phase 2, we tested the accuracy of an early warning system for identifying cases at risk for treatment failure. In this process, we used half the selected sample to establish cutoff scores intended to signal at risk cases. We then used the second half of the sample to evaluate correspondence between the cutoffs' outcome predictions and the actual outcomes observed in the archive.

Method

Participants and Procedure

We analyzed data selected from the archives (years 1997–2008) of an outpatient public community mental health system located in the Intermountain West of the United States. This community system covers approximately 1.5 million lives, with clientele typically of average to low socio–economic status. The psychotherapy services provided in this setting included individual and family psychotherapy, psychosocial skill-building groups, and medication management visits. Although a broad range of therapeutic approaches were used, therapists reportedly employed family therapy and cognitive strategies more frequently than psychodynamic or behavioral techniques.

Outcome data were collected at this institution as part of routine services. Parents or guardians completed the Youth Outcome Questionnaire (Y-OQ; described below) at check-in when presenting their children for outpatient treatment, typically requiring less than 10 min to complete. At intake, parents or guardians completed a form requesting basic demographics, some of which were later used in this study (e.g., sex and date of birth). We selected our data sample from an original Y-OQ archive having complete data for 1,782 cases with at least one treatment session. These cases had missing values for less than 10% of the Y-OQ's 64 items. In instances of missing values, we substituted values calculated using item-specific regression models. We limited our sample to cases within the appropriate Y-OQ age range of 4–17, which was 99% of the archive. Selecting cases with at least three measurement occasions further reduced our sample to 22% of the original archive. Selecting cases not having extremely long treatment episodes (i.e., below the 90th percentile: 83 weeks or fewer) further reduced our sample to 20%. For each case, we selected only the first treatment episode meeting inclusion criteria, with episodes delimited by 90+ day breaks in treatment or by changes in treatment setting (e.g., outpatient to day treatment).

Table 1 presents descriptive statistics for our selected sample of 363 cases and their 115 therapists providing services. Of these cases, the mean age was 10.8 years old, 38% were female, 62% were male, 51.8% were receiving Medicaid, and 31.1% were minorities. Unfortunately, the data archive was limited in specifying each minority group, but the largest group was Hispanic. The median treatment length was 14 sessions (33.3 weeks), with a Y-OQ outcome measurement at every 3.8 sessions on average (median). Primary diagnoses for these cases are listed in Table 1.

According to t tests and χ^2 tests, our selected sample differed significantly from the original archive with a lower mean age (11.5 vs. 12.4), a higher baseline Y-OQ score (86.4 vs. 82.2), a lower percentage of cases with reported alcohol and drug usage (9% vs. 15%), a higher percentage of cases receiving medication treatment (72% vs. 53%), a lower percentage of cases on Medicaid (52% vs. 58%), and a higher percentage of severely emotionally disturbed cases (94% vs. 87%; SED status was rated by the clinician and defined as emotional and mental disturbance that severely limits the individual's development and welfare). The sample did not differ significantly from the archive in percentages of females, minorities, or cases with previous treatment.

In phase 1 of the study, we used the total sample of 363 cases to create a model that would predict scores over time and identify related predictor variables. In phase 2, we tested the accuracy of an early warning system for identifying cases at risk for treatment failure. In this process, we used half the selected sample ($n = 181$) to establish cutoff scores intended to signal at risk cases. We then used the second half of the sample ($n = 182$) to evaluate correspondence between the cutoffs' outcome predictions and the actual outcomes observed in the archive. We created these two subsamples by random assignment. Usage of two separate subsamples was an attempt to avoid inflated estimates that could result from predictions being created from and tested on a single sample. To control for any potential bias due to random assignment of the two samples, we repeated the random assignment and analysis 10 times and reported the mean results for our analyses of the warning system's prediction accuracy.

Measures

The Youth Outcome Questionnaire-2.01 (Y-OQ; Burlingame et al. 2001, 2004, 2005) is a parent- or guardian-completed questionnaire requiring 8–10 min for completion. In contrast to other commonly used child behavior questionnaires, the Y-OQ was specifically designed to be sensitive

Table 1 Descriptive statistics for selected sample

	M	SD	Mdn	Range		*n*	%
n Y-OQs per case	3.9	1.3	3.0	3–11	Female	138	38
Weeks between Y-OQs	9.5	4.7	8.7	.3–26.5	Previous treatment	122	34
Sessions between Y-OQs	4.5	3.2	3.8	.3–19.3	Hispanic	37	10.2
Treatment episode number	1.9	2.0	1.0	1–24	Minority (includes Hispanic)	113	31.1
Treatment episode length (weeks)	36.4	18.9	33.3	.9–80.1	Medicaid	188	51.8
Treatment episode length (sessions)	17.7	15.2	14.0	1–104	Alcohol and drug use	33	9.1
Age	10.8	3.5	10.4	4.2–17.8	Cases on medications	260	71.6
					SED	341	93.9

Primary diagnoses[a]	*n*	%	Therapists	*n*	%
Attention-deficit/hyperactivity disorders	98	27.0	Social workers	81	70.4
Mood disorders	74	20.4	Psychologists	12	10.4
Adjustment disorders	33	9.1	Licensed professional counselors	9	7.8
Posttraumatic stress disorder	30	8.3	Psychiatrists	4	3.5
Oppositional defiant disorder	28	7.7	Marriage and family therapists	2	1.8
Substance abuse/dependence	27	7.5	Other/unknown	7	6.1
Abuse/neglect of child	22	6.1			
Anxiety-related disorders	15	4.1			
Conduct disorders	11	3.0			
Autistic disorders	8	2.2			
Other/unknown	17	4.6			

[a] 88% of cases had multiple diagnoses

to changes in symptom levels over the course of treatment, as opposed to classifying or categorizing child psychopathology. Its 64 items use 5-point Likert-type scaling with scores ranging from 0 to 4 (e.g., "My child is more fearful than other children of the same age."). Higher scores indicate greater distress. Eight of these items are scored in reverse to tap "healthy" behaviors and are weighted differently, with scores ranging from 2 to −2 (e.g., "My child cooperates with rules and expectations"). Different weights for adaptive behavior items are used because for this measure of psychosocial distress, endorsement of behavioral dysfunction is given slightly more emphasis than the absence of adaptive behavior.

The measure uses summative scoring and total scores may range from −16 to 240. Scores higher than the established clinical cut score of 46 are considered in the clinical range for level of distress (Burlingame et al. 2005). Although the current study used only Y-OQ total scores, the Y-OQ's items also form six subscales corresponding to behavioral domains useful in identifying youth with behavioral problems: (a) Intrapersonal Distress, (b) Somatic, (c) Interpersonal Relations, (d) Critical Items, (e) Social Problems, and (f) Behavioral Dysfunction.

The Y-OQ has a four-week test–retest reliability of .83 and an internal consistency reliability of .97. The concurrent validity of the Y-OQ with the Child Behavior Checklist (CBCL; Achenbach 1991) and the Conners' Parent Rating Scale (CPRS; Conners et al. 1998) ranges from the .80s to the low .90s. The Y-OQ is effective at distinguishing between clinical and non-clinical samples and it has been widely accepted for tracking treatment outcome and assessing psychosocial distress (Burlingame et al. 2004).

Analysis

Phase 1: Change Trajectory Model

We used MLM to create a model of Y-OQ scores over time and to identify any predictor variables for these change trajectories (LMER procedure, R software, version 2.7.2; Singer and Willett 2003). MLM is a form of regression that can be used to predict a subject's score at any particular time (dependent variable) using a number of independent variables, including a *time variable* (e.g., weeks in treatment). MLM estimates the starting point (i.e., intercept) and rate of change during treatment (i.e., slope) for each participant. Additionally, we estimated random effects that allow us to estimate the extent to which the intercepts and slopes varied across participants and therapists. Given that some participants received services from different therapists on different occasions, the LMER procedure of *R* software calculated these random effects while permitting cross-nesting of cases within therapists.

We used weeks in treatment as the basis for our time variable because of precedents in the child treatment literature failing to demonstrate a significant dose-response relationship for sessions attended and treatment outcome (Andrade et al. 2000; Bickman et al. 2002; Salzer et al. 1999). We theorized a curvilinear trajectory in which clients' rate of symptom level reduction (i.e., slope) is most rapid initially and tapers off over time. Similar to precedents in the literature (e.g., Finch et al. 2001; Lambert et al. 2002a; Spielmans et al. 2006), we modeled this trajectory shape using a logarithmic transformation of weeks in treatment (i.e., LNWEEKS = $\log_e$[weeks + 1]). Compared to other transformations, including polynomial functions and no transformation at all, this transformation also yielded superior model fit to our data (using indices such as the Deviance statistic and the Bayesian Information Criterion; for information regarding variable transformation, see Singer and Willett 2003, sections 6.2–6.3).

Our hypothesized model (Model A) predicted Y-OQ total scores using the log of weeks as a main effect. The model also included the following predictor variables we hypothesized as likely associated with the change trajectory: prior treatment recorded in data archive (1 = yes, 0 = no), total dose of treatment recorded in data archive (i.e., total number of sessions; Baldwin et al. in press), age (continuous variable calculated at the time of each measurement; e.g., session 1 age = 12.32 years, session 4 age = 12.46 years), and sex (1 = female, 0 = male). We did not test a diagnosis variable in the model because of potential diagnostic inaccuracies that likely would have limited its usefulness (Jensen and Weisz 2002) and because other research has indicated that diagnosis contributes little to predicting speed of recovery once initial symptom level is taken into account (Brown et al. 2005).

The model evaluated main effects for our hypothesized variables in order to assess their association with trajectory intercept. The model also evaluated these variables in interaction with the log of weeks in order to assess their association with trajectory slope. To facilitate interpretation and reduce multicollinearity, we centered all covariates around their grand means (e.g.,age − $\overline{\text{age}}$). We used stepwise deletion of predictor variables from this hypothesized model to create a *final* change trajectory model omitting any non-significant parameters (Model B; confirmed by stepwise addition).

Phase 2: Warning System

We created the warning system to predict which cases would experience negative outcome and be part of the deterioration outcome class. We determined the deterioration class and other outcome classes by calculating overall change scores for each client (i.e., difference between first and last Y-OQ scores), then comparing these change scores with the Y-OQ's reliable change index of 13 (RCI; Jacobson and Truax 1991). The RCI is an index of the minimum change in scores that is still distinguishable from measurement error.

The outcome classes were *deterioration* if the final score was at least 13 points worse than baseline, *no reliable change* if the final score differed from baseline by less than 13 points, *improvement* if the final score was at least 13 points better than baseline, or *recovery* if meeting criteria for improvement and the final score was in the subclinical range (i.e., less than 46). Cases whose scores worsened by 13 points or more and remained subclinical at treatment termination fell in a subclinical form of the deterioration outcome class. As described below, deterioration rates—the percentages of cases deteriorating—played a role in creating the prediction intervals and cutoffs that the warning system used to identify cases at risk for negative outcome.

The warning system we tested in this study used cutoff scores at each measurement occasion to identify at risk cases (Bybee et al. 2007; Cannon et al. 2009; Finch et al. 2001). To understand the concept of these cutoffs, imagine a sample consisting of cases with similar baseline scores. Given a hypothetical deterioration rate of 10% for the sample, final scores above the 90th percentile (i.e., final scores in the most extreme 10%) would belong to cases in the deterioration outcome class. Consider the rationale that the percentile rank of each case's final score would likely be similar to the percentile rank of any earlier score from each case. If the rationale holds, cutoffs set at the 90th percentile of scores at each session could identify cases heading for a final outcome of deterioration. Cases whose scores exceed such warning system cutoffs at any session would be in the most extreme 10% and would be more likely than other cases to be in the 10% of the sample that comprises the deterioration outcome class.

We created such warning system cutoffs using the reference sample ($n = 181$, subsample 1), then tested how accurately the cutoffs predicted deterioration in the validation sample (subsample 2). We used two steps to create these cutoffs from the reference sample. First, we created a multilevel model (Model C) of predicted Y-OQ total scores over time using only main effects for the log of weeks and initial score (the latter centered around its mean). We explain why we used only these two main effects after describing the second step in creating the warning system cutoffs.

In our second step for creating cutoffs, we created prediction intervals (i.e., t type confidence intervals) around these predicted scores. We set the confidence level of each prediction interval to correspond to the deterioration rate calculated for the overall sample. For example, had the deterioration rate been 10%, we would have used an 80% confidence level—the interval encompasses 80% of scores at any point in treatment—which would distinguish the highest and lowest 10% of cases above and below the interval, respectively. Thus the upper boundary of the interval provides the warning system cutoffs that identify cases exhibiting the most extreme symptomatology and who are likely at risk for deterioration. We did not include cases from the subclinical deterioration outcome class in our calculations of the deterioration rates that helped us determine these cutoffs. Ultimately, these interval boundaries or cutoffs for deterioration could be displayed in a single reference chart, enabling clinicians to identify predicted final outcome given their client's current score and number of weeks in treatment.

Our purpose in including only main effects for log of weeks and baseline score in the model for predicted scores was to ensure that the prediction intervals—and warning system cutoffs—created around these predicted scores would not vary by the values of any variable other than cases' baseline scores. This ensured that cutoffs adjusted up and down according to cases' baseline scores, but still corresponded to the deterioration rate from the overall sample, the only rate we could calculate with reliability without a larger sample. Unfortunately, we did not have a sufficiently large data set to calculate deterioration rates for various demographic subsamples and create associated cutoffs by including related predictors in the model.

With warning system cutoffs created using the reference sample, we next calculated the accuracy of the warning system's cutoffs in predicting outcomes in the validation sample. We based these calculations on the comparison of predicted outcomes with observed outcomes. Scores from the validation sample that exceeded the cutoffs on any measurement occasion other than the first or last signaled cases as predicted to have final outcomes of deterioration. We did not use first or last measurements to predict deterioration in the interest of methodological rigor, because those same measurements produced the criterion for actual deterioration (deterioration = final score 13+ points worse than baseline score). We identified the number of true positives (TPs; i.e., deterioration prediction was accurate), false positives (FPs), true negatives (TNs), and false negatives (FNs), ultimately calculating indices such as the sensitivity (percentage of actual deteriorators correctly predicted) and specificity (percentage of actual non-deteriorators correctly predicted).

Results

Phase 1: Change Trajectory Model

In phase 1 of this study we used MLM to create a model of Y-OQ scores over time and to identify any predictor variables for these change trajectories. Not all of our predictor variables were significant in our hypothesized model

Table 2 Change trajectory models

	Model A (with all covariates)		Model B (with significant covariates only)		Model C (For warning system prediction interval)	
	Estimate	SE	Estimate	SE	Estimate	SE
Fixed effects						
Intercept						
Intercept[a]	85.642*	2.038	85.762*	2.039	86.203*	.990
Prior treatment	9.356*	4.360				
Total sessions	.152	.134				
Age	−1.414*	.567	−1.105*	.490		
Female	−2.304	4.145				
Baseline					.867*	.022
Slope (interaction with *LNWEEKS*)						
Intercept[a]	−2.737*	.518	−2.751*	.509	−2.938*	.587
Prior treatment	−2.183	1.122				
Total sessions	−.009	.031				
Age	.130	.147				
Female	−1.192	1.050				
Random effects	Estimate	SD	Estimate	SD	Estimate	SD
Intercept	924.82*	30.41	940.57*	30.67	<.00	<.00
Slope (*LNWEEKS*) correlation	17.70*	4.21	17.49*	4.18	71.03*	8.43
Intercept × *LNWEEKS*	−.10		−.12		.00	
Residual	538.82*	23.21	540.12*	23.24	381.51*	19.53
Goodness of fit	Estimate		Estimate		Estimate	
Deviance	13,701		13,713		13,053	
Akaike information criterion	13,722		13,723		13,071	
Bayesian information criterion	13,796		13,760		13,108	

[a] Estimates for the Intercept parameter reflect the mean intercept and slope overall because all variables are centered around their grand mean. Estimates for all other parameters are merely deviations from the intercept constant

* $p < .05$

(see Table 2, Model A). We used stepwise deletion of non-significant parameters to arrive at our final model, shown in Table 2 as Model B. We also confirmed this model using a stepwise addition approach. The estimates for this model indicate that the average trajectory intercept was 85.8. The average rate of change was an improvement of 2.8 points for every unit increase in the log of weeks. This represents the improvement in scores after the first 1.7 weeks in treatment (where $LNWEEKS = 1$, weeks = 1.7), given the log transformation equation $LNWEEKS = \log_e(\text{weeks} + 1)$. Note that improvements of this size require increasingly longer periods of time as treatment progresses (e.g., where $LNWEEKS = 2$, weeks = 6.4, where $LNWEEKS = 3$, weeks = 19.1), as is expected with the curvilinear trajectory.

The fixed effects for intercept and slope in Model B (see Table 2) exhibited a correlation of −.506, suggesting that higher intercepts (i.e., more severe initial symptom levels) were associated with steeper slopes (i.e., faster rates of improvement). The only additional predictor that was significant in this model was the main effect for age. For every year that clients were older than the mean age, their trajectory intercept was an average of 1.1 points lower. The predictor variable for prior treatment, as a main effect and in interaction with LNWEEKS, was on the border between significance and non-significance in both Model A and B. The main effect was only significant when the interaction was also included, yet had we included the interaction in Model B, it would have had a p value of .0505. In addition, inclusion of these two extra parameters would only have lowered the Deviance statistic by 6.4 points. This difference of 6.4 points can be tested on a χ^2 distribution at 2 degrees of freedom (equal to number of parameters differing between the nested models), yielding a p value of .0408. Although we opted for parsimony by omitting the predictor for prior treatment from Model B, future studies may do well to examine it further.

There is still variability that remains unexplained by Model B, as indicated by the random effects estimates that remain statistically significant. The Intercept and Slope estimates indicate the between-persons variability in intercept and slope. The Residual estimate indicates the within-person variability. The random effects estimates for variability between therapists were not statistically significant in any model in Table 2, indicating that the variability attributable to therapists was not significantly different from zero. Thus we omitted random effects for therapists from all models in the table. This non-significance may be due, at least in part, to the cross-nesting of cases within therapists. Regarding the goodness of fit estimates listed in Table 2, values closer to zero indicate better fit. Singer and Willett (2003) offer further information about how such estimates play into model estimation.

Phase 2: Warning System

In phase 2 of this study we evaluated the accuracy of a warning system's cutoffs in identifying cases at risk for deterioration. We first identified RCI-based outcome classes of 21.2% deterioration, 30.0% no reliable change, 30.0% improvement, 17.7% recovery, and 1.1% subclinical deterioration. We next used the reference sample to calculate predicted scores over time using MLM. Model C of Table 2 presents estimates for this model. We then created a prediction interval around these predicted scores, the interval having a 57.6% confidence level such that the interval's upper boundary would identify a percentage of cases equal to the deterioration rate of 21.2%. This boundary then served as the warning system's cutoffs for identifying cases at risk for deterioration. Figure 1 offers a visual representation of the average predicted scores and the associated cutoffs for an example case having the mean baseline score of 86. This information could also be displayed in a table for clinicians to reference. The cutoffs increase over time, which appears to be a statistical artifact of increasing variability in scores as treatment progresses.

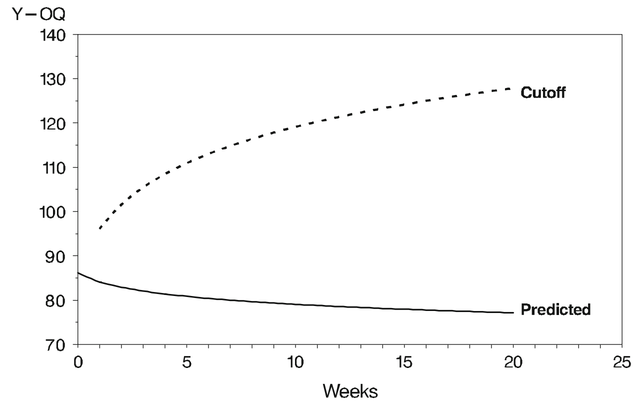

Fig. 1 Predicted scores and cutoffs for an individual with the mean baseline score of 86

Having created the warning system cutoffs from the reference sample, we next evaluated their accuracy in identifying deteriorators in the validation sample. Table 3 presents the warning system's deterioration predictions in comparison with the actual or observed outcomes. The system correctly identified 71% of the actual deteriorators (sensitivity). The system correctly identified 76% of the non-deteriorators (specificity). The system was correct 75% of the time in its overall classifications of deterioration/non-deterioration (hit rate). Cases signaled for deterioration by the system were 3.02 times more likely to end in deterioration than not (likelihood ratio). Of the cases that the system incorrectly predicted to deteriorate, 48% ended in the no reliable change outcome class.

Table 3 Warning system accuracy in predicting deterioration

		Predicted	Actual			
			Deterioration		Non-deterioration	
Sensitivity	.71	Deterioration	TP		FP	
Specificity	.76		28	15%	34	19%
Hit rate	.75	Non-deterioration	FN		TN	
Likelihood ratio	3.02		12	7%	108	59%
FP non-improvers	48%					

TP true positives, *FP* false positives, *TN* true negatives, *FN* false negatives, *FP non-improvers* percentages of false positives that showed no reliable change

Discussion

In phase 1 of this study we created a model for predicted Y-OQ scores over time. Age was the only significant predictor variable, with older clients exhibiting slightly lower trajectory intercepts. Prior treatment was nearly significant as a predictor variable, suggesting that future studies may find it to be associated with higher intercepts and steeper rates of change. In phase 2 of this study we developed a reasonably accurate warning system for identifying youth psychotherapy patients at risk for treatment failure. We developed the warning system using empirically-derived change trajectories and prediction algorithms based on a patient's deviation from expected progress at a given treatment session. The 71% sensitivity in identifying eventual treatment failures is considerably higher than estimates of therapists' accuracy in predicting such cases (e.g., 2.5% in a study by Hannan et al. 2005), and emphasizes the potential value of using this type of warning system to enhance clinical decision-making (Grove and Meehl 1996).

The overall hit rate of the warning system in this study (i.e., 75% accuracy in overall classifications of deterioration/non-deterioration) was nearly as high as rates in similar adult and youth studies. For example, in their study of adults, Lambert et al. (2002a) reported hit rates of 79 and 83% for rationally-derived and empirically-derived approaches, respectively. In child and adolescent populations, Bishop et al. (2005) reported an overall hit rate of 81% using a rationally-derived approach, and Bybee et al. (2007) reported a hit rate of 88% using empirically-derived methods. This study also appears consistent with previous child/adolescent studies in its sensitivity for accurately identifying treatment failures (71% in the present study compared to 77 and 72% in the Bishop et al. and Bybee et al. studies).

The current study may be conservative in its report of the warning system's prediction accuracy. Given that we determined actual deterioration/non-deterioration by comparing scores from the first and last measurements, we calculated the system's accuracy in the validation sample using alert signals produced on measurement occasions other than the first or last. Our purpose was to avoid using the same measurements to produce both the criterion and the prediction. However, clinicians using the warning system would often be unaware of which measurement occasions would be the last, and could also benefit from signal alerts occurring on the final measurement occasions. Used in such a manner, and given that some cases would produce their first signal alert on their final measurement occasions, the system would generally demonstrate a higher accuracy than reported in this study.

Although the warning system demonstrated an acceptable level of sensitivity, it is helpful to examine the cases whose outcomes the system predicted incorrectly. Of the system's predictions in the validation sample, 7% were false negatives—patients predicted not to deteriorate but who did (29% of deteriorating cases). It is regrettable that the warning system would fail to identify any case at risk for treatment failure and hopefully continued research in this area will improve on the system we tested in this study. The system's other incorrect predictions were the false positives comprising 19% of the validation sample—patients predicted to deteriorate but who did not. In the field of medicine, false positives from an analogous warning system could be potentially costly and dangerous to the patient (e.g., prompting unnecessary and invasive medical tests or treatments). Fortunately, such risks are less likely in psychotherapy. By definition, patients identified by the warning system are not making expected progress—relative to other patients—given their initial symptom level and their current stage in treatment. In practice, we expect that alerting clinicians to these cases will almost always be in the patient's best interests. In the present study, of those cases that were inaccurately predicted to end in treatment failure, 48% ended treatment with no reliable change. In other words, cases flagged by this warning system are very likely to be in need of some change in the approach to treatment if a positive outcome is to be achieved.

A number of other observations should be made about our findings. First, it is noteworthy that age was the only significant predictor variable in the change trajectory model. Significant results may have been observed for other variables with a larger sample; however, the overall impact of such variables on rate of change could be relatively small. As it stands, the change trajectory model developed in the present study provides a reasonably accurate, parsimonious, and practical foundation for evaluating ongoing progress in child/adolescent community mental health settings.

Another unexpected and sobering finding was that over half of the children and adolescents in this public community mental health sample did not achieve a positive outcome in therapy. In the total sample, based on parent/guardian-report, 21% had significantly higher symptoms at the end of treatment than when they began, and an additional 30% did not achieve any reliable change in symptom levels. Although discouraging, these findings appear consistent with most reviews and meta-analyses of traditional child psychotherapy outcomes in usual care settings which report little to no effect of treatment compared to controls (Bickman 1996; Weiss et al. 1999; Weisz 2004; Weisz et al. 1995). As we conducted this study using a patient-focused research framework, our purpose was not to evaluate the overall effectiveness of the community mental health system serving these youth. However, the observed 21% deterioration rate among patients in the total sample underscores the need for a valid system to help clinicians identify youth at risk for negative outcomes in usual care settings.

Some limitations of the available data and the treatment setting warrant discussion. A limitation to the study's generalizability was the lack of information about specific race categories for the sample's 31% minorities. Another noteworthy limitation may have been the relative infrequency with which the outcome measure was administered: every 3.8 sessions, on average (median). Session-by-session Y-OQ administration would have increased modeling accuracy and, possibly, warning system sensitivity (by increasing the number of potential signal alerts). Although previous child/adolescent studies in this area did not provide detailed information on the frequency of outcome measure administration, available information suggests that the slightly higher prediction accuracy in those studies could be attributable to more frequent outcome measurement (Bishop et al. 2005; Bybee et al. 2007). The infrequent measurement imposed perhaps the greatest limitation on the size of our selected sample. Whereas our selected sample included only 20% of the archive, it would have included 61% of the archive had the Y-OQ been administered at every treatment session. Results from a larger sample size such as this would have been more reliable in general and would have been more reflective of the archive's overall population. The Participants and Procedures section above describes demographic differences between our selected sample and the archive. However, the frequency of outcome assessment in this organization appears to be higher than what is typically observed in regular clinical practice, and demonstrates that such a system for tracking outcomes can be successfully employed and maintained in a large community mental health setting.

The use of a single parent-report measure for assessing outcome was also a possible limitation of the study. In a separate study, our research group is currently examining possible differences in deterioration rates, change trajectories, and warning system accuracy for parent versus adolescent self-report of outcome to evaluate the circumstances under which adolescent self-report of symptoms may be more appropriate for the warning system. In addition, the inclusion of supplemental outcome measures in other domains (e.g., consumer satisfaction, youth self-efficacy, parent stress) could have yielded a more complete picture of the impact of treatment. However, it is unknown whether the inclusion of such measures would significantly improve the accuracy of the warning system. In addition, the simplicity of using a single measure may maximize the interpretability and sustainability of the warning system approach, particularly in larger community mental health systems where these efforts may yield the greatest benefits.

A caveat for interpretation is required given the split-sample approach we used in phase 2 of the study. We created warning system cutoffs using subsample 1 and then tested the cutoffs' prediction accuracy in subsample 2. Coming from the same archive, these two subsamples exhibited inevitable similarities. If applied to a sample from a different institution, the warning system cutoffs from this study could yield rather different prediction accuracies. Where possible, an ideal application of the system would be for institutions to use their own archives to identify deterioration rates and create predictive cutoff scores specific to their institutions.

This study provides a foundation for a number of clinical practice applications and highlights several areas for future research, many of which have been raised in discussing adult applications of the patient-focused paradigm. Consistent with guidelines on evidence-based practice (APA 2006), predicted change trajectories and early warning systems can be used in child and adolescent psychotherapy to monitor outcomes and alert therapists to cases that may require a change in the treatment approach. Lambert and colleagues have developed an outcome monitoring system that provides immediate feedback to clinicians on patient progress, and the benefits of this system have been well-documented in adult studies (e.g., Lambert et al. 2001, 2002b). Research to date has not evaluated the impact of providing feedback on patient progress to clinicians (and/or parents) in child and adolescent psychotherapy.

The benefits of providing feedback have been enhanced in adult studies through the use of "clinical support tools"—problem-solving strategies and resources provided to clinicians to help them attend to certain factors known to be related to positive treatment outcomes (Harmon et al. 2007; Whipple et al. 2003). In adult treatment settings in which this approach is used, clinicians are alerted when a patient is "not on track" (i.e., identified as being at risk for treatment failure), and the clinician is provided with a decision tree designed to assess several outcome-related factors such as the patient's readiness for change, social support network, and the therapeutic relationship. A brief measure of these factors is completed by the patient, and the clinician can use this information to adjust the treatment approach as necessary. Similar procedures have not yet been developed for children and adolescents, but they could be particularly valuable if linked to putative mediators of treatment outcome and empirically supported interventions. For example, using the warning system described in this study, an alert could prompt additional assessment of the patient in areas believed to be related to treatment outcome in children and youth such as the therapeutic alliance, parent and youth motivation for treatment, the youth and family social support network, or recent stressful life events. Based on this information, the clinician could modify the treatment approach to address problems or deficits in those areas. In addition, alerts could prompt clinicians and supervisors to examine more closely whether empirically supported interventions for the client's concerns have been appropriately considered and utilized. The adult clinical support tools described above were developed after patient-focused warning systems were found to be accurate and feasible used in adult treatment settings; the results of the current study lay the foundation for the development of similar clinical support tools for child and adolescent cases.

Finally, future research is needed to address a number of issues related to the development and accuracy of child/adolescent change trajectories and the warning system described in this study. For example, results from the Bishop et al. (2005) study suggest that the accuracy of a warning system may vary as a function of the type of treatment setting (e.g., outpatient, residential, inpatient, etc.). Change trajectories and warning system accuracy may also differ based on respondent (e.g., youth self-report vs. parent/guardian-report of outcome). In addition, important differences in client population, services provided, and deterioration rates appear to exist between public community mental health systems and private managed care systems (Bishop et al. 2005; Bybee et al. 2007). As such, research is needed to examine potential differences in change trajectories and warning system accuracy across treatment settings, reporters of outcome, and systems of care. Future research could also explore alternative means to creating warning system cutoffs, experimenting perhaps with flat or descending cutoffs, in contrast to the current study's ascending cutoffs created using prediction intervals.

References

Achenbach, T. M. (1991). *Manual for the child behavioral checklist/4–18 and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Andrade, A. R., Lambert, E. W., & Bickman, L. (2000). Dose effect in child psychotherapy: Outcomes associated with negligible treatment. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 161–168. doi:10.1097/00004583-200002000-00014.

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *The American Psychologist, 61*, 271–285. doi:10.1037/0003-066X.61.4.271.

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

Bickman, L. (1996). A continuum of care: More is not always better. *The American Psychologist, 51*, 689–701. doi:10.1037/0003-066X.51.7.689.

Bickman, L., Andrade, A. R., & Lambert, E. W. (2002). Dose response in child and adolescent mental health services. *Mental Health Services Research, 4*, 57–70. doi:10.1023/A:1015210332175.

Bishop, M. J., Bybee, T. S., Lambert, M. J., Burlingame, G. M., Wells, M. G., & Poppleton, L. E. (2005). Accuracy of a rationally derived method for identifying treatment failure in children and adolescents. *Journal of Child and Family Studies, 14*, 207–222. doi:10.1007/s10826-005-5049-1.

Breslin, F., Sobell, L. C., Buchan, G., & Cunningham, J. (1997). Toward a stepped-care approach to treating problem drinkers: The predictive validity of within-treatment variables and therapist prognostic ratings. *Addiction (Abingdon, England), 92*, 1479–1489. doi:10.1111/j.1360-0443.1997.tb02869.x.

Brown, G. S., Lambert, M. J., Jones, E. R., & Minami, T. (2005). Identifying highly effective psychotherapists in a managed care environment. *The American Journal of Managed Care, 11*, 513–520.

Burlingame, G. M., Cox, J. C., Wells, M. G., Lambert, M. J., Latkowski, M., & Ferre, R. (2005). *The administration and scoring manual of the Youth Outcome Questionnaire*. Salt Lake City, UT: American Professional Credentialing Services.

Burlingame, G. M., Mosier, J. I., Wells, M. G., Atkin, Q. G., Lambert, M. J., Whoolery, M., et al. (2001). Tracking the influence of mental health treatment: The development of the Youth Outcome Questionnaire. *Clinical Psychology & Psychotherapy, 8*, 361–379. doi:10.1002/cpp.315.

Burlingame, G. M., Wells, A., Lambert, M. J., & Cox, J. (2004). Youth Outcome Questionnaire: Updated psychometric properties. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (3rd ed., Vol. 4, pp. 235–274). Mahwah, NJ: Lawrence Erlbaum Associates.

Bybee, T. S., Lambert, M. J., & Eggett, D. (2007). Curves of expected recovery and their predictive validity for identifying treatment failure. *Dutch Journal of Psychotherapy, 33*, 419–434.

Cannon, J. A. N., Warren, J. S., Nelson, P. L., & Burlingame, G. M. (2009). *Identifying youth at risk for psychotherapy treatment failure: Expected change trajectories for the Youth Outcome Questionnaire self-report.* Manuscript submitted for publication.

Conners, C. K., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998). The revised Conners' parent rating scale (CPRS–R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26*, 257–268. doi:10.1023/A:1022602400621.

Finch, A. E., Lambert, M. J., & Schaalje, B. G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology & Psychotherapy, 8*, 231–242. doi:10.1002/cpp.286.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2*, 293–323. doi:10.1037/1076-8971.2.2.293.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., et al. (2005). A lab test and algorithms for identifying cases at risk for treatment failure. *Journal of Clinical Psychology, 61*, 155–163. doi:10.1002/jclp.20108.

Harmon, S. C., Lambert, M. J., Smart, D. M., Hawkins, E., Nielsen, S. L., Slade, K., et al. (2007). Enhancing outcome for potential treatment failures: Therapist—client feedback and clinical support tools. *Psychotherapy Research, 17*, 379–392. doi:10.1080/10503300600702331.

Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K., & Tuttle, K. (2004). The therapeutic effects of providing client progress information to patients and therapists. *Psychotherapy Research, 10*, 308–327. doi:10.1093/ptr/kph027.

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *The American Psychologist, 51*, 1059–1064. doi:10.1037/0003-066X.51.10.1059.

Institute of Medicine. (2006). *Improving the quality of health care for mental and substance-use conditions: Quality chasm series*. Washington, DC: National Academies Press.

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. doi:10.1037/0022-006X.59.1.12.

Jensen, A. L., & Weisz, J. R. (2002). Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology, 70*, 158–168. doi:10.1037/0022-006X.70.1.158.

Kazdin, A. E. (1996). Dropping out of child therapy: Issues for research and implications for practice. *Clinical Child Psychology and Psychiatry, 1*, 133–156. doi:10.1177/1359104596011012.

Kazdin, A. E. (2005). Evidence-based assessment for children and adolescents: Issues in measurement development and clinical application. *Journal of Clinical Child and Adolescent Psychology, 34*, 548–558. doi:10.1207/s15374424jccp3403_10.

Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 143–189). New York: John Wiley and Sons.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69*, 159–172. doi:10.1037/0022-006X.69.2.159.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., et al. (2004). *Administration and scoring manual for the outcome questionnaire -45*. Orem, UT: American Professional Credentialing Services.

Lambert, M. J., Whipple, J. L., Bishop, M. J., Vermeersch, D. A., Gray, G. V., & Finch, A. E. (2002a). Comparison of empirically-derived and rationally-derived methods for identifying patients at risk for treatment failure. *Clinical Psychology & Psychotherapy, 9*, 149–164. doi:10.1002/cpp.333.

Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., et al. (2002b). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology & Psychotherapy, 9*, 91–103. doi:10.1002/cpp.324.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2*, 53–70. doi:10.1111/j.1745-6916.2007.00029.x.

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. doi:10.1037/0022-006X.69.2.150.

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. doi:10.1037/0022-006X.67.4.571.

Mohr, D. C. (1995). Negative outcome in psychotherapy: A critical review. *Clinical Psychology: Science and Practice, 2*, 1–27.

National Advisory Mental HealthCouncil. (2001). *Blueprint for change: Research on child and adolescent mental health. A report by the national advisory mental health council's workgroup on child and adolescent mental health intervention development and deployment*. Bethesda. MD: National Institutes of Health/National Institute of Mental Health.

Ringel, J. S., & Sturm, R. (2001). National estimates of mental health utilization and expenditures for children in 1998. *J. Behav. Health Serv. Res., 28*, 319–333. doi:10.1007/BF02287247.

Salzer, M. S., Bickman, L., & Lambert, E. W. (1999). Dose-effect relationship in children's psychotherapy services. *Journal of Consulting and Clinical Psychology, 67*, 228–238. doi:10.1037/0022-006X.67.2.228.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford.

Spielmans, G. I., Masters, K. S., & Lambert, M. J. (2006). A comparison of rational versus empirical methods in the prediction of psychotherapy outcome. *Clinical Psychology & Psychotherapy, 13*, 202–214. doi:10.1002/cpp.491.

Weiss, B., Catron, T., Harris, V., & Phung, T. M. (1999). The effectiveness of traditional child psychotherapy. *Journal of Consulting and Clinical Psychology, 67*, 82–94. doi:10.1037/0022-006X.67.1.82.

Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples*. Cambridge: Cambridge University Press.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between lab and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 688–701. doi:10.1037/0022-006X.63.5.688.

Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problem solving strategies in routine practice. *Journal of Counseling Psychology, 50*(1), 59–68. doi:10.1037/0022-0167.50.1.59.

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology, Research and Practice, 24*, 190–195. doi:10.1037/0735-7028.24.2.190.