

Marquette University

e-Publications@Marquette

Psychology Faculty Research and Publications

Psychology, Department of

6-2006

Empirically and Clinically Useful Decision Making in Psychotherapy: Differential Predictions With Treatment Response Models

Wolfgang Lutz
University of Trier

Stephen M. Saunders
Marquette University, stephen.saunders@marquette.edu

Scott C. Leon
Loyola University Chicago

Zoran Martinovich
Northwestern University

Joachim Kosfelder
University of Bochum

See next page for additional authors

Follow this and additional works at: https://epublications.marquette.edu/psych_fac



Part of the [Psychology Commons](#)

Recommended Citation

Lutz, Wolfgang; Saunders, Stephen M.; Leon, Scott C.; Martinovich, Zoran; Kosfelder, Joachim; Schulte, Dietmar; Grawe, Klaus; and Tholen, Sven, "Empirically and Clinically Useful Decision Making in Psychotherapy: Differential Predictions With Treatment Response Models" (2006). *Psychology Faculty Research and Publications*. 301.

https://epublications.marquette.edu/psych_fac/301

Authors

Wolfgang Lutz, Stephen M. Saunders, Scott C. Leon, Zoran Martinovich, Joachim Kosfelder, Dietmar Schulte, Klaus Grawe, and Sven Tholen

Marquette University

e-Publications@Marquette

Psychology Faculty Research and Publications/College of Arts and Sciences

This paper is NOT THE PUBLISHED VERSION; but the author's final, peer-reviewed manuscript. The published version may be accessed by following the link in the citation below.

Psychological Assessment, Vol. 18, No. 2 (2006): 133-141. [DOI](#). This article is © American Psychological Association and permission has been granted for this version to appear in [e-Publications@Marquette](#). American Psychological Association does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from American Psychological Association.

Empirically and Clinically Useful Decision Making in Psychotherapy: Differential Predictions With Treatment Response Models

Wolfgang Lutz

Department of Psychology, University of Berne, Berne, Switzerland;

Stephen M. Saunders

Department of Psychology, Marquette University

Scott C. Leon

Department of Psychology, Loyola University Chicago

Zoran Martinovich

Department of Psychiatry, Northwestern University Medical School

Joachim Kosfelder

Department of Psychology, University of Bochum, Bochum, Germany

Dietmar Schulte

Department of Psychology, University of Bochum, Bochum, Germany

Klaus Grawe

Department of Psychology, University of Berne, Berne, Switzerland

Sven Tholen

Department of Psychology, University of Berne, Berne, Switzerland

Acknowledgement

This work was partially supported by Swiss National Science Foundation Grants PP001–102651/1 and 1114-064884.01. We sadly note the death of Klaus Grawe during the revision process of this article. We thank Ken Howard and Klaus Grawe.

Abstract

In the delivery of clinical services, outcomes monitoring (i.e., repeated assessments of a patient's response to treatment) can be used to support clinical decision making (i.e., recurrent revisions of outcome expectations on the basis of that response). Outcomes monitoring can be particularly useful in the context of established practice research networks. This article presents a strategy to disaggregate patients into homogeneous subgroups to generate optimal expected treatment response profiles, which can be used to predict and track the progress of patients in different treatment modalities. The study was based on data from 618 diagnostically diverse patients treated with either a cognitive–behavioral treatment protocol ($n = 262$) or an integrative cognitive–behavioral and interpersonal treatment protocol ($n = 356$). The validity of expected treatment response models to predict treatment in those 2 protocols for individual patients was evaluated. The ways such a procedure might be used in outpatient centers to learn more about patients, predict treatment response, and improve clinical practice are discussed.

Keywords

expected treatment response, patient-focused research, outcomes management, adaptive decision making, differential predictions

Introduction

The scientist-practitioner gap related to the lack of practical implications of psychotherapy research is regularly noted in the field of psychotherapy (e.g., [Goldfried & Wolfe, 1998](#); [Newman & Tejada, 1996](#)). Clinicians criticize researchers for not producing clinically relevant research results, whereas researchers complain that few research findings and empirically supported treatments find their way into clinical practice.

Therapy-focused researchers attempt to redress the scientist-practitioner gap by developing and validating treatments that, ideally, will be applied in real-world clinical settings. Such research asks whether a specific treatment works. [Rounsaville, Carroll, and Onken \(2001\)](#) proposed a stage model of therapy research wherein treatments can be developed in research settings and ultimately transferred to clinical settings through a series of stages, including pilot investigations, randomized clinical trials, and evaluations in clinic settings. Hence, this model proposes that the internal validity of new treatments be established under well-controlled experimental conditions, whereas external validity evaluations require the use of more generalizable clinical conditions, samples, and designs (cf. [Kazdin, 2003](#)). The stage model of therapy research thus proposes how newly developed treatments can be validated in both the laboratory and in clinical practice.

Although this approach may address some of the issues causing the scientist-practitioner gap, an ongoing major difficulty is that therapy-focused research is not readily adopted by clinics and clinicians. Practitioners perceive the practical utility and external validity of newly developed treatments to be poor, as such treatments were developed and tested on groups of patients not representative of real-world clinical settings. Treatments must be both acceptable and feasible in real-world settings before they can or will be widely used in clinical practice. In contrast, newly developed treatments often entail complex protocols that require extensive training and supervision to guarantee minimal competency.

A different and supplemental approach to bridging the scientist-practitioner gap is to encourage clinics and clinicians to use *patient-focused* research methodologies ([Lambert, 2001](#)). Whereas treatment-focused research asks whether a specified intervention (e.g., cognitive-behavioral treatment [CBT] or interpersonal treatment) can work, patient-focused research seeks to determine the extent to which the ongoing treatment is working for a given patient ([Howard, Moras, Brill, Martinovich, & Lutz, 1996](#)). Patient-focused research concentrates on maximizing the treatment success of individual patients and thereby has many similarities to therapy as it is actually conducted in clinical settings. It is likely to be both more acceptable and more feasible within clinical settings ([Lambert, Hansen, & Finch, 2001](#); [Lambert et al., 2003](#)).

Expected Treatment Response (ETR)

Prior patient-focused research has used the ETR method ([Lambert et al., 2003](#); [Lueger et al., 2001](#)), which evolved from the dose-effect model ([Hansen, Lambert, & Forman, 2002](#); [Howard, Kopta, Krause, & Orlinsky, 1986](#)). The ETR method uses data and research findings that are generated as part of a specific clinical practice to inform and supplement decision making within that clinical practice. Essentially, patient outcomes assessment data collected at an agency are conceptualized as a documentation of prior experience. Each clinic collects its own data over a period of time. These data are used to generate expected treatment response curves (ETR curves) for that clinic's future patients. The ETR method applies research findings from the psychotherapy literature to such data to learn and improve from that prior experience (cf. [Fishman, 1999](#)). Using aggregated longitudinal data collected at the clinic, the ETR method uses growth curve analyses to generate an ETR curve for each individual patient.

For example, in one study, individual differences in change were predicted using seven indices of patient status measured at intake: self-report indices of well-being, symptoms and functioning, previous therapeutic experience, duration of the problem, patient's expectation of improvement in psychotherapy, and evaluation of the patient's functional status by the therapist ([Lutz, Martinovich, & Howard, 1999](#); [Lutz, Martinovich, Howard, & Leon, 2002](#)). Each patient's expected change over treatment was modeled on the log-transformed number of sessions. That is, an ETR curve (the expected patient improvement on the outcome measurement over time) was generated using patient information from the seven predictors. The ETR curves were compared with the patients' actual responses to treatment to evaluate treatment progress. The resulting model explained approximately 22% of the interindividual differences in the estimated individual growth parameters. Pretreatment distress, which was assessed with the self-report indexes of well-being and symptoms and functioning, accounted for 18% of the variance, and the other variables combined only accounted for 4% of the variance.

The reliability of the initial predictions has subsequently been investigated and supported ([Leon, Kopta, Howard, & Lutz, 1999](#); [Lueger et al., 2001](#); [Lueger, Lutz, & Howard, 2000](#)). The clinical usefulness of the ETR method has also been evaluated. Research indicates that this method can be validly applied to different diagnostic groups and a variety of symptom pattern presentations ([Lutz, Lowry, Kopta, Einstein & Howard, 2001](#)). [Lutz, Rafaeli, Howard, and Martinovich \(2002\)](#) found that the predictive accuracy of ETR curves was enhanced by correcting them on the basis of the change a patient has experienced up to that point. [Lambert et al. \(2001\)](#) compared a rational method of predicting patient treatment failure with a statistical ETR technique and showed essential

equivalence, though the statistical approach was marginally more accurate. [Finch, Lambert, and Schaalje \(2001\)](#) modeled recovery curves separately for initial level of disturbance in small score bands.

The Present Study

Building on prior research, the present study extends and refines the original ETR method. The aim was to extend earlier methods by using more patient-specific sampling strategies. In previous versions of this methodology, the validity of the use of ETR study prediction weights for any particular patient depended on the extent to which the study sample was representative of the population of which that patient was a member. For this reason, most predictors worked only for specific subsets of patients ([Krause, Howard, & Lutz, 1998](#)). To address this problem, the present study used *nearest neighbor* techniques, which are drawn from research strategies used to estimate the probabilities of avalanches occurring ([Brabec & Meister, 2001](#)). To be specific, this strategy identifies those patients (who have already been treated) that most closely match the target patient (hence nearest neighbors) on a number of criteria or variables. The strategy selects a homogeneous subgroup on the basis of initial patient information and bases the prediction of treatment progress on those patients ([Beutler, 2001](#)).

This article presents the results of a study evaluating the predictive validity and the clinical utility of ETRs generated using this homogeneous subsample method (HSM; herein referred to as the ETR via HSM method). The advantage of the ETR via HSM method over the traditional ETR method in predicting the course of treatment for patients in outpatient psychotherapy on the basis of initial psychometric information has recently been investigated ([Lutz et al., 2005](#)). Thus, the ETR via HSM method can be used to forecast individual patient progress, and this expected progress can be compared with actual progress to evaluate change on a patient level.

Moreover, the ETR via HSM method might further be used in clinical settings either to determine what treatment modality (e.g., medication, psychotherapy, a combination of medication and psychotherapy) is most likely to result in positive change or to evaluate the progress of an individual patient by contrasting actual progress to expected progress in that treatment modality. Whereas in most clinical cases, progress is evaluated by comparing current scores with pretreatment scores, this method allows comparison of predicted (ETR curves) and actual progress of patients. For example, if expected progress is minimal, subsequent actual minimal progress might be considered successful. In contrast, expected progress might be substantial, such that actual moderate progress might be considered insufficient.

To assess whether such a system could provide a valid forecast for the optimal treatment modality for a given patient, we considered (a) whether the ETR via HSM method identified differential optimal treatment predictions for different patients and (b) whether, among cases with a clear HSM-identified preferred treatment, actual outcomes were correlated with a match to the preferred treatment.

Method

Participants

A sample of 618 patients treated at the outpatient clinics at the Departments of Psychology of the University of Berne (Switzerland) and the University of Bochum (Germany) were recruited. Patients at the Berne setting were treated using an integrative cognitive-behavioral and interpersonal treatment (IT) protocol ([Grawe, 1998, 2002](#)), and patients at Bochum were treated via a CBT protocol ([Schulte & Eifert, 2002](#); [Schulte, Künzel, Pepping, & Schulte-Bahrenberg, 1991](#)). In both settings, therapists (postgraduate psychologists) had 2–5 years of training using the treatment protocols.

IT sample

The Berne patient sample ($n = 356$) in the IT protocol were two thirds women and had an average age of 35.5 years. The average length of treatment in this protocol was 29.1 sessions. Diagnoses were based on the Structured Clinical Interview for *DSM-IV* (SCID; [First, Spitzer, Gibbon, & Williams, 1995](#)). Accordingly, patients in the IT protocol met diagnostic criteria for a mood disorder (30%), an anxiety disorder (40%), an adjustment disorder (9%), a somatoform disorder (6%), a sexual dysfunction (2%), or another psychological disorder (13%). Patients were treated by 63 therapists, who saw 1–25 patients. About two thirds (59.3%) of the therapists were women, and one third (40.7%) were men. The average age of the therapists was 42.7 years ($SD = 6.79$).

CBT sample

The Bochum patient sample ($n = 262$) in the CBT protocol were two thirds women and had an average age of 36.4 years. The average length of treatment in the CBT protocol was 23.5 sessions. On the basis of the SCID, patients in the CBT protocol were diagnosed with an anxiety disorder (48%), a mood disorder (17.8%), or comorbid anxiety and depressive disorders (34%). Patients were treated by 26 therapists, who saw 2–17 patients. About two thirds (62.3%) of the therapists were women, and one third (37.7%) were men. The average age of the therapists was 33.3 years ($SD = 5.62$).

Measures

In addition to the SCID, the intake status of all patients was evaluated using various measures and associated scales. All measures are part of the standard assessment procedure at the clinics.

Emotional distress

Patients' general psychological and emotional distress was assessed at intake and over the course of treatment using the Inventory of Emotional Distress (EMI; [Ullrich de Muynck & Ullrich, 1977](#)). Respondents were asked to report how they felt during the last week by endorsing a list of 70 bipolar adjectives (e.g., *tense* vs. *relaxed*) on a 6-point scale (e.g., 1 = *very tense*, 6 = *very relaxed*). The EMI yields an overall score and seven subscales. In this study, only the EMI overall score was used. The EMI overall score has good psychometric properties. It has high internal consistency ($\alpha = .98$) and correlates highly (.69) with the General Symptom Index score of the Symptom Checklist-90—Revised (SCL-90-R; [Derogatis, 1977](#)). The test-retest reliabilities (comparing scores from Weeks 2 and 4) ranged from .63 (Loneliness subscale) to .89 (Inhibition subscale). To better make use of the EMI, we converted the overall score into T scores ($M = 50$, $SD = 10$) on the basis of available normative data. Higher EMI T scores represent better psychological health ([Lutz, 2002](#)).

Symptomatology

Patients' symptoms at intake were assessed using the SCL-90-R. The scale contains 90 items that are rated on a 5-point scale ranging from *not at all* to *extremely*. The study used eight of the primary symptom subscales: Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, and Paranoid Ideation. Most subscales have good psychometric properties and established validity. The subscale Psychoticism, however, does not have good reliability, so it was excluded from this study (see also [Lutz, 2002](#)). The SCL-90-R has often been used to assess a range of psychotherapy outcomes and has demonstrated adequate test-retest and internal consistency reliability ([Lambert & Ogles, 2003](#)).

Interpersonal problems

Patients' role functioning at intake was measured using a revised, shortened version of the Inventory of Interpersonal Problems (IIP; [Horowitz, Rosenberg, Baer, Ureño, & Villasenor, 1988](#)) created for this study. The shortened IIP was created from the original 64 items and was derived from analyses of baseline data from 377 psychotherapy outpatients at the University of Berne clinic. Several factor analyses (principal component with varimax rotation) were conducted. At each sequential step, items were selected with a loading of at least .6 on

one of the components, without loading on another component more than .3. This procedure was repeated four times with progressively smaller item sets to produce short and reliable scales (Lutz, 2002). These analyses resulted in four subscales of three items each with good internal consistency reliabilities and low interscale correlations: Conflict Proneness ($\alpha = .86$), Lack of Assertiveness ($\alpha = .77$), Problems with Intimacy ($\alpha = .75$), and Overly Trustful ($\alpha = .74$). These four subscales were used in the present study.

Procedure

All patients seen at the outpatient clinics at the Departments of Psychology of the University of Berne and the University of Bochum were recruited to the study. As part of normal clinic routine, all patients completed intake questionnaires and the SCID. Treatment progress was evaluated with the EMI and session reports from both the therapists and the patients. To be included in the study, patients had to undergo the intake assessment procedure of the clinics, had at least 2 sessions of treatment, and had at least 2 data points (outcome assessments). Study participants had a minimum of 2 (pre- and posttreatment) and a maximum of 15 assessments with the EMI during the course of treatment. This resulted in unequal numbers of patients in the two treatment groups.

Patients were 16- to 73-years-old. Table 1 shows the intake characteristics (predictor variables) of the patient sample as a whole and by patient groups (IT and CBT). As can be seen, the samples differ somewhat in intake status with respect to several of the intake predictors. The patient-specific sampling strategy of collecting a homogeneous subgroup for each patient is particularly valuable given intake differences between patient groups.

Table 1. Sample Characteristics at Intake for the Full Sample and the Patient Groups (CBT and IT)

Predictor variable	Full sample (N=618)		CBT (n=262)		IT (n=356)		T
	M	SD	M	SD	M	SD	
Somatization	0.95	0.71	1.04	0.76	0.89	0.67	-2.57*
Obsessive-compulsive	1.40	0.80	1.32	0.83	1.46	0.78	2.06*
Interpersonal sensitivity	1.38	0.89	1.31	0.96	1.43	0.83	1.46
Depression	1.74	0.90	1.57	0.92	1.86	0.86	3.99**
Anxiety	1.39	0.87	1.40	0.92	1.38	0.84	-0.33
Hostility	1.00	0.76	0.97	0.77	1.02	0.75	0.77
Phobic anxiety	0.99	0.98	1.17	1.09	0.86	0.87	-3.79**
Paranoid ideation	1.11	0.84	1.07	0.92	1.13	0.78	0.80
Conflict proneness	0.81	0.81	0.87	0.76	0.76	0.84	-1.63
Lack of assertiveness	2.30	0.99	2.38	1.05	2.24	0.95	-1.79
Problems with intimacy	1.53	0.96	1.58	1.04	1.49	0.90	-1.19
Overly trustful	1.39	0.97	1.51	0.97	1.30	0.95	-2.62**
Inventory of Emotional Distress	45.21	8.89	45.33	9.73	45.12	8.24	-0.28

For this study, the EMI overall score was the dependent variable (the outcome variable). ETR curves relating EMI to session number were generated for individual patients and the accuracy of the ETR via HSM method was evaluated by comparing those curves with actual, subsequent progress.

The specific analyses and procedures used for determining, applying, and evaluating the ETR via HSM method are described in detail in the Results section.

Results

The results of a random intercept and slope model of growth in emotional distress (based on results of EMI scores) as a function of treatment modality (CBT or IT) using a hierarchical linear modeling approach can be seen in [Table 2](#) (e.g., [Raudenbush, 2001](#); [Raudenbush & Bryk, 2001](#)). Growth was modeled on EMI scores (dependent variable) over time, regressing EMI scores on the logarithm of the number of sessions (independent variable of Level 1 in the growth-curve terminology of [Raudenbush, 2001](#); for more information about the use of log-transformed number of sessions variable see [Lambert et al., 2001](#) and [Lutz et al., 1999](#)). Therefore, the intercept of slopes in [Table 2](#) represents the average growth rate over both treatments. As can be seen in [Table 2](#), there is significant growth over the course of treatment of, on average, about a one-half standard deviation in the first 10 sessions (expressed in T scores, 4.83 per log 10 of session number). [There is no significant impact of treatment modality on intercept \(Level 1 predictor\) or change \(Level 2 predictor\). This indicates that there is no significant difference between treatment modalities in average change rates or intake scores.](#)

Table 2. Random Intercept and Slope Model of Growth in Emotional Distress (Modeled on EMI Scores) as a Function of Treatment Modality (CBT or IT)

Fixed effect	Coefficient	SE	T	P
For intercept (π_{0i})				
Intercept (π_{00})	44.76	0.41	108.00	<.01
Treatment modality (β_{01})	-0.36	0.56	-0.64	.52
For slope (π_{1i})				
Intercept (β_{10})	4.83	0.59	8.70	<.01
Treatment modality (β_{11})	0.60	0.75	0.25	.43

Note. N = 618. Random effect variation component for slope (v_{1i}) = 56.35; random effect variation component level 1 error (r_{it}) = 27.49. EMI = Inventory of Emotional Distress; CBT = cognitive-behavioral treatment; IT = integrative cognitive-behavioral and interpersonal treatment.

The ETR via HSM method allows exploration of potential treatment differences at the individual patient level. From the therapy-focused perspective, there was, on average, no difference between treatment modalities detected. The ETR via HSM method addresses the patient-focused question of whether treatment modality makes a difference for some patients.

Identifying Homogeneous Subsamples

The ETR via HSM method required identifying homogeneous subsamples for each patient. To define the homogeneous subsamples, we calculated euclidean distances among all cases using differences in responses to the intake measures (see [Table 1](#)): eight of the SCL-90-R symptom subscales, the four shortened subscales of the IIP, and the EMI overall score. All scores were standardized (converted to z scores, $M = 0$, $SD = 1$) so that they could be combined despite different metrics. The euclidean distance between two cases was calculated as the square root of the sum of the squared differences in values for each of the standardized variables. The euclidean distances among all participants was plotted into a 618 × 618 matrix (the number of participants). [Table 3](#) shows a portion of the matrix. Higher values represent more distance between the cases, whereas lower values represent less distance. Note that the values comparing a patient with him- or herself are zero, indicating identical presentations. In this manner, the 30 most similar patients from the IT group (the IT homogeneous subsample) and the 30 most similar patients from CBT group (the CBT homogeneous subsample) for each of the 618 patients were identified.

Table 3. Excerpt of the Euclidean Distances Analysis Matrix

Patient	1	2	3	4
---------	---	---	---	---

1.	--	9.14	8.01	10.10
2.		--	8.81	12.66
3.			--	7.77
4.				--

Note. Each score represents the distance scores between each case over all the standardized independent variables. Higher values represent more distant patients, lower values represent closer patients.

Calculating ETRs via HSM

Two growth curve models (ETR curves) were computed for each patient: one using the treatment responses of the CBT homogeneous subsample and one using the treatment responses of the IT homogeneous subsample. ETR curves were calculated using the same growth curve model as described above. The homogeneous subsamples for each patient do not include the patient, which makes it possible to validate this procedure on the full sample.

Validation Procedure

To evaluate the validity of the ETR via HSM method, we selected cases for which (a) the ETR curve predicted for at least one treatment modality (IT or CBT) was not positive and (b) the ETR curves predicted for each treatment modality were meaningfully different from each other.

Determining differential predictions for both modalities on a patient level

First, any case wherein the ETR curve slopes for both treatments exceeded 6 (which represents approximately the 99% confidence interval upper bound above the average slope of 4.83, see [Table 2](#)) were excluded, even when there was a significant difference between the two treatments, because such a slope indicated that both treatments had a clear positive prediction. In such a case differential predictions are not meaningful, because the predictions for both modalities are clearly positive, indicating that it does not make a difference which treatment strategy is followed. This resulted in 461 remaining cases.

Second, cases were selected in which the difference in predicted slopes for the two treatment modalities was reliably different. To be specific, the slopes for each patient were categorized as either “predict better outcome with CBT” or “predict better outcome with IT.” This was operationalized as cases in which the slope values for the two treatment modalities differed by at least 1.65, which represents approximately one third of the average slope for the whole sample (average slope = 4.83, see [Table 2](#)). If the slope of the ETRs differed by 1.65 or more, this represented a difference in predicted T scores of 1.5 at Session 8 and a 2-point T score difference at Session 15 between both modalities. Thus, the two HSM samples were required to demonstrate rates of change that implied a “greater than small” difference by Session 15 (i.e., $SD = 0.2$). This difference occurred in about one third (36.66%) of the remaining cases (169 of 461) and 27.35% of the full sample (169 out of 618).

Evaluating the validity of the ETR via HSM method for differential predictions of treatment modality

For each of the remaining 169 cases, the predictions (regarding which treatment would be better) were compared with the actual outcomes of that patient in the actual treatment modality. Treatment success for the actual outcome was defined as in a prior study ([Lutz et al., 1999](#)) with a reliability-based improvement criterion (cf. [Jacobson & Truax, 1991](#); [Martinovich, Howard & Saunders, 1996](#); [Saunders, Howard, & Newman, 1988](#)). That is, reliable improvement was operationalized as a change of 1.28 SEMs from reliability-adjusted intake EMI scores to EMI scores at treatment termination. (When calculated from internal consistency reliability, standard error of the mean represents the standard deviation of observed scores around a theoretical true score. Assuming normally distributed measurement error, one may be 90% confident that if an observed score falls more than 1.28 SEMs above some value, the true score is indeed above that value.)

Applying the reliable improvement criterion, each comparison of the ETR curve and actual results was categorized as either supporting or not supporting the ETR via HSM method. There were two possible outcomes supporting and two possible outcomes not supporting the ETR via HSM method for each case. To illustrate, consider a case treated with CBT. Two outcomes comparing the ETR curve and actual results were deemed to support the ETR via HSM method: First, the ETR curve predicted that a patient would have better success with CBT than IT, and success was indeed observed for that patient in CBT; second, the ETR curve predicted that a patient would not be as successful in CBT as in IT, and no success was observed in CBT. Two outcomes would not support the ETR via HSM method: First, the ETR curve predicted that a patient would experience better progress with CBT than IT, but no success was observed in CBT; second, the ETR curve predicted that the patient would not be as successful in CBT as in IT, but success in CBT was observed. The same four options were evaluated for each case treated in IT. This procedure is a stringent test of the hypothesis, which is that the predictions will be accurate regarding the more successful treatment modality. An obvious limitation of this approach is that there is no actual comparison with the treatment modality not used. That is, one cannot know how the patient would have done if treated with the other modality. But the accuracy of the predictions regarding the preferable modality can be evaluated.

The results showed that the ETR via HSM method correctly predicted treatment modality in 100 of 169 (59.17%) cases, $\chi^2(1, N = 169) = 5.69, p = .02$. This means that a chance or expected distribution for the prediction of treatment modality would have been a 50%–50% distribution equally favoring and not favoring the prediction model, and we observed a statistically different distribution of 59.17% to 40.83% favoring the ETR via HSM method.

Examples

Example 1

An example of applying this ETR via HSM method is presented in [Figure 1](#). This patient is a 49-year-old woman diagnosed with major depression and social phobia. She attended 33 sessions of IT treatment. She had elevated scores on the Depression, Anxiety, Phobic Anxiety, and (to a lesser extent) Hostility subscales of the SCL-90-R.

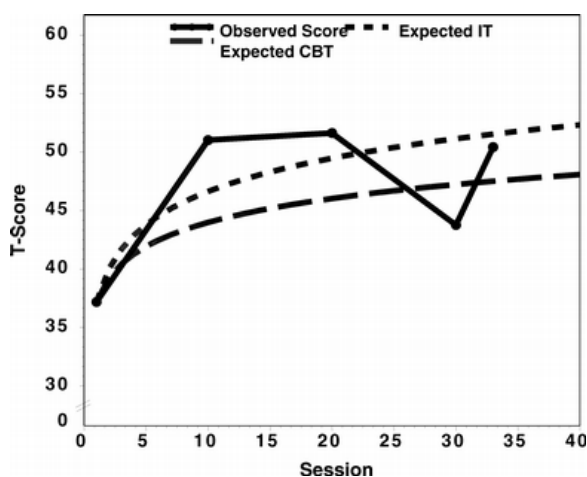


Figure 1. Predicted and actual treatment course for Example 1. Observed score = Inventory of Emotional Distress score for integrative cognitive-behavioral and interpersonal treatment (IT); expected IT = expected treatment response curve for IT; expected CBT = expected treatment response curve for cognitive-behavioral treatment

[Figure 1](#) shows the ETR curve, on the basis of the homogenous subsample method, for the EMI scale (using T scores) for both CBT and IT. On the basis of the ETR via HSM method, this patient is expected to do better in IT than CBT. [Figure 1](#) also shows the patient's actual EMI scores over IT treatment. As can be seen, the patient's

progress is somewhat better than expected through Session 20. Between Sessions 20 and 30, however, her EMI T scores declined precipitously and contrary to expectations. However, she was within expected range again by Session 33.

The ETR via HSM method generates other useful information. [Table 4](#) presents additional statistics about the homogeneous subsamples derived from the CBT and IT patients for this case. [Table 4](#) shows the average age, gender, and diagnostic composition of each treatment modality's subsample for this case. It also shows the average distance between the case and the 30 members of the subsample. In this example, the case had an overall distance score of 6.6 for the IT subsample and 7.7 for the CBT subsample, indicating that she is more similar to the IT subsample. Such information might be taken into consideration when deciding on the representativeness of the homogeneous subsample for the incoming patient. This information might be also useful for a therapist to consider. For example, he or she could seek advice from a therapist who had already treated several patients in that most representative group.

Table 4. Additional Information on Homogeneous Subsamples for Patient Example 1

	Sample demographics							
	Age (years)					Ex. 1 distance		
Treatment	M	SD	Gender	Diagnoses	M	SD	SE	Reliable change (n)
IT (n=30)	35	12	24 F 6 M	5 anxiety 10 depression 1 comorbid	6.6	0.9	6.2	+26 ±3 -1
CBT (n=30)	33	3.0	16 F 14 M	5 anxiety 10 depression 15 comorbid	7.7	1.4	4.7	+17 ±11 -2

Note. Diagnoses summarize the diagnostic composition of the 30 patients; distance measures euclidean distances between the case and the 30 members of each subsample; SE measures the expected treatment response function for the homogeneous subsamples; reliable change reflects the treatment outcomes of the 30 previously treated patients defined by the reliability-based improvement criterion used in this study. F = female; M = male; + = reliable improvement; ± = no reliable change; - = reliable deterioration; IT = integrative cognitive-behavioral and interpersonal treatment; CBT = cognitive-behavioral treatment.

[Table 4](#) also shows the treatment outcomes of the 30 previously treated patients most similar to the Example 1 patient (note that such information is available at intake). This is shown in [Table 4](#), in the column reliable change, for the two modalities with the previously described reliability-based improvement criterion for each of the 30 patients (+ = reliable improvement; ± = no reliable change; - = reliable deterioration). The standard error of estimate for the ETR slope coefficient is indicated in column SE in [Table 4](#). This provides an estimate of the observed variation in scores around the predicted trajectory. Higher SE scores indicate that patients like the patient being evaluated show relatively more fluctuation around the general trend during their treatment course. If substantial variation is demonstrated in the homogeneous ETR sample, some positive or negative extreme scores for a case might be expected. That is, a therapist perhaps should not be too optimistic in response to a single extreme positive score but also not too pessimistic in response to an isolated negative score.

Example 2

Another example of the ETR via HSM method is shown in [Figure 2](#). This patient is a 41-year-old woman diagnosed with an anxiety disorder (panic disorder with agoraphobia). The ETR curve predicted that she would obtain clinically reliable success in CBT and that CBT would be more successful than IT. She was successfully treated with CBT. [Table 5](#) shows the additional information about the homogeneous subsamples for this patient. For example, the CBT homogeneous subsample includes more patients with an anxiety disorder. In addition, the expected variation (*SE* in [Table 5](#)) around the general trend is slightly less than for the IT group, suggesting less fluctuation around a general trend during treatment for this group of previously treated patients.

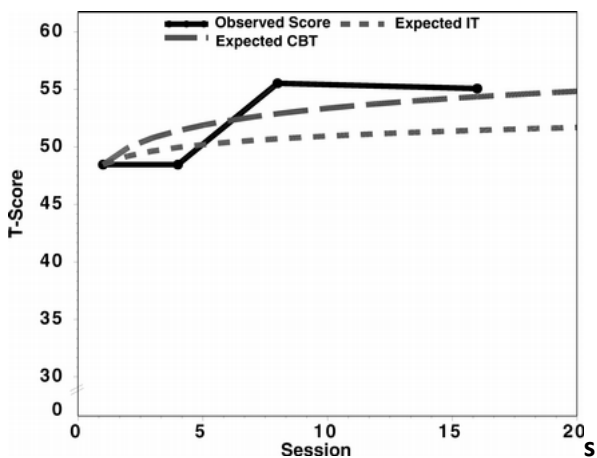


Figure 2. Predicted and actual treatment course for Example 2. Observed score = Inventory of Emotional Distress score for cognitive-behavioral treatment (CBT); expected IT = expected treatment response curve for integrative cognitive-behavioral and interpersonal treatment; expected CBT = expected treatment response curve for CBT

Table 5. Additional Information on Homogeneous Subsamples for Patient Example 2

	Sample demographics			Ex. 2 distance				
	Age (years)			Diagnoses	M	SD	SE	Reliable change (n)
Treatment	M	SD	Gender					
IT (n=30)	40.4	2.7	15 F	8 anxiety	8.7	1.4	5.7	+15
			15 M	2 depression				±13
				2 comorbid				-2
CBT (n=30)	41.1	3.3	19 F	19 anxiety	8.4	1.8	4.6	+19
			11 M	2 depression				±7
				9 comorbid				-4

Note. Diagnoses summarize the diagnostic composition of the 30 patients; distance measures euclidean distances between the case and the 30 members of each subsample; SE measures the predicted courses for the homogeneous subsamples; reliable change reflects the treatment outcomes of the 30 previously treated patients defined by the reliability-based improvement criterion used in this study. F = female; M = male; + = reliable improvement; ± = no reliable change; - = reliable deterioration; IT = integrative cognitive-behavioral and interpersonal treatment; CBT = cognitive-behavioral treatment.

Discussion

This article presents a strategy to identify homogeneous subsamples for individual patients to generate optimal ETR profiles that can be used to both predict and evaluate their progress in different treatment protocols. The 30 most similar participants from each of the groups (CBT and IT) were selected as homogeneous subsamples

that were used to generate expected trajectories for each case under each treatment modality. The ETR via HSM method created meaningful and different predictions between the treatment modalities for about 27% of the patients. Validation analyses compared the prediction from the ETR via HSM method with actual outcomes within the conducted treatment modality. Results supported the validity of the ETR via HSM method.

Thus far, empirically based treatment decisions have been informed by guidelines supported by research examining average effects on relatively heterogeneous patient groups. The ETR via HSM method, in contrast, presumes that a more accurate predicted treatment response can be generated by using homogeneous subsamples of previously treated patients. It is argued that recent advances in health care technology have made ETR via HSM methodology increasingly feasible. Technologically advanced information systems are being increasingly incorporated into health service settings. Likewise, there is a growing interest in empirical outcomes assessment. As a result, large naturalistic databases documenting response to psychotherapy across time are becoming more readily available. These databases may include information on a variety of variables that may impact outcomes, including treatment, patient, clinician, and setting characteristics as well as contractual or financial information. Reasonably large, homogeneous subsamples can be culled from larger databases on the basis of similarity for any particular patient. The ETR via HSM method thereby enables patient-specific statistics to assist in treatment planning and to predict and assess ongoing outcomes. Patient-specific sampling can be used to describe what “treatment as usual” is for patients most resembling any particular case. In addition, optimal and suboptimal treatment strategies and settings, clinician matches, and contractual conditions for a given case can be identified. Such information could be used for intake decisions, decisions about ongoing treatment, and the improvement of clinical practice.

In this study, we used treatment modality as a major determinant of ETRs. However, we acknowledge that treatment modality is a problematic variable in the field of psychotherapy research. Moreover, it is a largely ignored variable in psychotherapy as conducted in the real world. In both research and practice, treatment modality as a variable presents problems with treatment compliance and integrity. However, it is a primary variable in the field, especially with respect to randomized clinical trials. We believe that psychotherapy research should move beyond an average comparison between treatment modalities (such as psychotherapy vs. pharmacotherapy). Research should instead strive to identify and explore, if possible, differences on an individual patient level. We believe this article demonstrates the potential of such an approach.

ETR via HSM methodology also can guide future research topics. For example, given large databases, specific shapes of change could be investigated (see [Barkham et al., 2001](#)). As one example, ETR via HSM could be used to predict the likelihood that a client would experience a sudden gain or a sudden deterioration in mental health status ([Stiles et al., 2003](#); [Tang & DeRubeis, 1999](#)).

However, this research is in an early stage and many questions remain to be answered. The applicability of this method is probably not limited to the instruments used in this study, but this has to be investigated. Indeed, it may be shown that ETR via HSM is even more useful and accurate when used with measures that were specifically developed for outcomes management (e.g., [Lambert et al., 2003](#)).

In future research, some other metric besides euclidean distances could be used to define the homogeneous neighbors. Variables could be differentially weighted in the analysis to reflect estimates of their contribution to relevant similarity. In this study, initial disturbance was weighted heavily because the predictors were symptom intensity measures (subscales of SCL-90-R, IIP, and EMI). From a practical point of view, to make these models more easily applicable in routine care, there might be reason to use a single measure of initial distress rather than an extensive battery of tests ([Finch et al., 2001](#)).

In addition, there are problems associated with the ETR via HSM validation procedure with this data set. Consider the case in which a patient is treated with CBT: The ETR via HSM method predicts that the preferable

treatment was IT, and there was indeed a poor outcome with CBT. This result was determined to support the validity of the ETR via HSM method. However, by definition, it cannot be known what would have happened if the other modality had been used for that patient. In the chi-square test, it was assumed that such error would be normally distributed in both groups equally favoring the model and not favoring it. The only study that could actually evaluate this issue would need to use a prospective design, in which during the course of treatment the modality could be changed (if no progress was made with the first chosen modality and the alternative treatment option could directly be tested). Another limitation of this method concerns the availability of needed computer software. ETR via HSM depends on a database of patients that can easily be continuously and accurately updated with new information. This is specifically necessary because in clinical practice advanced statistical analyses tools as used in this study are usually not available. The use of the system further requires a practitioner or scientist practitioner network (e.g., [Borkovec, Echemendia, Ragusea, & Ruiz, 2001](#)), which would make the results available for feedback to other practitioners. Thus, at this stage a single clinician would likely be unable to apply such a system.

Additionally, future studies must investigate whether such a system can indeed improve actual clinical services. Likewise, it must be determined whether the limitations of such a system (e.g., potential wrong classifications) are not excessively detrimental. In the present study, there was a meaningful differential prediction for about one fourth of the cases, and the system was not correct in all of those cases. In other words, the actual clinical utility of this approach must still be evaluated.

Further research also has to be done to define decision rules on the basis of such models. In addition, it must be determined whether such prediction models are better at identifying negative treatment courses than, for example, rationally derived rules ([Lambert et al., 2001](#)). If this were shown, then these models would be useful for ongoing decision making and quality assurance during the course of treatment.

Despite the preliminary nature of the results of this study, such a system could be extremely helpful. It is already possible to generate forecasts that could empirically support clinical decisions concerning treatment progress of individual patients as well as differential decisions concerning treatment modalities. Both the clinician and the researcher can benefit from this endeavor: The clinician gains the possibility of integrating research information into everyday clinical duties and the researcher gains a large research database of treated cases to improve models and advance the scientific integrity of the field.

Footnotes

¹ This statement holds only for the first 10 sessions. Because of the log transformation of the session variable, the predictions follow a negatively accelerating curve (see also [Lutz et al., 1999](#)).

References

- Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., et al. (2001). Service profiling and outcomes benchmarking using CORE-OM: Toward practice-based evidence in psychological therapies. *Journal of Consulting and Clinical Psychology, 69*, 184–196.
- Beutler, L. (2001). Comparisons among quality assurance systems: From outcome assessment to clinical utility. *Journal of Consulting and Clinical Psychology, 69*, 197–204.
- Borkovec, T. D., Echemendia, R. J., Ragusea, S. A., & Ruiz, M. (2001). The Pennsylvania Practice Research Network and future possibilities for clinically meaningful and scientifically rigorous psychotherapy effectiveness research. *Clinical Psychology: Science and Practice, 8*, 155–167.
- Brabec, B., & Meister, R. (2001). A nearest-neighbor model for regional avalanche forecasting. *Annals of Glaciology, 32*, 130–134.
- Derogatis, C. R. (1977). *SCL-90*. Administration, scoring and procedures. Manual-I for the R(evised) version and other instruments of the psychopathology rating scale series. Baltimore: Johns Hopkins University School of Medicine.

- 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 and Psychotherapy, 8*, 231–242.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). *Structured Clinical Interview for Axis I DSM-IV Disorders—Patient edition (SCID-I/P)*. New York: Biometrics Research Department, New York State Psychiatric Institute.
- Fishman, D. B. (1999). *The case for pragmatic psychology*. New York: New York University Press.
- Goldfried, M. R., & Wolfe, B. E. (1998). Toward a more clinically valid approach to therapy research. *Journal of Consulting and Clinical Psychology, 66*, 143–150.
- Grawe, K. (1998). Research-informed psychotherapy. *Psychotherapy Research, 7*, 1–19.
- Grawe, K. (2002). *Psychological therapy*. Seattle, WA: Hogrefe.
- Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*, 329–343.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villasenor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885–892.
- Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-response relationship in psychotherapy. *American Psychologist, 41*, 159–164.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). The evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *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.
- Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Needham Heights, MA: Allyn & Bacon.
- Krause, M. S., Howard, K. I., & Lutz, W. (1998). Exploring individual change. *Journal of Consulting and Clinical Psychology, 66*, 838–845.
- Lambert, M. (2001). Psychotherapy outcome and quality improvement: Introduction to the special section on patient-focused research. *Journal of Consulting and Clinical Psychology, 69*, 147–149.
- Lambert, M., 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.
- Lambert, M. J., & Ogles, B. M. (2003). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., (pp. 139–193). New York: Wiley.
- Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10*, 288–301.
- Leon, S. C., Kopta, S. M., Howard, K. I., & Lutz, W. (1999). Predicting patients' responses to psychotherapy: Are some more predictable than others? *Journal of Consulting and Clinical Psychology, 67*, 698–704.
- Lueger, R. J., Howard, K. I., Martinovich, Z., Lutz, W., Anderson, E., & Grissom, G. (2001). Assessing treatment progress with individualized models of predicted response. *Journal of Consulting and Clinical Psychology, 69*, 150–158.
- Lueger, R. J., Lutz, W., & Howard, K. I. (2000). The predicted and observed course of psychotherapy for anxiety and mood disorders. *Journal of Nervous and Mental Disease, 188*, 127–134.
- Lutz, W. (2002). Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirical based clinical practice. *Psychotherapy Research, 12*, 251–273.
- Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., et al. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology, 73*, 904–913.
- Lutz, W., Lowry, J., Kopta, M., Einstein, A. D., & Howard, K. I. (2001). Prediction of dose-response relations based on patient characteristics. *Journal of Clinical Psychology, 57*, 889–900.

- Lutz, W., Martinovich, Z., & Howard, K. I. (1999). Patient profiling: An application of random coefficient regression models to depicting the response of a patient to outpatient psychotherapy. *Journal of Consulting and Clinical Psychology, 67*, 571–577.
- Lutz, W., Martinovich, Z., Howard, K. I., & Leon, S. C. (2002). Outcomes management, expected treatment response and severity adjusted provider profiling in outpatient psychotherapy. *Journal of Clinical Psychology, 59*, 1291–1304.
- Lutz, W., Rafaeli, E., Howard, K. I., & Martinovich, Z. (2002). Adaptive modeling of progress in outpatient psychotherapy. *Psychotherapy Research, 12*, 305–327.
- Martinovich, Z., Howard, K. I., & Saunders, S. (1996). Some comments on “Assessing clinical significance.” *Psychotherapy Research, 6*, 124–132.
- Newman, F. L., & Tejada, M. J. (1996). The need for research that is designed to support decisions in the delivery of mental health services. *American Psychologist, 51*, 1040–1049.
- Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology, 52*, 501–525.
- Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park: Sage.
- Rounsaville, B. J., Carroll, K. M., & Onken, L. S. (2001). A stage model of behavioral therapies research: Getting started and moving on from stage 1. *Clinical Psychology: Science and Practice, 8*, 133–142.
- Saunders, S. M., Howard, K. I., & Newman, F. L. (1988). Evaluating the clinical significance of treatment effects: Norms and normality. *Behavioral Assessment, 10*, 207–218.
- Schulte, D., & Eifert, G. H. (2002). What to do when manuals fail? The dual model of psychotherapy. *Clinical Psychology: Science and Practice, 9*, 312–328.
- Schulte, D., Künzel, R., Pepping, G., & Schulte-Bahrenberg, T. (1991). Tailor-made versus standardized therapy of phobic patients. *Advances in Behaviour Research and Therapy, 14*, 67–92.
- Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., et al. (2003). Early sudden gains in psychotherapy under routine clinic conditions: Practice-based evidence. *Journal of Consulting and Clinical Psychology, 71*, 14–21.
- Tang, Z. T., & DeRubeis, R. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*, 894–904.
- Ullrich de Muynck, R., & Ullrich, R. (1977). *Das Emotionalitätsinventar als Befindlichkeitsmaß Testmanual EMI-B*. [The Inventory of Emotional Distress—Manual EMI-B]. Munich, Germany: Pfeifer.