

**Marquette University**  
**e-Publications@Marquette**

---

Psychology Faculty Research and Publications

Psychology, Department of

---

1-1-2009

# Methodological Background of Decision Rules and Feedback Tools for Outcomes Management in Psychotherapy

Wolfgang Lutz  
*University of Trier*

Niklaus Stulz  
*Psychiatric Services Aargau*

Zoran Martinovich  
*Northwestern University*

Scott Leon  
*Loyola University Chicago*

Stephen M. Saunders  
*Marquette University, [stephen.saunders@marquette.edu](mailto:stephen.saunders@marquette.edu)*

---

Accepted version. *Psychotherapy Research*, Vol. 19, No. 4-5 (2009): 502-510. DOI. © 2009 Society for Psychotherapy Research. Used with permission.

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.

*Psychotherapy Research*, Vol. 19, No. 4-5 (July-September 2009): 502-510. [DOI](#). This article is © Society for Psychotherapy Research and permission has been granted for this version to appear in [e-Publications@Marquette](#). Society for Psychotherapy Research does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Society for Psychotherapy Research.

# Methodological background of decision rules and feedback tools for outcomes management in psychotherapy

**Wolfgang Lutz**

Department of Psychology, University of Trier, Trier, Germany

**Niklaus Stulz**

Department of Psychiatry, University of Pennsylvania, Philadelphia, PA

**Zoran Martinovich**

Department of Psychiatry and Behavioral Sciences, Northwestern University, Chicago, IL,

**Scott Leon**

Department of Psychology, Loyola University Chicago, Chicago, IL,

**Stephen M. Saunders**

Department of Psychology, Marquette University, Milwaukee, WI

## Abstract

Systems to provide feedback regarding treatment progress have been recognized as a promising method for the early identification of patients at risk for treatment failure in outpatient psychotherapy. The feedback systems presented in this article rely on decision rules to contrast the actual treatment progress of an individual patient and his or her expected treatment response (ETR). Approaches to predict the ETR on the basis of patient intake characteristics and previous treatment progress can be classified into two broad classes: Rationally derived decision rules rely on the judgments of experts, who determine the amount of progress that a patient has to achieve for a given treatment session to be considered "on track." Empirically derived decision rules are based on expected recovery curves derived from statistical models applied to aggregated psychotherapy outcomes data. Examples of each type of decision rule and of feedback systems based on such rules are presented and reviewed.

## Keywords

patient-focused **psychotherapy research**; longitudinal data analysis; outcomes management; treatment feedback; treatment progress; decision making

Patient-focused **psychotherapy research** is concerned with the monitoring, prediction, and evaluation of individual treatment progress during the course of therapy by means of the repeated assessment of outcome variables, the evaluation of these outcome variables through decision rules, and the feedback of this information to therapists and patients (e.g., Howard, Moras, Brill, Martinovich, & Lutz, [13]; Lutz, [24]). Such quality management efforts have been recognized as a promising method to identify patients at risk for treatment failure, to support adaptive treatment planning during the course of treatment, and, as a result, to enhance the likelihood of positive treatment outcomes (Lambert et al., [20]).

The cornerstone of such systems is the identification of whether or not a patient is making sufficient progress. The evaluation of sufficient progress largely depends on the idiosyncratic presentation of the patient. For example, minimal progress by the eighth session might be considered insufficient for most patients. However, for a highly symptomatic patient with high levels of impairment (e.g., multiple problems in the familial and vocational areas), such moderate progress might be considered a harbinger of success. Clearly, not all patients have the same prospect of treatment success, at least not within the same amount of time. As a result, decision rules and feedback systems to support clinical decision making in psychotherapy should not only consider the actual treatment progress of a patient but should also take into account an estimation of the amount of change or improvement that can be expected for a specific patient. By contrasting the actual treatment progress of a patient and his or her expected treatment response (ETR), the sufficiency of progress to date can be put into perspective. This information can then be used to develop decision rules for treatment management and to generate feedback to therapists.

Two distinct approaches have been used to provide feedback to therapists regarding an individual patient's treatment response. These approaches use different versions of decision rules regarding a particular patient's expected level of progress, and they fall into two broad classes (cf. Lambert, Whipple, Bishop, et al., [22]; Lutz, Lambert, et al., [25]). Rationally derived methods are based on predefined clinical judgments about poor progress. Empirically derived methods, in contrast, are based

on statistically derived ETR curves. Examples of such decision rules and feedback systems are presented.

## Rationally Derived Methods

Rationally derived methods of patient-focused research use information gleaned from clinicians or a standardized measure (e.g., the Beck Depression Inventory; Beck, Ward, Mendelson, Mock, & Erbaugh, [4]) to make a pretreatment determination about the patient's clinical status, which serves as a benchmark for his or her expected change. Using rationally derived methods, the patient is judged to have achieved success if he or she achieves an a priori criterion. Thus, many of the rationally derived models have used strategies that are similar to what in the testing and assessment literature would be termed "criterion referenced" (e.g., Barkham et al., [1]; Kordy, Hannover, & Richard, [16]).

A classic example of the rationally derived method of patient-focused research comes from Jacobson and Truax ([14]) and their concept of clinically significant change. In this method, a patient is judged to have made clinically significant change if pretreatment scores on normed measures indicated that he or she was in the patient range of scores and posttreatment scores indicated that he or she was in the nonpatient range of scores (e.g., within the range of scores obtained from a community sample). Clinical significance is a version of rationally derived methodology because the patient is judged (based on a decision rule defined by the clinical significant change concept) to have achieved success if he or she achieves, by the end of treatment, a greater statistical likelihood of outcome scores being representative of a community sample rather than a patient sample.

We present an example of rationally derived methods for evaluating progress, which is more complex than the concept of clinical significant change, especially in terms of the used decision rules and the feedback given to the therapist: The example (1) shows how change on measures during treatment can be evaluated based on a rational derived method and (2) demonstrates how a system of feedback might be implemented in a treatment clinic.

### Example 1: A Method of Rating Treatment Progress

A large-scale study conducted by a German health insurance company (Lutz, Tholen, et al., [33]) illustrates a rationally derived decision rule based on an adaptation of clinically significant change concepts (for details see, e.g., Jacobson & Truax, [14], and later discussion). Feedback to the therapists was based on patients' presentation at intake and on their amount of change by a certain session. To implement the system, patients were first classified into categories according to their initial impairment on a specific instrument (e.g., the Brief Symptom Inventory [BSI]; Derogatis, [8]). Patients categorized as initially highly impaired reported an initial score on the instrument above the mean of a reference sample of psychotherapy outpatients. Initially moderately impaired patients scored below the mean of that reference sample but above a predetermined cut-off score that distinguished scores with a higher probability of having been sampled from a clinical sample than from a nonclinical sample (e.g., Jacobson & Truax, [14]). Finally, initially minimally impaired patients scored below that cut-off score and, hence, had a higher probability of belonging to the nonclinical sample (i.e., they were already in the functional range at the beginning of treatment).

The rating of treatment progress went as follows: For the minimally impaired patients, every amount of positive change on the instrument resulted in a positive evaluation of treatment progress. For

moderately impaired patients, however, progress was rated positive only if improvement exceeded reliable change criterion, that is, if the amount of change surpassed expectations based on the measure's reliability (e.g., Jacobson & Truax, [14]). Last, treatment progress of initially highly impaired patients was viewed as positive only if they both improved reliably and passed the cut-off score (i.e., if their postscore was in the functional range). Statistically reliable deterioration on the outcome instrument led to a negative rating of progress, irrespective of the initial score. All other outcomes resulted in neutral ratings (e.g., if a highly impaired patient improved reliably but did not pass the cut-off score or if a moderately impaired patient improved but did not reach reliable change).

#### Example 2: A Procedure for Providing Feedback

Figure 1 illustrates a possible strategy for providing feedback of patient progress to therapists. In this example, at every measurement occasion, all patients complete three instruments: the BSI, the Inventory for Interpersonal Problems (IIP; Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, [11]), and a third instrument that specifically measures symptoms related to their primary diagnosis. For example, a patient diagnosed with a depressive disorder would complete a depression measure, a patient with a diagnosis of bulimia nervosa would complete a measure of eating disorder symptoms, and so forth.

Graph: Figure 1. (A) Integration of the three repeatedly assessed instruments into a global feedback through a rationally derived decision rule (IIP = Inventory of Interpersonal Problems; BSI = Brief Symptom Inventory). (B) Example for content, operational criteria, and recommendations based on rationally derived decision rules. Note that the disorder-specific instruments are the Beck Anxiety Inventory (Beck, Epstein, Brown, & Steer, [3]) for anxiety disorders, the Eating Disorders Inventory-2 (Garner, [10]) for eating disorders, the Hamburger Zwangsinventar (Klepsch, Zaworka, Hand, Lünenschloss, & Jauering, [15]) for obsessive-compulsive disorders, the Screening Instrument for Somatoform Symptoms (Rief, Hiller, & Heuser, [36]) for somatoform disorders, and the Beck Depression Inventory (Beck et al., 1961) for affective disorders and all other disorders not mentioned previously.

As can be seen in Figure IA, patients are rated on each of the three instruments according to the rules described previously (positive, neutral, or negative). These ratings are then integrated into a global score by summing them. For example, the patient in Figure IA received a positive rating on the BSI, had deteriorated reliably on the IIP, but had not changed (rated "neutral") on the disorder-specific instrument. These three ratings resulted in an overall rating of 0, which corresponds to a neutral global feedback to the therapist at this point in treatment (see Figure IA). The therapist could also be informed of the stability of treatment progress by reporting on progress over several measurements (see Figure IB). For example, the global feedback "stable-positive" might be provided if a patient earns a positive overall feedback in two subsequent periods of assessment. Finally, feedback could be supplemented with additional information about each of the measures comprising the global feedback. For example, therapists could be provided a graphic representation of progress in all three instruments.

In summary, rationally derived methods use theoretically defined pretreatment and change criteria to make determinations about patient progress. Historically, these determinations are made based on distinct, not empirically derived, but intuitively clinically meaningful methods. For example, in the clinical significance example (Jacobson & Truax, [14]), therapy is judged successful if, on completion of

treatment, the patient is more likely to have been sampled from the nontreatment population. Using this global, meaningful criterion provides relevant stakeholders (e.g., therapists, patients, evaluators) with information about the patient's psychological status relative to meaningfully distinct populations of people (e.g., "This patient now functions more similarly to people in the general community than the population of people seeking treatment").

## Empirically Derived Methods

As noted previously, rationally derived methods are similar to criterion-referenced methods often used in psychological assessment. In contrast, empirically derived methods are most similar to norm-referenced methods of psychological assessment, because determination of the patient's expected course of change is based on prior patients with similar intake characteristics. The resulting expectation for change is determined on the basis of a statistical prediction. Further, confidence intervals can be assigned to predicted courses of improvement such that empirically derived methods can estimate how much a patient's actual progress is diverging from expected progress. As with rationally derived methods, a patient's actual treatment progress can be compared with his or her ETR curve to evaluate progress.

Empirically derived ETRs are calculated based on the dose–effect model (Howard, Kopta, Krause, & Orlinsky, [12]), which assumes a linear relationship between the logarithm of the session number (the amount of treatment) and the benefit of psychotherapy (measured by an outcome variable). Negatively accelerated ETR curves can be generated after the intake session on the basis of an individual patient's unique profile of characteristics, using latent growth curve models derived from analyses of large longitudinal data sets. An example of this is shown in Figure 2. Hierarchical linear models (HLMs; e.g., Bryk & Raudenbush, [6]) assume an underlying trajectory of change over time and are able to deal with the nested structure of longitudinal psychotherapy data (i.e., repeated assessments nested within each patient and patients nested within therapists). In a Level 1 model, a repeatedly observed outcome variable is related to time or a time-related variable (e.g., the number of treatment sessions). Individual variation in growth is captured by the fact that growth factors (e.g., slopes in a linear model) are random coefficients that vary across individuals. In a second step of analysis (Level 2 model), the variation in these growth parameters is related to potential predictors (e.g., patient intake characteristics) by treating the growth parameters as dependent variables. With a sufficiently large and relevant database of previous patients, regression weights from a multilevel model can be used to generate ETR trajectories for new patients with characteristics that are similar to those of previous patients in the database.

Graph: Figure 2. Observed and predicted (i.e., expected) treatment course of an imaginary patient. In Session 12, the patient's score falls below the failure boundary around the predicted treatment course, resulting in a warning signal for the therapist that the therapy might not be successful (MHI = Mental Health Index [Sperry et al., [37]]; ETR = expected treatment response).

In an application of empirically derived ETRs, Lutz, Martinovich, and Howard ([29]) analyzed data from 890 psychotherapy outpatients and identified a set of seven indices of status at intake (e.g., patient's expectation of improvement) that allowed prediction of individual change. The resulting measure was called the Mental Health Index (MHI; Sperry, Brill, Howard, & Grissom, 1996). Each patient's expected

MHI change over treatment was modeled as a log-linear function of the session number using information from the seven predictors.

Figure 2 shows the ETR profile and the actual treatment progress of an imaginary patient. In Session 12 the actual treatment progress is much worse than expected according to the intake characteristics. Moreover, it falls below a predefined failure boundary (i.e., outside of a prediction interval around the predicted treatment course and, hence, is significantly different from the ETR). As the number of observed values falling below this failure boundary increases, the probability of treatment failure increases (e.g., Finch, Lambert, & Schaalje, [9]; Lambert, Whipple, Bishop, et al., [22]; Lueger et al., [23]; Lutz et al., [29]). For the imaginary patient, for example, the deviation from the ETR profile in Session 12 might result in a "warning" feedback signal to the therapist and supervisors or other involved clinicians. Based on this, the treatment strategy could be reevaluated and the odds of a positive treatment outcome might be enhanced.

Different approaches to ETR models provide varied kinds of information that may contribute to understanding patient progress and improving treatment strategies. For example, research has supported the application of ETR models to differentiating effects associated with diagnostic groups, particular symptom patterns, and varying outcome instruments (Lutz, Lowry, Kopta, Einstein, & Howard, [28]). Finally, in a recent study, this concept was extended to a three-level model in order to determine the influence of therapist variables on patient change (Lutz, Leon, Martinovich, Lyons, & Stiles, [32]).

There are variations on determining ETRs. The first concerns using similarly characterized subpopulations of patients to predict ETRs for specific patients. The second concerns adjusting ETRs so that a presumption of a certain "shape" of change is not required.

#### Nearest-Neighbors Techniques to Generate ETR Curves

The HLM method referred to in the previous example implies that the predictors identified in a large and heterogeneous sample of patients are predictive for all individuals. That is, it must be assumed that a particular intake characteristic has the same impact on treatment progress in all patients. However, an important intake characteristic for a specific case may be quite uncommon or may have an impact specific to that case. The characteristic might be rare, or the interaction of the characteristic with other patient attributes might be rare. Thus, the global linear effects estimated from HLM models derived from large and diverse samples may not identify relevant effects. The validity of the use of growth curve prediction weights for any particular patient depends on the extent to which the study sample (reference group) is representative of the population of which that patient is a member (Krause, Howard, & Lutz, 1998).

To address this issue and to further refine the prediction of ETR curves, Lutz, Leach, et al. ([26]) introduced an extended growth curve methodology that uses nearest-neighbors (NN) techniques. The NN approach is derived from research in which large databases with many kinds of potentially relevant parameters (e.g., temperature and barometric pressure) recorded on a daily basis are used to make predictions of alpine avalanche risk for a given day (e.g., Brabec & Meister, [5]). By choosing the most similar days based on these climatic parameters and using the relative frequency of avalanches among the NN as a prediction, the probability of alpine avalanche occurrence for a given day is estimated. This

NN methodology was adapted by Lutz, Leach, et al. (2005) to make forecasts in psychotherapy based on the completed courses of the most similar treated patients. In a sample of 203 psychotherapy outpatients seen in the United Kingdom, subsamples of those previously treated patients who most closely match the incoming patient on a series of six demographic and psychometric intake variables (e.g., age, severity of depression) were chosen. Closeness among cases was defined in terms of Euclidean distances between standardized values on these six variables. Hypothesizing that the incoming patient's course will resemble the treatment courses of a subset of similar patients, in a subsequent step of analysis, the slopes of the growth models of the (10, 20, or 50) most similar treated patients were then aggregated, and the resulting mean slope served as prediction of treatment progress for the target patient. Lutz, Leach, et al. (2005) showed that, in the database mentioned previously, the NN method generated more accurate predictions of individual treatment courses than the HLM approach using the same six pretreatment measures as predictors. Therefore, a strategy of basing predictions on small subsamples of cases with similar pretreatment characteristics seems to be superior to a strategy of basing predictions on optimally weighted combinations of pretreatment characteristics.

In a further study, Lutz, Saunders, et al. ([30]) tested the predictive validity and clinical utility of the NN approach in generating optimal ETR profiles for different treatment modalities (cognitive-behavioral therapy [CBT] versus an integrative CBT and interpersonal treatment [IPT] protocol). The NN method created meaningful and different patient-specific predictions between the treatment modalities for about 27% of the patients, even though no significant difference between the two modalities was found. Using a sample of 4,365 outpatients in the United States, Lutz, Lambert, et al. ([25]) further demonstrated the NN technique to be superior to a rationally derived decision rule with respect to the prediction of the probability of treatment success, failure, and duration by using the Outcome Questionnaire (e.g., Lambert, [18]).

In summary, these findings suggest the NN technique to be a viable alternative to predict individual treatment progress and to identify patients at risk for treatment failure. This method might be used in clinical settings either to evaluate the progress of an individual patient in a particular treatment modality or to determine what treatment modality (e.g., CBT or IPT) or treatment setting (e.g., individual or group) is most likely to result in positive change. Furthermore, if used in a team of service providers, the approach could also be used to identify therapists with experience of similar patients (defined by the NN technique) who might provide treatment, consultation, or even supervision for a specific patient.

It is worth noting the implementation barriers related to extensive computational effort and the lack of an easy-to-handle software supporting the NN technique. Implementation of the NN technique in clinical outcomes management would likely require some substantial investment in programming and statistical support services.

#### Growth Mixture Models to Detect Patterns of Change and Variability

The statistical models discussed in the previous sections take account of individual differences in change, but they are built on the assumption of one specific shape of change (e.g., log-linear) for all patients in the data set. Although this assumption makes sense in order to estimate a general trend over time, it has been noted that patient change may follow highly variable temporal courses and that



this variation may be clinically important (e.g., Barkham et al., [1]; Barkham, Stiles, & Shapiro, [2]; Krause et al., [17]).

Growth modeling procedures described previously assume that all individuals are drawn from the same population with common population parameters (i.e., with a common average growth curve), and they capture individual heterogeneity in development over time by individual variation in continuous latent growth factors (e.g., Lutz et al., [29]). Recently introduced growth mixture models (GMMs) relax this single-population assumption and allow for parameter differences across unobserved subpopulations by implementing a categorical latent variable into the latent growth modeling framework (Muthén, [34], [35]). This technique assumes that individuals tend to cluster into distinct subgroups with respect to development over time and, hence, allows estimation of different growth curves within a set of homogenous latent classes or subpopulations. Such GMMs have been used to analyze psychotherapy data deriving from naturalistic studies (Lutz, Stulz, Smart, & Lambert, 2007; Stulz & Lutz, [38]; Stulz, Lutz, Leach, Lucock, & Barkham, [39]) and from randomized controlled trials (RCTs; Lutz, Stulz, & Köck, 2008).

Figure 3 shows an example application of a GMM in a sample of 192 patients drawn from the UK database mentioned previously (Stulz et al., [39]). These patients completed the short-form version of the Clinical Outcome in Routine Evaluation–Outcome Measures (CORE-SF; Cahill et al., [7]). In this example, shapes or patient clusters of early change (up to Session 6) have been identified to predict later outcome and treatment duration. Figure 3A shows five different shapes of early change identified with the GMM, including an "early improvement" cluster of patients (11%), who start with high scores on the CORE-SF but improve rapidly and substantially; a "high impairment" cluster (23.1%) with little or no early change; and a "low impairment" cluster of patients (14.6%), who, like the "high impairment" cluster, demonstrate little or no early change.

Graph: Figure 3. The five different shapes of early change (up to Session 6) identified in a sample of 192 psychotherapy outpatients using growth mixture models (A), and the observed individual growth curves for the continuous group (B) and the discontinuous group (C). (CORESF = Clinical Outcome in Routine Evaluation–Outcome Measures–Short Form).

The two remaining clusters in Figure 3A include two moderately impaired groups with similar average growth curves but very different individual treatment courses. Figure 3B and C shows the plots of the actual individual treatment courses around the average growth curves in these two groups. Figure 3B shows the growth curves of the 27.6% of patients categorized into the "continuous" group, who demonstrated fairly modest session-to-session variation in the early phase of treatment. These can be contrasted to the individual growth curves displayed in Figure 3C. These patients (23.6%) were categorized into the "discontinuous" group because they demonstrated fairly substantial session-to-session variation. Although the average growth curves for these two groups were similar and the two groups did not differ in terms of pretreatment to posttreatment effect sizes ( $d = 0.55$  vs.  $d = 0.54$ , *ns*), comparison of individual growth curves suggests the importance of examining individuals rather than groups. If using the reliable change criterion (Jacobson & Truax, [14]) to evaluate pretreatment to posttreatment change in these two groups, results revealed a higher rate of reliably improved patients in the discontinuous group than in the continuous group (44% vs. 19%), although the treatments were not longer in the discontinuous group ( $M=24.63$  vs.  $M=25.55$  sessions, *ns*). However, conversely, the

rate of reliably deteriorated patients was also higher in the discontinuous patient group relative to the continuous group (13% vs. 0%). Following this, instability during early treatment phases seems to result in higher chances for positive treatment outcomes but also higher risk for negative treatment outcomes. Based on this, on the one hand, therapists may be reassured to have empirical support that discontinuity and instances of significant worsening can be expected over the course of therapy and do not necessarily forecast treatment failure. However, on the other hand, it might also be hypothesized that some other patients with relatively large variation in early session-to-session progress scores are less likely to benefit from treatment, are more likely to need longer treatment to exhibit positive response, or are more likely to require adjunctive (e.g., medications) treatment. Given replication of these findings, future research should aim at identifying variables that differentiate between achievers and failures in the patient group with high instability in early treatment progress. Concerning differentiation between the "continuous" and the "discontinuous" group, higher pretreatment anxiety suggests an increase in the probability of experiencing high instability in early treatment phases (Stulz et al., [39]). This new line of research may lead to more individualized decision rules in terms of specific patterns or shapes of change, especially early in treatment, as well as the identification and meaning of discontinuous treatment courses.

In summary, empirically derived methods use data from large samples of patients to create statistical predictions of an individual clients' treatment trajectory. By using an actuarial method for calculating treatment expectations, the empirically derived approach pegs judgments about the progress of an individual patient's therapy to samples of prior patients with similar intake characteristics. As such, judgments about treatment progress are relatively more normative in the empirically derived approach; that is, the expectations they derive are based on the average response to treatment for prior patients with similar intake characteristics. As a result, patients can be judged as performing above or below expectations for their "norm" group. Therefore, the empirically derived approach is particularly useful for helping to understand how patients are doing in therapy relative to other patients; it is not surprising that the empirically derived method has had success as an approach for providing ongoing feedback to patients and therapists through the course of therapy. Further, by statistically adjusting treatment expectations, the empirically derived approach is also appropriate when an additional goal is to compare therapists or organizations, because these approaches can be extended to provide more specific risk adjustment with respect to patient intake characteristics (Lutz, Leon, et al., [27]).

## Discussion

The repeated assessment of a patient's status during the course of therapy enables immediate feedback of treatment progress to therapists and patients. There are numerous advantages to utilizing these decision rules and evaluation tools to evaluate treatment progress, including alerting clinical staff to nonresponse or negative response, informing therapists about the progress of treatment, and determining when treatment has been sufficient (or is expected to be sufficient). The reviewed methods are at different stages of development, and there are hurdles to implement the monitoring of individual treatment progress (e.g., administrative costs, the need to bring clinicians and administrative staff on board, the need for sufficient expertise to produce timely decision algorithms and graphs like those illustrated). The utility of the techniques, which are based on recently developed statistical

techniques, is substantial, however. For example, these decision rules can be used when data are missing, either by design (e.g., when patient progress is monitored only irregularly during therapy) or by accident (e.g., when a questionnaire is not completed for some of the sessions).

The differences between approaches are fairly stark, and some research suggests that rationally derived and statistical approaches may be similar in their practical impacts. For example, Lambert, Whipple, Bishop, et al. ([22]) compared a rationally derived method to predict patient treatment failure with a statistical growth curve technique based on the HLM method. The results showed that the methods were essentially equivalent, although the statistical approach was somewhat more accurate. Other research indicates that empirically derived methods may be superior (e.g., Lutz, Lambert et al., [25]; Lutz, Saunders, et al., [30]). One difference between the methods is fairly stark, however. Rationally derived methods generate a single criterion by which all patients are measured, which may or may not be desirable. Empirically derived methods, in contrast, generate individually based criterion (expected growth curves) for each patient. Both methods utilize normative samples, but the latter is more idiosyncratic. Thus, it may appeal more to clinicians.

Initial applications of rationally derived rules have already found their way into specific computer programs and into routine care (e.g., Lambert et al., [21]; Lambert, Whipple, Vermeersch, et al., [19]; Lutz, Tholen, et al., [33]; Whipple et al., [40]). In a meta-analysis of feedback studies based on rationally derived decision rules, Lambert et al. ([20]) demonstrated beneficial effects of giving therapists feedback on their patients' progress. A somewhat different type of rationally derived decision rule was presented earlier in this article and is currently under examination in a large-scale study of a German health insurance company.

The class of statistically derived decision rules includes the HLM approach, the NN technique, and GMMs to generate ETR curves. Whereas in most clinical cases treatment progress is evaluated by comparing current scores in an instrument with pretreatment scores, these methods allow for comparisons of the predicted and the actual progress of an individual and have more potential to predict treatment progress with respect to the shapes of change. The more complex NN and GMM approaches are still in development and need optimizing before routine use becomes practical. However, they suggest how the focus on unique patient characteristics may be further improved, by comparing the current patient with an even more tightly specified reference group. For the future, such methods will be helpful in capturing the complexity of individual patient change and variability of change among patients. The experiences and results from multiple implementations of these tools might be also used for revising treatment procedures and manuals and to make them more appropriate for the treatment under routine conditions, e.g., by identifying with the NN approach as described in Lutz, Saunders, et al., [30], specific manuals or guidelines that can be particularly helpful for subsets of patients with similar intake conditions or progress. In that sense, those approaches might, for example, also be used to identify ingredients of interventions that are associated with more or less positive responses to treatment.

All methods presented in this article narrow the gap between research and practice in a real-time manner, with the main aim of improving practice as it happens. These methods provide a practical supplement to enhance treatment decisions that extends beyond the limited information value inherent to RCTs. Because RCTs tend to apply particular manualized treatments to highly selected

samples in experimental settings, they have limited relevance for guiding decisions about the changing treatment strategies in naturalistic settings. Clinical trials evaluate the potential appropriateness of new treatments, but the applications discussed here have potential for impacting clinical practice in ways that are more immediate and particularly tuned to the unique attributes of individual patients. In recent years, clinical research has generated a lot of knowledge about the psychopathology as well as the treatment of specific diagnostic groups, but the search for more knowledge about the heterogeneity of treatment courses as well as how to use this information to improve ongoing treatment has just begun.

## Acknowledgements

This work was partially supported by Swiss National Science Foundation Grants PP001-102651 and 1114-064884.01.

## References

1. [1](#) Barkham, M., Connell, J., Miles, J. N. V., Evans, C., Stiles, W. B., Marginson, F. and Mellor-Clark, J. 2006. Dose-effect relations and responsive regulation of treatment duration: The good enough level. *Journal of Consulting and Clinical Psychology*, 74: 160–167.
2. [2](#) Barkham, M., Stiles, W. B. and Shapiro, D. A. 1993. The shape of change in psychotherapy: Longitudinal assessment of personal problems. *Journal of Consulting and Clinical Psychology*, 61: 667–677.
3. [3](#) Beck, A. T., Epstein, N., Brown, G. and Steer, R. A. 1988. An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 56: 893–897.
4. [4](#) Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. and Erbaugh, J. 1961. An inventory for measuring depression. *Archives of General Psychology*, 4: 53–63.
5. [5](#) Brabec, B. and Meister, R. 2001. A nearest-neighbor model for regional avalanche forecasting. *Annals of Glaciology*, 32: 130–134.
6. [6](#) Bryk, A. S. and Raudenbush, S. W. 1992. *Hierarchical linear models: Applications and data analysis methods*, Newbury Park, CA: Sage.
7. [7](#) Cahill, J., Barkham, M., Stiles, W. B., Twigg, E., Rees, A., Hardy, G. E. and Evans, C. 2006. Convergent validity of the CORE measures with measures of depression for clients in brief cognitive therapy for depression. *Journal of Counseling Psychology*, 53: 253–259.
8. [8](#) Derogatis, L. R. 1993. *Brief Symptom Inventory. Administration, scoring, and procedures manual*, 4th ed., Minneapolis, MN: National Computer Systems.
9. [9](#) Finch, A. E., Lambert, M. J. and 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.
10. Garner, D. M. 1991. *Eating Disorders Inventory-2: Professional manual*, Lutz, FL: Psychological Assessment Resources.
11. Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G. and Villaseñor, V. S. 1988. Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56: 885–892.
12. Howard, K. I., Kopta, M., Krause, M. S. and Orlinsky, D. E. 1986. The dose-effect relationship in psychotherapy. *American Psychologist*, 41: 159–164.

13. Howard, K. I., Moras, K., Brill, P., Martinovich, Z. and Lutz, W.1996. The evaluation of psychotherapy. *American Psychologist*, 52: 1059–1064.
14. Jacobson, N. S. and Truax, P.1991. Clinical significance: A statistical approach to defining meaningful change in **psychotherapy research**. *Journal of Consulting and Clinical Psychology*, 59: 12–19.
15. Klepsch, R., Zaworka, W., Hand, I., Lünenschloss, K. and Jauerling, G.1993. *Hamburger Zwangsinventar–Kurzform (HZI-K)*. Manual, Göttingen, , Germany: Beltz.
16. Kordy, H., Hannover, W. and Richard, M.2001. Computer-assisted feedback-driven quality management for psychotherapy: The Stuttgart-Heidelberg model. *Journal of Consulting and Clinical Psychology*, 69: 173–183.
17. Krause, M. S., Howard, K. I. and Lutz, W.1998. Exploring individual change. *Journal of Consulting and Clinical Psychology*, 66: 838–845.
18. Lambert, M. J.2007. Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. **Psychotherapy Research**, 17: 1–14.
19. Lambert, M. J., Whipple, J. L., Bishop, M. J., Vermeersch, D. A., Gray, G. V. and Finch, A. E.2002. Comparison of empirically-derived and rationally-derived methods for identifying patients at risk for treatment failure. *Clinical Psychology and Psychotherapy*, 9: 149–164.
20. Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L. and 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.
21. Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L. and Hawkins, E. J.2001. The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced?. **Psychotherapy Research**, 11: 49–68.
22. Lambert, M. J., Whipple, J. M., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L. and Goates, M.2002. Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology and Psychotherapy*, 9: 91–103.
23. Lueger, R. J., Howard, K. I., Martinovich, Z., Lutz, W., Anderson, E. E. and Grissom, G.2001. Assessing treatment progress of individual patients using expected treatment response models. *Journal of Consulting and Clinical Psychology*, 69: 150–158.
24. 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.
25. Lutz, W., Lambert, M. J., Harmon, S. C., Tschitsaz, A., Schürch, E. and Stulz, N.2006. The probability of treatment success, failure and duration—What can be learned from empirical data to support decision making in clinical practice?. *Clinical Psychology and Psychotherapy*, 13: 223–232.
26. Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B. Evans, C.2005. Predicting change for individual psychotherapy clients based on their nearest neighbors. *Journal of Consulting and Clinical Psychology*, 73: 904–913.
27. Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S. and Stiles, W. B.2007. Therapist effects in outpatient psychotherapy: A three-level growth curve approach. *Journal of Counseling Psychology*, 54: 32–39.

28. Lutz, W., Lowry, J., Kopta, M., Einstein, A. D. and Howard, K. I. 2001. Prediction of dose-response relations based on patient characteristics. *Journal of Clinical Psychology*, 57: 1–12.
29. Lutz, W., Martinovich, Z. and 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.
30. Lutz, W., Saunders, S. M., Leon, S. C., Martinovich, Z., Kosfelder, J., Schulte, D. 2006. Empirically and clinically useful decision making in psychotherapy: Differential predictions with treatment response models. *Psychological Assessment*, 18: 133–141.
31. Lutz, W., Stulz, N., & Köck, K. in press. Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*.
32. Lutz, W., Stulz, N., Smart, D. W. and Lambert, M. J. 2007. Die Identifikation früher Veränderungsmuster in der ambulanten Psychotherapie [Patterns of early change in outpatient therapy]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 36(2): 93–104.
33. Lutz, W., Tholen, S., Kosfelder, J., Tschitsaz, A., Schürch, E. and Stulz, N. 2005. Evaluation und störungsspezifische Rückmeldung des therapeutischen Fortschritts in der Psychotherapie. [Evaluation and feedback of psychotherapeutic progress in psychotherapy]. *Verhaltenstherapie*, 15: 168–175.
34. Muthén, B. O. 2001. "Second-generation structural equation modeling with a combination of categorical and continuous latent variables". In *New methods for the analysis of change*, Edited by: Collins, L. M. and Sayer, A. G. 291–332. Washington, DC: American Psychological Association.
35. Muthén, B. O. 2004. "Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data". In *Handbook of quantitative methodology for social sciences*, Edited by: Kaplan, D. 345–368. Newbury Park, CA: Sage.
36. Rief, W., Hiller, W., & Heuser, J. 1997. SOMS—Das Screening für somatoforme Störungen (Manual zum Fragebogen) [SOMS—The screening for somatoform disorders (Manual for the questionnaire)]. Bern, , Germany: Huber.
37. Sperry, L., Brill, P. L., Howard, K. I. and Grissom, G. R. 1996. *Treatment outcomes in psychotherapy and psychiatric interventions*, New York: Brunner/Mazel.
38. Stulz, N. and Lutz, W. 2007. Multidimensional patterns of change in outpatient psychotherapy: The phase model revisited. *Journal of Clinical Psychology*, 63(9): 817–833.
39. Stulz, N., Lutz, W., Leach, C., Lucock, M. and Barkham, M. 2007. Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, 75: 864–874.
40. Whipple, J., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L. and Hawkins, E. 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: 59–68.

# Figures

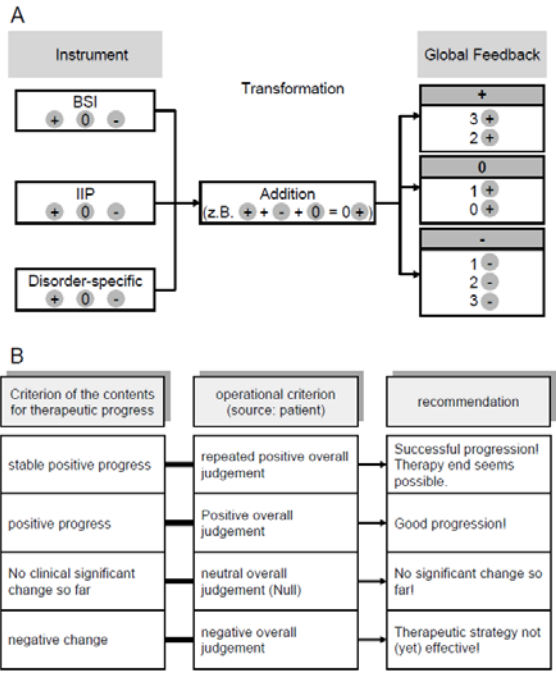


Figure 1. (A) Integration of the three repeatedly assessed instruments into a global feedback through a rationally derived decision rule (IIP=Inventory of Interpersonal Problems; BSI=Brief Symptom Inventory). (B) Example for content, operational criteria, and recommendations based on rationally derived decision rules. Note that the disorder-specific instruments are the Beck Anxiety Inventory (Beck, Epstein, Brown, & Steer, 1988) for anxiety disorders, the Eating Disorders Inventory-2 (Garner, 1991) for eating disorders, the Hamburger Zwangsinventar (Klepsch, Zaworka, Hand, Luñenschloss, & Jauering, 1993) for obsessive-compulsive disorders, the Screening Instrument for Somatoform Symptoms (Rief, Hiller, & Heuser, 1997) for somatoform disorders, and the Beck Depression Inventory (Beck et al., 1961) for affective disorders and all other disorders not mentioned previously.

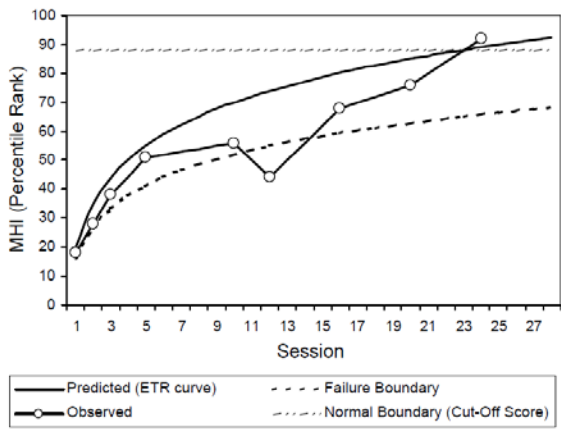


Figure 2. Observed and predicted (i.e., expected) treatment course of an imaginary patient. In Session 12, the patient's score falls below the failure boundary around the predicted treatment course, resulting in a warning signal for the therapist that the therapy might not be successful (MHI=Mental Health Index [Sperry et al., 1996]; ETR=expected treatment response). *Statistical decision rules* 505

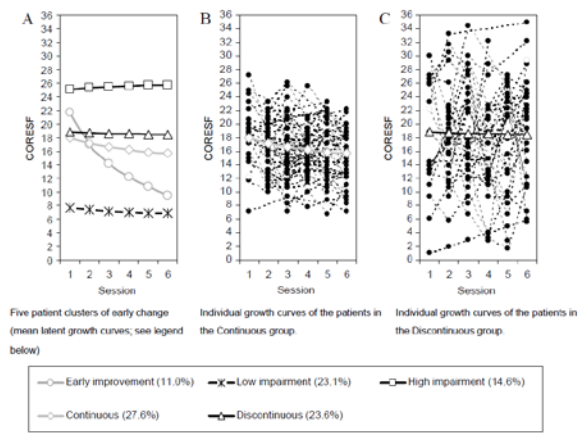


Figure 3. The five different shapes of early change (up to Session 6) identified in a sample of 192 psychotherapy outpatients using growth mixture models (A), and the observed individual growth curves for the continuous group (B) and the discontinuous group (C). (CORESF=Clinical Outcome in Routine Evaluation-Outcome Measures)