6-1-2015

Defining Early Positive Response to Psychotherapy: An Empirical Comparison Between Clinically Significant Change Criteria and Growth Mixture Modeling

Julian Rubel
University of Trier

Wolfgang Lutz
University of Trier

Stephen mark Kopta
University of Evansville

Katharina Köck
University of Koblenz-Landau

Takuya Minami
University of Massachusetts Boston

See next page for additional authors

Accepted version. *Psychological Assessment*, Vol 27, No. 2 (June 2015): 478-488. DOI.

Defining Early Positive Response to Psychotherapy: An Empirical Comparison Between Clinically Significant Change Criteria and Growth Mixture Modeling

Julian Rubel
Department of Clinical Psychology and Psychotherapy, University of Trier, Trier, Germany
Wolfgang Lutz
Department of Clinical Psychology and Psychotherapy, University of Trier, Trier, Germany
Stephen Mark Kopta
Department of Psychology, University of Evansville, Evansville, IN
Katharina Köck
Outpatient Clinic for Psychotherapy, University of Koblenz-Landau Koblenz-Landau, Germany

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

Takuya Minami
College of Education and Human Development, University of Massachusetts Boston,
Boston, MA
Dirk Zimmermann
Department of Clinical Psychology and Psychotherapy, University of Trier,
Trier, Germany
Stephen M. Saunders
Department of Psychology, Marquette University,
Milwaukee, WI

The investigation of patterns of change in psychological treatments has recently emerged as a topic in the research literature. Most treatment concepts and protocols so far have the implicit assumption of a linear or log-linear change course as the common pattern for all patients (e.g., Howard, Kopta, Krause, & Orlinsky, 1986; Lambert, Hansen, & Finch, 2001). However, interindividual differences in change over the course of the treatment might reflect different mechanisms and processes of change (Kazdin, 2007). Furthermore, knowledge about differences in change profiles might enable researchers and clinicians to maximize treatment outcomes for individual patients (Barlow, 2010; Lambert, 2007; Lutz, 2002). Therefore, research on early change is not only related to the debate on the optimal "dosage" of therapy. It is also an important issue related to the growing interest in routine outcome monitoring and practice-oriented research (Castonguay, Barkham, Lutz, & McAleavy, 2013; Lambert, 2013; Newnham & Page, 2010; Shimokawa, Lambert, & Smart, 2010). However, to enable therapists to derive decisions about patients' improvement or nonimprovement from feedback information, rules based on scientific considerations and empirical tests are necessary.

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

Different methods and criteria for the definition of such decision rules have been proposed (e.g., Lambert et al., 2002; Lutz, Stulz, Martinovich, Leon, & Saunders, 2009). These different concepts can be broadly classified into two general classes: (a) those that take information from two time point assessments into account and (b) those that are able to consider information from the whole treatment course.

Decision rules based on only two assessments are relatively simple comparisons between impairment scores on a certain instrument for two time points. These rules define how large the difference between these scores has to be to consider that change improvement or deterioration. These definitions could, for example, rely on a priori–defined expert judgments about good and poor treatment progress. Regularly, these judgments rely on the psychometric properties of an instrument in different reference samples. These properties guide the decision on how much change must have been achieved, given a certain intake score, to consider a treatment successful, unhelpful, or even harmful. An often-applied method of this kind is the concept of *clinically significant change* introduced by Jacobson, Follette, and Revenstorf (1984) and extended Jacobson and Truax (1991). In this approach, to be considered clinically significantly improved from Time Point 1 to Time Point 2, patients' scores on an instrument have to meet two criteria: (a) the scores have to move from a range that is more probable for a sample of clinically impaired patients into a range that is more probable for a nonclinical reference sample, and (b) the difference between the scores has to be statistically significant and, thus, not just a result of imprecise measurements. If only the second criterion is met, an observed improvement is evaluated as reliable (i.e., statistically significantly different from zero) but not *clinically significant*, because the impairment score after the treatment is still highly probable for impaired reference samples. This concept of clinically significant change has great appeal to practitioners, because it can easily be applied in everyday clinical practice.

In comparison, decision rules taking into account the entirety of change course information are, for example, based on statistically derived response predictions based on repeated assessments of already treated patients. With the growing availability of large datasets

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

including repeated measurements over the course of treatment and the growing capacity of computers, sophisticated approaches based on intensive longitudinal methods have been more often developed. Modern statistical tools of growth curve modeling have been applied to generate *expected treatment response* (ETR) curves. These predictions can be compared with the actual change course of a patient (e.g., Finch, Lambert, & Schaalje, 2001; Lutz et al., 2005). On this basis, treatment response patterns can be detected. Specifically, *growth mixture modeling* (GMM) has been demonstrated to be useful for the identification of early change patterns (e.g., Cuijpers, van Lier, van Straten, & Donker, 2005; Lutz et al., 2014; Rubel, Lutz, & Schulte, 2013; Stulz, Lutz, Leach, Lucock, & Barkham, 2007). GMM is a latent variable cluster analytic method. This method allows the categorization of patients into classes with shared treatment response over a defined time period (Nagin & Odgers, 2010).

Both of the just-described methods have been used to identify early change patterns, support therapists in the evaluation of their patients' treatment progress, and guide them to adapt their treatment planning accordingly (e.g., Lutz, Böhnke, Köck, 2011).

Several studies have identified subgroups of clients showing substantial improvements early in treatment. Most of these studies suggest that these fast-responding patients are able to maintain their initial success in that they show markedly positive outcomes (e.g., Haas, Hill, Lambert & Morrell, 2002; Lutz et al., 2014; Lutz, Stulz, & Köck, 2009). Despite the observation of early positively responding patients in different studies, there is no consistent definition of the phenomenon of "early positive response." For example, Stewart et al. (1998) operationalized it as psychopathology being absent or minimal after 2 weeks of treatment. Other studies used a minimum percentage of improvement in the relevant outcome measure to identify early positively responding patients (Hayes et al., 2007; Renaud et al., 1998). Again, others used ETR curves to define early positive change by comparing these predictions with the actual session-to-session ratings of patients' symptomatology (Haas et al., 2002; Leon, Kopta, Howard, & Lutz, 1999). In summary, definitions of early positive response have been dependent on the researchers' divergent judgments on the essential aspects of this construct.

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

Recently, GMM has been repeatedly used for the investigation of patterns of early change in psychotherapy, and it has consistently revealed a pattern of early improving patients (e.g., Lutz et al., 2014). However, GMM is a rather complex statistical method with computationally demanding model-estimating algorithms. Given that, an important question not yet answered is whether GMM-identified early positive responders are a more informative subgroup than those identified with less complex change evaluations (e.g., clinically significant change). The aim of this study was to compare the concept of clinically significant change (Jacobson & Truax, 1991) with a GMM-based approach regarding their shared and distinct characteristics for the identification of early positive treatment response. Consequently, the following research questions were addressed in this study: First, how are the differentially identified early positive response groups related to each other regarding the following variables?: number of patients identified, overlap of subgroups, intake impairment, therapy outcome, and therapy length. Second, how stable are the differentially identified early improvements in the course of the treatment? Third, is the more complex GMM approach more advantageous than simple clinically significant change criteria in terms of specificity and sensitivity for the detection of early positive responders who also show positive treatment outcomes?

Method

Patients

The complete study sample consisted of 5,484 patients treated between June 2006 and December 2011 for at least four sessions in 26 centers comprising 20 college counseling centers, four primary care medical centers, and two private mental health centers. A written informed consent to allow for the anonymous use of their data in research projects was given by clients prior to their first assessment. Patients were treated for different psychological problems, predominantly symptoms of depression and anxiety. The majority of patients were female (61.7%; 3.6% did not report), and all of them were 18 years of age or older. Most of the patients who gave information about their racial background described themselves as European American (40.7%). The further distribution of patients'

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

ethnicity was as follows: Asian American (4.1%), African American (3.7%), Latino/Hispanic (3%), Native American (0.5%), multiracial (0.4%), and other (7.9%; 39.7% did not report). Regarding relationship status, 41.7% indicated that they were single, 16.7% dating, 7.5% married, 1.5% separated, and 0.7% divorced (31.9% did not report).

Most of the patients (3,894; 71%) started treatment with global mental health (GMH) scores in the range of a clinically impaired reference sample with regard to the cutoff criterion *c* described by Jacobson et al. (1984) and Jacobson and Truax (1991).

Therapists and Treatment

Two hundred and forty therapists from different professional backgrounds (including psychologists, psychiatrists, clinical social workers, and trainees) provided the treatments. Therapists were predominantly female (65.8%; 8% did not report) and European American (64.6%; 18.3% did not report). Regarding degrees, most of the therapists had a master's (46.7%) or a doctorate (29.2%; 8.8% did not report). There was no requirement for therapists to follow a manualized treatment protocol. Treatment duration was not fixed to a strict time limit and varied between four and 109 sessions (M = 9.76, SD = 8.25, Mdn = 7.00).

Measures

Prior to each session, the Behavioral Health Measure–20 (BHM-20; Kopta & Lowry, 2002) was administered via a computer-based system, the CelestHealth System-MH (Bryan, Kopta, & Lowes, 2012). The BHM-20 is a 20-item self-report measure consisting of three scales that cover the proposed phases of psychotherapy outcome (Howard, Lueger, Maling, & Martinovich, 1993): well-being (three items), symptoms (13 items), and life functioning (four items). Respondents are asked to rate the items regarding how they have been feeling over the past 2 weeks on a Likert-type scale ranging from 0 (*extreme distress/poor functioning*) to 4 (*no distress/excellent functioning*). A GMH score is calculated by adding the scores for all 20 items and dividing this sum by the number of endorsed items. High scores in the

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

GMH indicate good psychological functioning. The internal consistency reported for GMH in a larger sample from which the present study sample is a subsample was reported as α = .91 (Stulz, Lutz, Kopta, Minami, & Saunders, 2013). A test–retest reliability for a 2-week interval between tests in a college student sample was reported as *r*tt = .80. With regard to discriminant validity, the instrument showed the ability to distinguish clinical from nonclinical groups. Concurrent validity was shown by high correlations between the GMH scale and other established measures, including the Outcome Questionnaire–45 (Lambert & Finch, 1999) and the Symptom Checklist–R–90 (Derogatis & Savitz, 1999), with *r*s = −.81 and −.85, respectively.

Data Analysis

Early positive response. As described earlier, the definition of early positive response varies considerably between studies. Besides the applied methods, the time criterion is also subject to this variation. As a consequence, there is no agreed upon time span that is universally defined as "early" in psychotherapy research. For the present study, we chose the time criterion taking into account clinical and methodological considerations. Obviously, clinicians need to take decisions right from the start of the treatment and continuously throughout its course. It has been repeatedly shown that decisions based on statistical predictions are at least equal to and often better than decisions based solely on clinical judgment (e.g., Grove, Zald, Lebow, Snitz, & Nelson, 2000; Meehl, 1954). Thus, from a clinical perspective, it is important to design decision rules that support clinicians in their decision-making process as early in the treatment as possible.

Methodologically however, GMM as a latent growth model needs at least three scores to model a log-linear trend that was repeatedly reported for individual change curves in the research literature (e.g., Stulz et al., 2013). Consequently, we decided to define the time span until the third assessment (session) as "early." This is the earliest time point that allows for modeling of a log-linear change trend. Application of this rationale resulted in a time criterion that was the same as the one chosen by Haas et al. (2002).

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

GMM. First, the assumption of a log-linear relationship between the amount of treatment and outcome was tested comparing an intercept-only, a linear, and a log-linear latent growth model. A log-linear (i.e., a negatively accelerated) association between number of sessions and change corresponds to the assumptions of the dose-response model (Howard et al., 1986), which is widely used in psychotherapy research (Kopta & Lowry, 2002). In the next step, typical patterns of early change in the GMH scores over the first three sessions were identified using GMM. This method enables the identification of unobserved groups of individuals with shared patterns of change over time in one or more outcome variables (Muthén, 2004). It is based on conventional latent growth models (LGMs) but relaxes (i.e., does not adhere to) the assumption that all individuals in a sample need to be drawn from a single population. Instead, by implementing a categorical latent variable into the LGM framework, GMM allows the identification of subpopulations (latent classes) of individuals that correspond to different shapes of growth curves. In GMM, the mean growth curves for each latent class as well as the individual variations around these growth curves in terms of growth factor variances are estimated. In this current application, a model was chosen for which variances around the class-specific slopes were fixed to zero within classes, whereas intercept variances were freely estimated but constrained to be constant between classes. Consequently, all differences in change over time had to be captured completely by the differences in mean slopes of different latent classes. This model was stable and emphasized the identification of heterogeneity in change over time.

In this study, GMMs were estimated using the Mplus software (Version 6.0; Muthén & Muthén, 2010). Mplus uses maximum likelihood estimates as well as an accelerated expectation maximization procedure and allows for the estimation of models with missing values in continuous outcome variables.

Prior research applying GMM to session-by-session psychotherapy data has repeatedly identified a subgroup of patients who start treatment highly impaired and improve in the first few sessions. Patients showing such a pattern are, in the following, referred to as *GMM—early positive change*.

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

Clinically significant change criteria. Patient change was additionally assessed using the concept of clinically significant change (Jacobson & Truax, 1991). This concept is composed of two conditions. The first condition to consider the change of a patient clinically significant is *reliable improvement.* A patient changed reliably (i.e., statistically significantly; $p < .05$) if the difference between the two scores is larger than the reliable change index (RCI) of the instrument. The second condition is the movement of the scores from the range that is more likely for a clinical reference sample into the range that is more likely for a nonclinical reference sample (crossed cutoff). For the comparison with the GMM-based approach, reliable improvement and clinically significant improvement are investigated as two separate methods. On the basis of their GMH scores from the first to the third session, patients were categorized in one of two groups: (a) *clinically significant improvement*, with the GMH score moving from a score below 2.92 (cutoff) before the first session to a score above 2.92 before the third session and the difference between these two scores being larger than 0.39 points (RCI), or (b) *reliable improvement*, with the difference between the first score and the third score being larger than 0.39 points but the cutoff value of 2.92 not being crossed.

For the evaluation of treatment outcome, the difference between the first and the last score is assessed using the same criteria. Two additional groups for the description of negative treatment outcomes were defined: (c) *no change*, with the difference between the first score and the last score being smaller than 0.39 points and (d) *deterioration*, with the difference between the first score and the last score being larger than 0.39 points but in the negative direction.

Results

Reliable and Clinically Significant Improvement

At Session 3, 1,918 (35.0%) out of the 5,484 patients met the criterion of reliable improvement. Eight hundred and ninety-two patients (16.3%) had achieved clinically significant improvement until Session 3, whereas 3,035 patients (55.3%) showed no statistically

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

reliable change from the first to the third session, and 531 (9.7%) had deteriorated until Session 3.

Early Change Patterns (GMM)

The Bayes information criterion (BIC; Schwartz, 1978) indicated the best fit for a log-linear model: intercept-only Model 111,923.96, linear Model 110,806.38, and log-linear Model 110,733.71. Accordingly, the subsequent growth mixture analyses assumed a log-linear relationship between the number of treatment sessions and outcome.

In the following analyses, the number of distinct patterns of early change was determined by means of GMM (Muthén, 2004). Starting with one latent class (i.e., with a conventional LGM), additional classes were entered into the GMM until the optimal number of latent classes was found. The decision on the number of latent classes was based on joint consideration of two typically applied indices. The BIC (Schwartz, 1978) steadily decreased from the one- through the seven-class solutions (21,742.17; 21,532.58; 21,318.14; 21,220.09; 21,116.84; 21,069.74; and 21,052.47), indicating a model with at least seven classes having the best fit. In comparison, the Lo–Mendell–Rubin likelihood ratio test of model fit (Lo, Mendell, & Rubin, 2001) showed that already the addition of a fifth class did not result in a significant improvement of model fit (three classes vs. four classes: $p < .01$; four classes vs. five classes: $p = .08$). Consequently, a model with four classes (see Figure 1) was considered the best solution and used for further analyses.

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

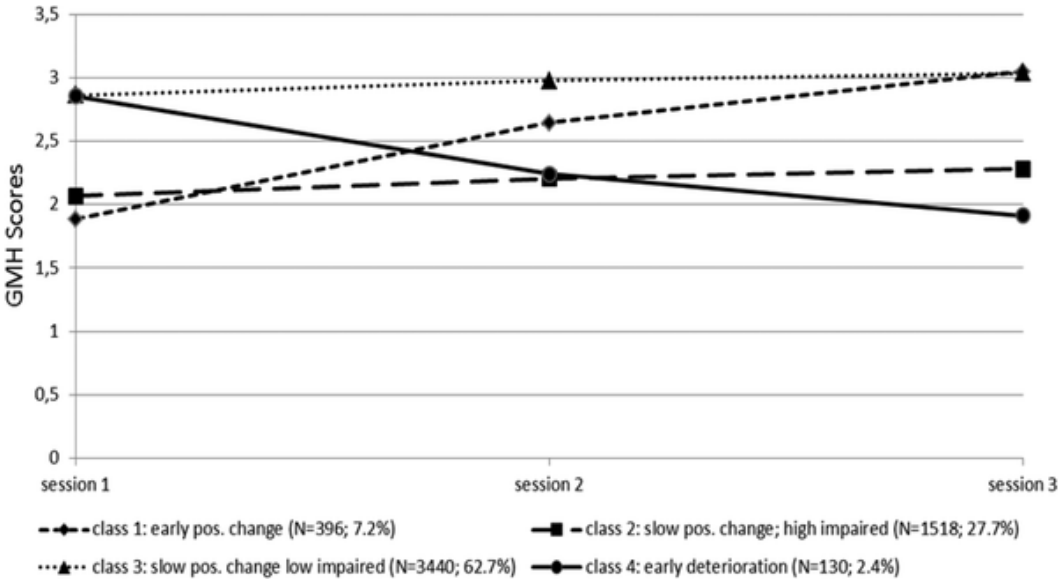

Figure 1. Estimated mean change trajectories over the first three sessions for a four-class growth mixture model solution.

The first subgroup comprised of 396 patients (7.2%) who started treatment with a relatively high average impairment (intake GMH score: *M* = 1.80, *SD* = 0.41) and improved relatively quickly until Session 3. Patients categorized in this group showed early positive response according to the GMM approach and constitute the GMM—early positive change group, as specified earlier. The second subgroup comprised of 1,518 patients (27.7%) who also started treatment relatively highly impaired (intake GMH score: *M* = 1.92, *SD* = 0.35) but improved relatively slowly until the third session. Both of these first two subgroups started treatment substantively more impaired than an average patient from a counseling (*M* = 2.68) and outpatient psychotherapy (*M* = 2.33) reference sample (Kopta & Lowry, 2002). The third subgroup was by far the largest, comprising 3,440 patients (62.7%). This class included patients who started with a relatively low initial impairment (intake GMH score: *M* = 2.89, *SD* = 0.42) and showed rather slow improvement until Session 3. The fourth subgroup comprised of a small number of patients deteriorating during the first three assessments (*n* = 130; 2.4%). The mean intake GMH score of the fourth subgroup was 2.92 (*SD* = 0.43). Comparing the initial impairment of Subgroups 3 and 4 with counseling and psychotherapy reference samples reveals that these subgroups started with comparatively low levels of impairment (Kopta & Lowry, 2002).

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

Overlap Among the Three Definitions of Early Positive Response

In a next step, the overlap and uniqueness of the differentially identified early positive response groups were investigated. The overall numbers and the overlap between the three groups, with percentages given in reference to each overall number, are displayed in Table 1.

Table 1
Numbers of Patients in the Differentially Identified Early Positive-Response Groups and Their Overlaps at Session 3

	Status at Session 3		
Status at Session 3	Reliable improvement	Clinically significant improvement	GMM—early positive change
Reliable improvement	1,918	892 (100%)	396 (100%)
Clinically significant improvement	892 (47%)	892	253 (64%)
GMM—early positive change	396 (21%)	253 (28%)	396

Note. Column percentages are shown in parentheses with reference to the main diagonal value of the respective column. GMM = growth mixture modeling.

Numbers of Patients in the Differentially Identified Early Positive-Response Groups and Their Overlaps at Session 3

Overall, the GMM approach identified many fewer patients as early positive responders than did the reliable improvement (about five times fewer) and clinically significant improvement (about two times fewer) criteria. However, considering the different group sizes, the three groups were largely overlapping (see Table 1). All patients in the GMM—early positive change group also improved reliably from intake to Session 3 (N = 396; 100%). Clinically significantly improved patients were 253 (64%) of these GMM—early positive change patients. Because of the overall group size differences, these numbers correspond to only 21% of reliably improved patients who were also

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

identified via the GMM approach and to 28% of clinically significantly improved patients.

Relations to Treatment Length, Intake Impairment, and Treatment Outcome

The three groups of early positively responding patients identified via different methods were compared with regard to treatment length, intake impairment, and treatment outcome. In terms of number of sessions in treatment, the three groups did not differ significantly (see Figure 2) from each other (GMM—early positive change: *M* = 8.32, *SE* = 0.41; reliably improved: *M* = 8.93, *SE* = 0.37; clinically significantly improved: *M* = 8.57, *SE* = 0.44). With regard to initial impairment, patients with early positive response identified via GMM (*M* = 1.79, *SE* = 0.03) started with lower GMH scores (indicating higher impairment) than early improving patients identified with the two other methods (reliable improvement: *M* = 2.10, *SE* = 0.03; clinically significant improvement: *M* = 2.38, *SE* = 0.03). The GMM—early positive change group also showed by far the highest pre- minus posttreatment differences (high values indicating large positive changes from pretreatment to posttreatment) in GMH scores (*M* = 1.28, *SE* = 0.03; reliably improved: *M* = 0.93, *SE* = 0.03; clinically significant improved: *M* = 0.85, *SE* = 0.04).

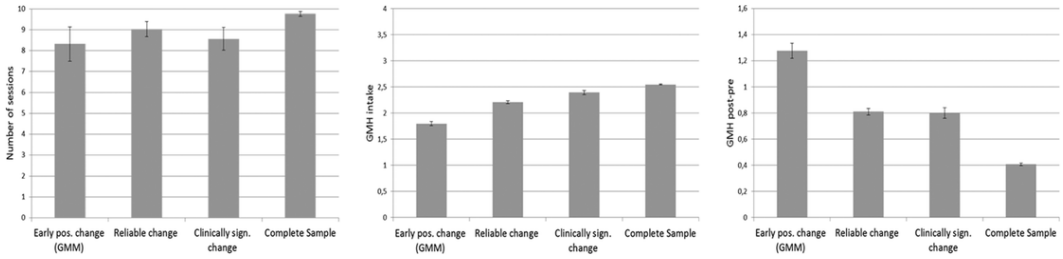

Figure 2. Mean numbers of sessions, mean intake general mental health (GMH) scores, mean differences between pre- and posttreatment GMH scores (high values indicating high positive changes from pretreatment to posttreatment), and 95% confidence intervals for each of the early positive response groups and the complete sample.

A more fine-grained examination of the relations among treatment outcome, early change, and therapy length is depicted in Table 2, which shows, the pre–post effect sizes (*d*s) and categorized change statuses after treatment (reliably improved, clinically

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

significantly improved, no change, and deterioration) for the three early positive change groups, depending on the number of sessions attended and in total. Irrespective of the number of sessions attended, the GMM—early positive change group showed the highest pre–post effect sizes (*d*s = 1.88–2.16) as well as the highest shares of reliably improved patients after the treatment (90%–93%). In comparison, the groups of patients identified via clinically significant change methods both showed smaller yet also high effects sizes (both between about 1.15 and 1.36) and shares of reliably improved patients at the end of the treatment (both between 74% and 82%). Regarding clinically significant change after the treatment, the GMM—early positive change group and the group of patients who had improved clinically significantly at Session 3 showed similar shares (both in the 65%–73% range). In comparison, a little less of the early reliably improved group achieved clinically significant change until the end of the treatment (51%–53%). The numbers of patients who showed no change or deterioration from pre- to post-treatment were slightly smaller in the GMM—early positive change group than in the groups defined via reliable and clinically significant change criteria.

Table 2
*Frequencies and Final Treatment Outcomes (*ds *and Categories) for All Patients and for Those Meeting the Respective Early Positive Response Criteria (Status After Session 3) Depending on Treatment Length and in Total*

Group and status after Session 3	Number of sessions				Total
	<7	7–12	13–20	>20	
RI (>0.38)					
N	1,095	464	232	127	1,918
d	1.27	1.33	1.33	1.33	1.29
RI	876 (80%)	359 (77%)	182 (78%)	101 (80%)	1,518 (79.1%)
CSI	556 (51%)	239 (52%)	123 (53%)	67 (53%)	985 (51.14)
NC	193 (18%)	89 (19%)	42 (18%)	23 (18%)	347 (18.1%)
Det.	26 (2%)	16 (3%)	8 (3%)	3 (2%)	53 (2.8%)
CSI (>0.38)					
N	538	205	96	53	892
d	1.31	1.22	1.30	1.15	1.28
RI	439 (82%)	155 (76%)	73 (76%)	39 (74%)	706 (79.1%)
CSI	395 (73%)	135 (66%)	67 (70%)	38 (72%)	38 (72%)
NC	86 (16%)	43 (21%)	21 (22%)	13 (25%)	163 (18.3%)
Det.	13 (2%)	7 (3%)	2 (2%)	1 (2%)	23 (2.6%)
GMM—early positive change					
N	242	98	36	20	396
d	2.00	2.12	2.16	1.88	2.04
RI	225 (93%)	91 (93%)	91 (93%)	18 (90%)	367 (92.7%)
CSI	157 (65%)	63 (64%)	26 (72%)	12 (60%)	258 (65.2%)
NC	15 (6%)	7 (7%)	3 (8%)	2 (10%)	27 (6.8%)
Det.	2 (1%)	0 (0%)	0 (0%)	0 (0%)	2 (0.5%)
All patients					
N	2,987	1,324	712	461	5,484
d	0.62	0.71	0.69	0.63	0.65
RI	1,095 (37%)	464 (35%)	232 (33%)	127 (28%)	1,918 (35%)
NCI	538 (18%)	205 (16%)	96 (14%)	53 (12%)	892 (16%)
NC	1,616 (54%)	722 (55%)	404 (57%)	293 (64%)	3,035 (55%)
Det.	276 (9%)	138 (10%)	75 (11%)	41 (9%)	531 (10%)

Note. RI = reliably improved; CSI = clinically significantly improved; NC = no change; Det. = deterioration; GMM = growth mixture modeling.

Frequencies and Final Treatment Outcomes (ds and Categories) for All Patients and for

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

Those Meeting the Respective Early Positive Response Criteria (Status After Session 3) Depending on Treatment Length and in Total

Compared with the effect sizes for each of the three early positive response groups, the average effect sizes for all patients in the sample were consistently smaller (between 0.62 and 0.75). On average, effect sizes for all patients were about half as high as those of the groups defined with clinically significant change criteria and one-third as high as the GMM-defined group. Accordingly, although the rates of reliably and clinically significantly improved patients at the end of the treatment were much lower (between 28% and 35% and between 12% and 18%, respectively) the rates of patients showing no change or deterioration over the course of the treatment were much higher (between 55% and 64% and between 9% and 11%, respectively).

To evaluate the predictive power of the different approaches for final treatment status, specificity and sensitivity values were calculated, and these are presented in Table 3. Although the GMM—early positive change group showed the highest specificities for predicting positive reliable change (0.989) and clinically significant change (0.964) from pre- to posttreatment, its sensitivities were the lowest for both outcome criteria (.135 for reliable and .157 for clinically significant improvement). Similarly, high specificity values for the prediction of reliable and clinically significant improvement were found for the early positive responders classified via clinically significant change criteria (.933 for both reliable and clinically significant improvement). Sensitivity values for this subgroup were higher but still low (.260 for reliable and .386 for clinically significant change). The highest sensitivity values were obtained for the reliable early improvement criterion (.559 for reliable .599 for clinically significant improvement). Conversely, specificity values were the lowest for this subgroup of early positive responders identified via reliable change (.856 for reliable and .757 for clinically significant improvement).

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

Table 3
Specificity and Sensitivity Values of the Three Classification Methods for the Prediction of Positive Reliable Change and Clinically Significant Change After Treatment

Early response classification method	Status at end of treatment: Positive reliable change	Status at end of treatment: Clinically significant change
Clinically significant improvement		
Specificity	.933	.933
Sensitivity	.260	.386
Reliable improvement		
Specificity	.856	.757
Sensitivity	.559	.599
GMM—early positive change		
Specificity	.989	.964
Sensitivity	.135	.157

Note. Specificity denotes the proportion of patients not reliably and not clinically significantly improved after treatment and not classified as early positive responders from all the patients not reliably/clinically significantly improved after the treatment. Specificity denotes the proportion of patients reliably and clinically significantly improved after the treatment and classified as early positive responders from all the patients reliably/clinically significantly improved after the treatment. GMM = growth mixture modeling.

Specificity and Sensitivity Values of the Three Classification Methods for the Prediction of Positive Reliable Change and Clinically Significant Change After Treatment

Stability of early improvements given the differential definition methods is illustrated in Figure 3, which shows the percentages of reliably improved patients after each of Sessions 4 through 13 and at the end of the treatment. Independent of session number, the rate of reliably improved patients was consistently highest in the GMM—early positive change group (about 90%). Only slight fluctuations could be observed over the course of the first 13 sessions. The rates for the two early improving groups defined with the clinically significant change criteria were similar to each other and consistently smaller than those for the GMM-defined group.

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

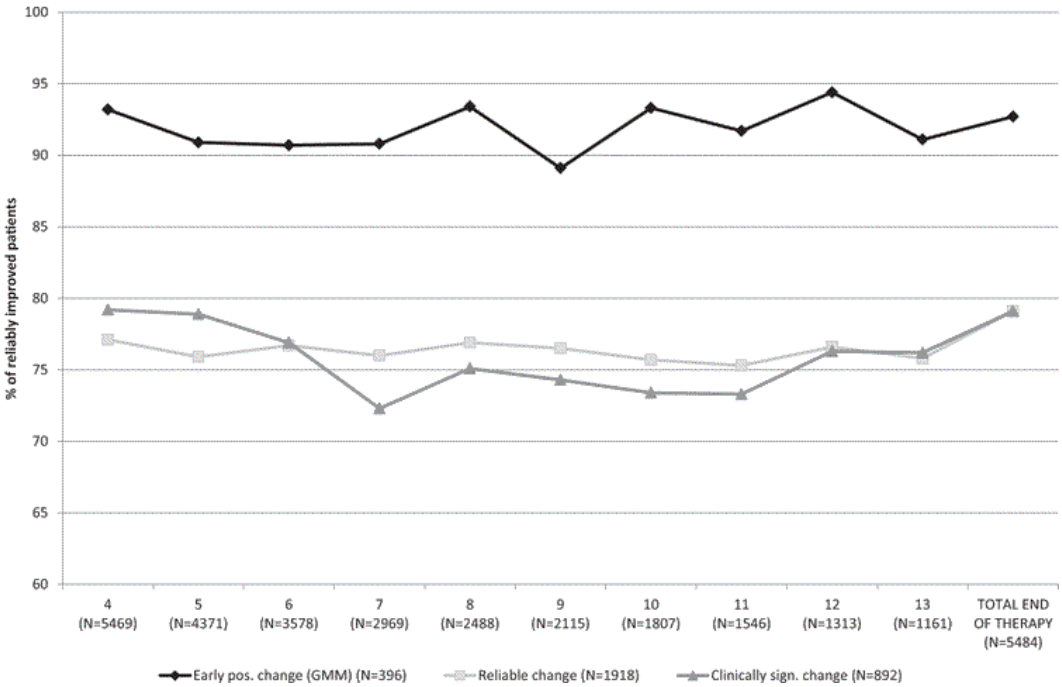

Figure 3. Percentages of reliably improved patients after Sessions 4–13 and after the end of the treatment for 5-patient subgroups defined on the basis of their change status after Session 3.

Discussion

In this study, three methods for the identification of early positive response to psychotherapy were compared with regard to overlap and uniqueness of the identified subgroups and their specific characteristics and predictive qualities. A GMM-based approach was compared with two methods from the concept of clinically significant change. Given the methodological definitions of the clinically significant change methods and GMM, there are some general differences, which can be deduced on a theoretical basis: Whereas for the clinically significant change methods, an a priori fixed amount of change is minimally required to meet one of the criteria (RCI), GMM is more flexible in this regard. How much change is needed to be identified by the GMM approach depends on the nature of the change courses within the whole patient sample and all of the available change course information. GMM is also more flexible with regard to intake and end state functioning. To be categorized as clinically significantly improved, a patient's score has to move from the range above an a priori defined cutoff score into the range below that cutoff score. Consequently,

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

patients who do not start the treatment within the range above the cutoff score can never improve "clinically significantly." As for the GMM approach, there are no such cutoff scores. Given that, theoretically every patient can be categorized as belonging to the improved group. Another important difference is the fact that the GMM approach takes into account the complete change course until a certain time point. Clinically significant change criteria, conversely, solely rely on the comparison of change from one time point to another.

The aim of this study was to compare these three methods for the identification of early positive response to psychotherapy on an empirical basis. For this purpose, these methods were applied to the first three scores of patients in a big naturalistic outpatient psychotherapy sample. The results of the comparison of the three methods provide evidence that the different identification methods have very specific characteristics when defining similar patients as early positive responders. In fact, all of the early positive responders identified via GMM were also detected by the reliable improvement method. Given that, the GMM—early positive change group was a subgroup of the patients reaching positive reliable change until Session 3. However, GMM categorized about five (positive reliable change) and two (clinically significant change) times fewer patients as early positive responders than did the other methods. Consequently, the GMM approach is more conservative in its identification of early positively changing patients than are clinically significant change methods.

Further, it could be shown that the GMM—early positive change group was characterized by higher average intake impairments and larger average pre- to posttreatment changes than the groups identified via clinically significant change criteria. As high intake scores are regularly connected to higher pre- to posttreatment changes, these results suggest that the difference between the early positive responders identified with the GMM approach and those defined via reliable change until Session 3 was mainly attributable to high intake values. As a consequence, one could argue that the GMM model is unnecessary if the amount of change from intake to Session 3 and the intake score are known. To test this hypothesis, a binomial logistic regression analysis was conducted. Being classified as an early positive responder with the GMM method (yes = 1, no = 0) was used as categorical dependent variable; changing reliably positively until

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

Session 3 (yes = 1, no = 0) and the pretreatment GMH score were used as predictor variables in the regression analysis. Only 78 (19.7%) of the 396 early change patients identified via GMM were correctly predicted by the logistic regression model using these predictor variables. Given that, GMM-identified early positive responders were not just a subgroup of reliably improved patients with very low intake scores (high intake impairment). Thus, the application of GMM for the identification of early positively responding patients supplies additional information that cannot be deduced alone from the intake score and the amount of change until Session 3. This might be due to the fact that GMM does not use only the information from two time points (Session 1 and Session 3). Because GMM takes each of the repeated assessments of individual change curves into account, this definition generally requires a more stable positive response pattern than do the clinically significant change criteria. There might be many patients starting with high initial impairment and changing reliably or even clinically significantly from the first to the third session but not meeting the GMM criteria because the score in the second session was not positive enough. This aspect is more pronounced the more assessments that are considered. In the case of the present study, in which only three assessments were taken into account also, rather instable change courses could result in an average early response pattern if the gain from the second to the third session was big enough.

With respect to outcome prediction, which is the basis for the formulation of decision rules, it could be shown that both the GMM approach and the computationally less demanding clinically significant change methods had their positive and negative aspects. Being identified as an early positive responder by the GMM approach was a highly reliable prognostic factor for being reliably improved after the treatment. However, this method showed itself to be very insensitive. As a consequence, many patients who improved reliably or clinically significantly from pre- to posttreatment would have been missed if only GMM had been applied.

Given their ease of use, it comes as somewhat of a surprise that clinically significant change criteria showed such a good performance in predicting ultimate treatment outcome. While being only slightly less specific than GMM in the prediction of treatment success, the

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

reliable improvement method in particular proved to be much more sensitive than the more complex GMM approach.

Given that, decision rules should not solely rely on GMM. Rather GMM-based approaches should be complemented by more sensitive reliable and clinically significant change methods. In practice, such an integrated approach could be implemented in feedback software tools by the means of a stepwise system with different probability estimates for positive outcomes depending on the method that classified a patient as an early positive responder.

However, one of the limitations of the present study concerns the results of the comparison of the methods regarding their predictive qualities for treatment outcome. One of the three compared methods is also used to assess treatment outcome. We chose the clinically significant change criteria for the evaluation of treatment outcome (see Tables 2 and 3) because they are widely used methods in clinical research and practice (cf. Ronk, Korman, Hooke, & Page, 2013). It should be noted that the predictive power of a method is regularly relatively high if it is used to define a state at two time points and the latter state is predicted from the first state. Compared with that, the predictive power is lower when two different methods are used to define the states at the two respective time points. Accordingly, because the reliable and clinically significant change criteria are more similar to each other than to the GMM approach, the present results might be biased to the disadvantage of GMM. Future investigations should consider evaluating the different methods by using a different instrument for the evaluation of treatment outcome than the one used here for the assessment of early positive change.

In addition, the generalizability of these results is reduced because only patients with at least four sessions were included in the analysis. Given that, the present results are only valid for patients who do not drop out before the fourth session. However, previous studies have shown that some patients experience substantial improvements in the first or first two sessions (Haas et al., 2002). Thus, there might be some early improving patients who were excluded from the current analysis because of a too early termination of the treatment.

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

Another shortcoming of the present study regards the definition of *early*, which is always a matter of debate and is related to theoretical orientations, national health care policies, and the actual number of sessions attended by each individual patient. It follows from that that, for patients being provided with 300 sessions of therapy, the early phase might rather be the first 30 sessions instead of the first three. But for patients who were provided with four sessions, the first three also cannot be doubtlessly defined as "early." Owing to these considerable differences, it simply would not be possible to define an early treatment phase that would be appropriate from all perspectives. Consequently, this definition has to be done on grounds of the specific characteristics of the investigated patient sample. In the current investigation, we decided to define as early the shortest possible time span that still enabled us to estimate a log-linear change trend with the GMM approach. Although Haas et al. (2002) chose the same interval, compared with most other investigations of early response, the first three assessments represents a rather short phase. In addition to the just-stated rationale, several other reasons support our decision to reduce the time span to this minimum. First, the treatments in this sample were rather short (M = 9.76 sessions). Thus, our early phase definition already covered, on average, about one-third of the complete treatment. In addition, the number of patients that could be taken into account was at its maximum when the required number of sessions was minimal. Thus, this approach enabled us to derive predictions for about 20% more patients than we could have if we had extended the early phase to Session 4 and 34% more patients than we could have if we had extended the early phase to Session 5. However, utility for clinical practice was the most important argument for choosing the shortest possible phase. Decision rules are designed to assist clinicians in their decision making. Therefore, it should be the aim of researchers to design decision rules so that they can be validly applied as early in treatment as possible.

It must also be admitted that a potential alternative explanation of early positive response in psychotherapy outcome studies is regression to the mean. Statistically, patients who start treatment rather highly impaired have more room to improve in their scores than do patients who start with relatively low impairment. For the present sample, this is also reflected in the significant negative correlation between the initial score and the change score from pre- to

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

posttreatment ($r = -.53$, $p = <.00$). In such cases, when the correlation between initial scores and amount of change is negative, the occurrence of regression to the mean is likely (Rogosa, Brandt, & Zimowski, 1982; Speer, 1992). The common clinically significant change concept introduced by Jacobson and Truax (1991), which was applied in the current study, does not take regression to the mean into account. Speer revisited the concept and presented a method that considers regression to the mean as being more conservative for more impaired patients (more distant from the mean). Therefore, all early change classes were additionally checked with this more conservative method proposed by Speer. All of the patients who were defined as early positive responder by the Jacobson and Truax method or by the GMM method also improved statistically significant ($p < .05$) according to the Speer method. Thus, it is unlikely that regression to the mean was the only factor that led to early positive improvements.

Despite these limitations, the current study may have potential implications for future research, health care services, and clinical practice. Considering the results of the current study, future research on early response might be better able to anticipate the implications connected with the different methods. For the evaluation of correlations between early response and treatment outcome, it is of central importance to know which methods were applied for the definition of early positive response and how specific and sensitive they are. However, replications in other samples, settings, and countries as well as with different instruments are needed to validate and generalize our results. Given the high rates of patients from the early positive response groups who showed positive ultimate treatment outcomes, psychometric progress monitoring and feedback seem to be important tools for health services to optimize the allocation of resources (i.e., treatment sessions). Patients who show positive response at such an early stage of the treatment might need fewer sessions than patients who need longer to show positive response (cf. Lambert, 2007). However, to deduce concrete suggestions for health care services, controlled clinical trials with follow-up assessments would be necessary to test the hypothesis that patients who improve early need fewer sessions to achieve stable positive outcomes than do more slowly improving patients. Regarding the design of feedback software systems, results suggest a combination of the different approaches. Whereas early positive

Psychological Assessment, Vol 27, No. 2 (June 2015): pg. 478-488. DOI.

responders identified via clinically significant change criteria had very high chances of a good treatment outcome, additional *GMM*-based information could supply additional assurance to therapists.

An important message for practitioners who will not or cannot use sophisticated feedback software is the very good performance of the clinically significant change criteria for the prediction of ultimate treatment outcome. Given the high predictive qualities of these easy-to-apply methods, the RCI and the cutoff score of an instrument should be mandatory information in every test handbook. Being provided with this information enables every therapist who tracks his or her patients' progress session by session to evaluate the chances for positive treatment outcome. Using the instrument from the present study in a similar sample, a therapist could also directly apply the findings from the present study. Therapists know, for example, that if one of their patients improves reliably until Session 3, the probability for this patient to be reliably or clinically significantly improved at the end of the treatment is more than doubled (from 33.6% to 79.1% for reliable and from 18.5% to 51.4% for clinically significant change).

Taken together, the findings of the present study illustrate the specific characteristics of three widely used approaches for the identification of early positive response in a large sample of psychotherapy outpatients. The findings underline not only the additional value provided by the computationally demanding GMM approach but also the surprisingly good validity of predictions that can be deduced on the grounds of simple clinically significant change criteria. For routine outcome monitoring and feedback systems, the results suggest that a combination of decision rules, a GMM-based approach, and clinically significant change methods might be a fruitful combination.

Footnotes

[1] Criterion *c* defines the cutoff point as the point that lies halfway between the mean of a functional and a dysfunctional population if variances are equal. Considering the means and standard deviations reported for the GMH score of the Behavioral Health Measure–20 in Kopta and Lowry (2002), *cutoff*GMH is calculated as follows (Jacobson et al., 1984; Jacobson & Truax, 1991):

$$cutoff_{GMH} = \frac{s_0 * M_1 + s_1 * M_0}{s_0 + s_1} = \frac{0.47 * 2.33 + 0.68 * 3.32}{0.51 + 0.62}$$

$$= 2.92,$$

where M0/s0 and M1/s1 are the mean/standard deviations of a community adult reference sample and a sample of psychotherapy outpatients, respectively. This criterion resulted in a *cutoff*GMH score of 2.92. Thus, patients with a GMH score below 2.92 are more likely ($p < .05$) to belong to a clinical population than to a nonclinical population.

[2] The RCI is calculated using the following formula (Jacobson & Truax, 1991): $RCI = 1.96 * \sqrt{2 * (SD * \sqrt{1 - r})^2}$, where *SD* is the standard deviation of the GMH score in a community adult sample (Kopta & Lowry, 2002), and *r* is the reliability (internal consistency; α = .91) of the instrument in a similar sample (Stulz et al., 2013). Internal consistency, instead of test–retest reliability, is used to calculate the RCI. Internal consistency has been recommended for clinical samples because test–retest reliabilities are likely to be deflated by real individual differences in treatment response and phenomena like spontaneous remission (Martinovich, Saunders, & Howard, 1996).

References

Barlow, D. H. (2010). Negative effects from psychological treatments: A perspective. *American Psychologist*, *65*, 13–20. 10.1037/a0015643

Bryan, C. J., Kopta, S. M., & Lowes, B. D. (2012). The CelestHealth System: A new horizon for mental health treatment. *Integrating Science and Practice*, *2*, 7–11.

Castonguay, L. C., Barkham, M., Lutz, W., & McAleavy, A. (2013). Practice-oriented research: Approaches and applications. In M. J.Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 85–133). New York: Wiley.

Cuijpers, P., van Lier, P. A. C., van Straten, A., & Donker, M. (2005). Examining differential effects of psychological treatment of depressive disorder: An application of trajectory analyses. *Journal of Affective Disorders*, *89*, 137–146. 10.1016/j.jad.2005.09.001

Derogatis, L. R., & Savitz, K. L. (1999). The SCL-90-R, Brief Symptom Inventory, and Matching Clinical Rating Scales. In M. E.Maruish (Ed.),

The use of psychological testing for treatment planning and outcomes assessment (2nd ed., pp. 679–724). Mahwah, NJ: Erlbaum.

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

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19–30. 10.1037/1040-3590.12.1.19

Haas, E., Hill, R. D., Lambert, M. J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains?*Journal of Clinical Psychology*, *58*, 1157–1172. 10.1002/jclp.10044

Hayes, A. M., Feldman, G. C., Beevers, C. G., Laurenceau, J.-P., Cardaciotto, L., & Lewis-Smith, J. (2007). Discontinuities and cognitive changes in an exposure-based cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, *75*, 409–421. 10.1037/0022-006X.75.3.409

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose–effect relationship in psychotherapy. *American Psychologist*, *41*, 159–164. 10.1037/0003-066X.41.2.159

Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology*, *61*, 678–685. 10.1037/0022-006X.61.4.678

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, *15*, 336–352. 10.1016/S0005-7894(84)80002-7

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. 10.1037/0022-006X.59.1.12

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, *3*, 1–27. 10.1146/annurev.clinpsy.3.022806.091432

Kopta, S. M., & Lowry, J. L. (2002). Dosage model. In M.Hersen & W.Sledge (Eds.), *Encyclopedia of psychotherapy* (pp. 655–660). New York: Academic Press. 10.1016/B0-12-343010-0/00079-9

Kopta, S. M., & Lowry, J. L. (2002). Psychometric evaluation of the Behavioral Health Questionnaire-20: A brief instrument for assessing global mental health and the three phases of psychotherapy outcome. *Psychotherapy Research*, *12*, 413–426. 10.1093/ptr/12.4.413

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. 10.1080/10503300601032506

Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. J.Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 169–218). New York: Wiley.

Lambert, M. J., & Finch, A. E. (1999). The Outcome Questionnaire. In M. E.Maruish (Ed.), *The use of psychological testing for treatment and planning outcomes assessment* (2nd ed., pp. 831–869). Mahwah, NJ: Erlbaum.

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

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

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. 10.1037/0022-006X.67.5.698

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767–778. 10.1093/biomet/88.3.767

Lutz, W. (2002). Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirical based clinical practice. *Psychotherapy Research*, *12*, 251–272. 10.1080/713664389

Lutz, W., Böhnke, J. R., & Köck, K. (2011). Lending an ear to feedback systems: Evaluation of recovery and non-response in psychotherapy in a German outpatient setting. *Community Mental Health Journal*, *47*, 311–317. 10.1007/s10597-010-9307-3

Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., & Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology*, *82*, 287–297. 10.1037/a0035535

Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., & Iveson, S. (2005). Predicting change for individual psychotherapy

clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology*, *73*, 904–913. 10.1037/0022-006X.73.5.904

Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, *118*, 60–68. 10.1016/j.jad.2009.01.019

Lutz, W., Stulz, N., Martinovich, Z., Leon, S., & Saunders, S. M. (2009). Methodological background of decision rules and feedback tools for outcomes management in psychotherapy. *Psychotherapy Research*, *19*, 502–510. 10.1080/10503300802688486

Martinovich, Z., Saunders, S., & Howard, K. (1996). Some comments on "assessing clinical significance."*Psychotherapy Research*, *6*, 124–132. 10.1080/10503309612331331648

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D.Kaplan (Ed.), *The Sage handbook of quantitative methodology for social sciences* (pp. 345–368). Thousand Oaks, CA: Sage. 10.4135/9781412986311.n19

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.

Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, *6*, 109–138. 10.1146/annurev.clinpsy.121208.131413

Newnham, E. A., & Page, A. C. (2010). Bridging the gap between best evidence and best practice in mental health. *Clinical Psychology Review*, *30*, 127–142. 10.1016/j.cpr.2009.10.004

Renaud, J., Brent, D. A., Baugher, M., Birmaher, B., Kolko, D. J., & Bridge, J. (1998). Rapid response to psychosocial treatment for adolescent depression: A two-year follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*, 1184–1190. 10.1097/00004583-199811000-00019

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92*, 726–748. 10.1037/0033-2909.92.3.726

Ronk, F. R., Korman, J. R., Hooke, G. R., & Page, A. C. (2013). Assessing clinical significance of treatment outcomes using the DASS-21. *Psychological Assessment*, *25*, 1103–1110. 10.1037/a0033100

Rubel, J., Lutz, W., & Schulte, D. (2013). Patterns of change in different phases of outpatient psychotherapy: A stage-sequential pattern analysis of change in session reports. *Clinical Psychology & Psychotherapy*. Advance online publication. 10.1002/cpp.1868

Schwartz, G. (1978). Estimating dimensions of a model. *Annals of Statistics*, *6*, 461–464. 10.1214/aos/1176344136

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, *78*, 298–311. 10.1037/a0019247

Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology*, *60*, 402–408. 10.1037/0022-006X.60.3.402

Stewart, J. W., Quitkin, F. M., McGrath, P. J., Amsterdam, J., Fava, M., Fawcett, J., & Roback, P. (1998). Use of pattern analysis to predict differential relapse of remitted patients with major depression during 1 year of treatment with fluoxetine or placebo. *Archives of General Psychiatry*, *55*, 334–343. 10.1001/archpsyc.55.4.334

Stulz, N., Lutz, W., Kopta, S. M., Minami, T., & Saunders, S. M. (2013). Dose–effect relationship in routine outpatient psychotherapy: Does treatment duration matter?*Journal of Counseling Psychology*, *60*, 593. 10.1037/a0033589

Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, *75*, 864–874. 10.1037/0022-006X.75.6.864

Acknowledgement: This research was supported in part by grants from the Humboldt Foundation (TransCoop Program, TCVERL-DEU/1133562).