Improving decision making in psychotherapy through typical patterns of treatment response

Inauguraldissertation zur Erlangung der Doktorwürde (Dr. rer. nat.) im Fach Psychologie, Fachbereich I an der Universität Trier

Vorgelegt im

August 2015

von

Julian Amadeus Rubel

Gutachter:

Prof. Dr. Wolfgang Lutz

Prof. Dr. Michael Schneider

Dissertationsort: Trier

DANKSAGUNG

Mein Dank gilt vor allem meinem Mentor und Chef Prof. Wolfgang Lutz für die vielen Möglichkeiten, die er mir über die Jahre immer wieder eröffnet hat, die Herausforderungen, welche er mir zugetraut hat, sowie seine Appelle Therapie stets unvoreingenommen und kritisch zu beforschen.

Weiterhin gilt mein Dank Prof. Michael Schneider für die Übernahme der Zweitbetreuerschaft und sein Interesse an dieser Arbeit.

Mein Dank gilt auch allen Kollegen, die mich auf dem Entstehungsweg dieser Arbeit begleitet haben. Es hat mir viel Spaß gemacht, in den unterschiedlichsten Projekten mit euch zusammen zu arbeiten. Ihr habt immer für die nötige Abwechslung und frischen Wind in der Abteilung und meinem Büro gesorgt!

Sicher nicht zuletzt gilt mein Dank meiner Freundin Violetta Schaan für das Lesen der Arbeit, ihre wertvollen Empfehlungen zu und über diese Arbeit hinaus sowie für den schönsten und kraftvollsten Baum, den man sich wünschen kann. Er wächst und gedeiht wirklich großartig.

Table of content

ABSTRACT 2
LIST OF ALL AUTHORED AND CO-AUTHORED PUBLICATIONS 4
LIST OF PUBLICATIONS FOR THE CUMULATIVE DISSERTATION 7
1 INTRODUCTION 8
2 THEORETICAL BACKGROUND 10
2.1 From treatment-focused to patient-focused psychotherapy research 10
2.2 Psychometric Feedback on individual change courses 12
2.3 Typical change patterns as benchmarks to evaluate response to psychotherapy 14
3 RESEARCH QUESTIONS 19
3.1 Study I 19
3.2 Study II 19
3.3 Study III 19
4 METHODOLOGICAL ASPECTS 20
4.1 Growth Mixture Modeling 20
5 STUDY I: EARLY CHANGE PATTERNS IN A RANDOMIZED CONTROLLED TRIAL FOR PATIENTS WITH PANIC DISORDER 23
5.1 Introduction 23
5.2 Methods 23
5.3 Results 24
5.4 Conclusions 26
6 STUDY II: PATTERNS OF CHANGE IN DIFFERENT PHASES OF OUTPATIENT PSYCHOTHERAPY 28
6.1 Introduction 28
6.2 Methods 28
6.3 Results 29
6.4 Conclusions 30
7 STUDY III: DEFINING EARLY POSITIVE RESPONSE TO PSYCHOTHERAPY 32
7.1 Introduction 32
7.2 Methods 32
7.3 Results 33
7.4 Conclusions 34
8 GENERAL DISCUSSION 36
8.1 General conclusions: 36
8.2 General Limitations and future research 38
8.3 Concluding remarks 40
9 REFERENCES 42
10 FIGURES 53
11 ORIGINAL PUBLICATIONS 54

ABSTRACT

Empirical psychotherapy research has shown in numerous original studies and meta-analyses that on average psychotherapy is effective. However, this average effectivity does not imply that a specific treatment is helpful for a specific patient. Therefore, it is important to accompany each treatment with an ongoing monitoring of patient progress through psychometric questionnaires and the feedback of this information to therapists. This seems to be especially important taking into account that often therapists perform badly in recognizing and foresee negative developments in their patients. With the help of psychometric feedback on the change course of their patients, therapists become more aware of potentially negative developments and can adjust their treatments accordingly. However, without additional information it might be difficult for therapists to evaluate the change course of their patients. Therapists need knowledge on what kind of change patterns are typically to be expected for a given patient. Comparing these benchmarks with the actual change course of a patient could enhance the evaluation of whether this patient is profiting or will profit from the provided treatment or not.

The present work summarizes three studies that deal with the identification of typical response patterns in an early phase of the treatment using growth mixture modeling (GMM). The first study investigates these patterns over the first five sessions of a cognitive-behavioral treatment of panic disorder patients (N=326). Four typical trajectories of change in panic disorder symptoms were identified that were significantly associated with treatment outcome and treatment length. Membership in these four classes was predictable on the basis of the patients' pretreatment functional impairment and separation anxiety.

The second study examined change patterns in session reports over the first six sessions, sessions seven to twelve, and sessions thirteen to eighteen in a heterogeneous outpatient sample (N=1229). The number of shared treatment response patterns identified via GMM reduced from five in the first phase, to three in the second, and two in the last phase. Pre-treatment interpersonal problems predicted class membership in phases one and three. With regard to treatment outcome, especially patterns of very early change (phase 1) were predictive of treatment outcome (symptom distress and interpersonal problems).

The third study compares GMM with a computationally easier method (clinically significant change) in the identification of patients who respond well in a very early treatment phase (first three sessions). In a naturalistic outpatient sample (N=5.484) "early positive responding patients" identified via GMM are compared with "early positive responding patients" identified via clinically significant change criteria. GMM identified early responding patients showed to

be a subgroup of reliably improved patients that also had good ultimate treatment outcome. Although only a little less specific for the prediction of positive treatment outcome, clinical significant change criteria performed much better than GMM in terms of sensitivity.

The results of the three studies are discussed with regard to their utility to support decision making in everyday clinical practice.

LIST OF ALL AUTHORED AND CO-AUTHORED PUBLICATIONS

Atzil-Slonim, D., Bar-Kalifa, E., Rafaeli, E., Lutz, W., Rubel, J., Schiefele, A. K., & Peri, T. (2015). Therapeutic Bond Judgments: Congruence and Incongruence. *Journal of Consulting and Clinical Psychology*. Advance online publication. doi: 10.1037/ccp0000015

Lambert, M.J., Rubel, J., & Lutz, W. (in press). Der Outcome-Questionnaire (OQ-45.2) [The outcome-questionnaire (OQ-45)]. In *Diagnostische Verfahren in der Psychotherapie*. Göttingen: Hogrefe.

Lutz, W., De Jong, K., & Rubel, J. (2015). Patient-focused and feedback research in psychotherapy: Where are we and where do we want to go? *Psychotherapy Research.* Advance online publication. doi: 10.1080/10503307.2015.1079661

Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M. A., Jorasz, C., ..., & Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research, 23*(1), 14-24.

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*(2), 287-297.

Lutz, W. & Rubel, J. (2015). Warum die Psychotherapieforschung in Deutschland ausgebaut werden sollte/könnte/müsste [Why psychotherapy research should and can be expanded in Germany]. *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie, 65*(5), 161-162.

Lutz, W. & Rubel, J. (2013a). Psychotherapieforschung [Psychotherapy research]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie*, 17. Aufl. (S. 1356). Bern: Hans Huber Verlag.

Lutz, W. & Rubel, J. (2013b). Psychotherapieforschung [Psychotherapy research]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie*, 16. Aufl. (S. 1268). Bern: Hans Huber Verlag.

Lutz, W., Rubel, J., & Böhnke, J. R. (in press). Der Fragebogen zur Evaluation von Psychotherapieverläufen (FEP) [The questionnaire for the evaluation of psychotherpay (QEP)]. In *Diagnostische Verfahren in der Psychotherapie*. Göttingen: Hogrefe.

Lutz, W., Rubel, J., Schiefele, A.-K., Zimmermann, D. Böhnke, J., & Wittman, W. W. (2015). Feedback and Therapist effects in the context of treatment outcome and treatment length. *Psychotherapy Research*. Advance online publication. doi: 10.1080/10503307.2015.1053553

Lutz, W., Wittmann, W. W., Böhnke, J. R., Rubel, J., & Steffanowski, A. (2012). Zu den Ergebnissen des Modellprojektes der Techniker-Krankenkasse zum Qualitätsmonitoring in der ambulanten Psychotherapie aus Sicht des wissenschaftlichen Evaluationsteams [Results from the pilot project of the Techniker Krankenkasse (TK) 'quality monitoring in outpatient psychotherapy': the evaluators' perspective]. *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie, 62*(11), 413-417.

Lutz, W., Wittmann, W., Böhnke, J., Rubel, J., & Steffanowski, A. (2013). Das Modellprojekt der TK zum Qualitätsmonitoring in der ambulanten Psychotherapie aus Sicht der Evaluatoren–Ein Plädoyer für mehr Psychotherapieforschung in Deutschland [The TK-Project Quality Monitoring in Outpatient Psychotherapy from the Perspective of the Evaluation Team-A Plea for More Psychotherapy Research in Germany]. *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie, 63*(6), 225-228.

Rubel, J., Ehrlich, T., & Lutz, W. (in press). Feedback-Ansätze zur Reduktion des Misserfolgsrisikos und zur Optimierung von Psychotherapie [Feedback appraoches for the reduction of treatment failure and the enhancement of psychotherapy]. *Psychotherapie im Dialog – PiD*.

Rubel, J. & Lutz, W. (2013a). Wachstumsanalyse [Growth Analysis]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie (*17th edition), (p. 1775). Bern: Hans Huber Verlag.

Rubel, J. & Lutz, W. (2013b). Phasenmodell psychotherapeutischer Veränderung [Phase model of psychotherapeutic change]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie (*17th edition), (p. 1262). Bern: Hans Huber Verlag.

Rubel, J. & Lutz, W. (2013c). Wachstumsanalyse [Growth Analysis]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie (*16th edition), (p. 1653). Bern: Hans Huber Verlag.

Rubel, J. & Lutz, W. (2013d). Phasenmodell psychotherapeutischer Veränderung [Phase Model of Psychotherapeutic Change]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie (*16th edition), (p. 1187). Bern: Hans Huber Verlag.

Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2014). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. *Psychological Assessment, 27*(2), 478-488.

Rubel J., Lutz W., & Schulte D. (2015). Patterns of change in different phases of outpatient psychotherapy: A stage-sequential pattern analysis of change in session reports. *Clinical Psychology & Psychotherapy, 22*(1), 1–14.

Strauss, B. M., Lutz, W., Steffanowski, A., Wittmann, W. W., Boehnke, J. R., Rubel, J., ..., & Kirchmann, H. (2015). Benefits and challenges in practice-oriented psychotherapy research in Germany: The TK and the QS-PSY-BAY projects of quality assurance in outpatient psychotherapy. *Psychotherapy Research, 25*(1), 32-51.

LIST OF PUBLICATIONS FOR THE CUMULATIVE DISSERTATION

1. Study I

   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*(2), 287-297.

2. Study II

   Rubel J., Lutz W. and Schulte D. (2015), Patterns of change in different phases of outpatient psychotherapy: A stage-sequential pattern analysis of change in session reports. *Clinical Psychology & Psychotherapy*, *22*(1), 1–14.

3. Study III

   Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2014). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. *Psychological Assessment*, *27*(2), 478-488.

1 INTRODUCTION

Taking the right decisions at the right moment is key to a successful clinical practice. Decision making in psychotherapy involves the assignment of specific strategies and techniques to a patient based on an elaborated indication process, the continuous evaluation of the success of the chosen approach, a timely detection of negative developments, and if necessary a modification of the treatment (Lutz, 2002). However, recent research on decision making in psychotherapy suggests that therapists perform badly in the detection of negative developments in their patients (e.g., Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010.). This is especially alarming when we consider that patients who show negative developments throughout the treatment course have an increased risk for bad treatment outcome (Haas, Hill, Lambert, 2002). Recent attempts to reduce the number of patients who fail to profit from psychotherapy use repeated assessments of patients' distress with psychometric questionnaires to feed this information back to the therapists. This feedback should support therapists in the continuous evaluation of their patients' progress. Often this feedback comprises a graphical representation of the patient's symptom scores over time. However, without benchmarks that indicate which developments are to be expected for a specific patient, it is difficult for clinicians to integrate this information into their clinical decision-making. Therefore, therapists need information on typical change patterns for specific patients to know if a certain course is more or less probable connected with success or failure. While much is known about the effectivity of different psychotherapeutic approaches, knowledge on individual change patterns is sparse (Castonguay, Barkham, Lutz, & McAleavy, 2013). In the last couple of years, this topic emerged in the field of psychotherapy research (e.g., Stulz et al., 2007). Different methodologies has been discussed that have the potential to reveal meaningful change patterns (e.g., Stulz et al., 2007; Pöhlmann et al., 2008; Quitkin et al., 1999; Tang & DeRubeis, 1999). One possibility to extract that kind of information from the bunch of individual response curves are pattern recognition procedures like Growth Mixture Modeling (GMM; e.g., Muthen, 2006). With GMM it is possible to group patients on the basis of shared treatment response patterns. In this dissertation, three applications of GMM for the detection of typical patterns of early treatment response are presented:

In the first study (Chapter 5), typical patterns of early treatment response are identified in a homogeneous sample of patients with panic disorder treated with a manualized cognitive behavioral treatment. Additionally, associations of these early change classes with patient characteristics, treatment outcome, and treatment length are examined.

Study two (Chapter 6) expands the lens from the early phase of the treatment to three treatment phases (sessions one to six; sessions seven to 12 and sessions 13 to 18) allowing a more detailed, partly non-linear, examination of typical change patterns in the first 18 sessions of outpatient psychotherapy.

The third study (Chapter 7) is an evaluation of the utility of GMM as a method to define early positive treatment response and compares it to computationally easier clinically significant change criteria (Jacobsen & Truax, 1991).

Before the three studies are described in more detail, a short theoretical overview is given from which the specific research questions that motivated these studies are deduced.

During my PhD studies I also authored and co-authored several additional articles and book chapters which are thematically related to the three studies summarized in this umbrella (Atzil-Slonim et al., 2014; Lambert, Rubel, & Lutz, in press; Lutz et al., 2013; Lutz, De Jong, & Rubel, in press; Lutz & Rubel, 2015; Lutz, Rubel, & Böhnke, in press; Lutz et al., 2015; Lutz, Wittmann et al., 2012, 2013; Rubel, Ehrlich, & Lutz, in press; Rubel & Lutz 2013a-d; Lutz & Rubel, 2013a,b; Strauß et al., 2015). However, these articles and chapters will not be discussed in detail in the present work.

2 THEORETICAL BACKGROUND

2.1 From treatment-focused to patient-focused psychotherapy research

The term psychotherapy is ubiquitous in everyday life. However, what do we mean when we use this term in clinical research? In 1986, Alain Kazdin already noted 400 variants of psychotherapy (Kazdin, 1986). This continuously increasing number shows that, depending on the respective context, very different things can be connected with this term (Lambert, 2013). The present text is based on a broad understanding of psychotherapy, since the models and applications discussed are applicable to all forms of psychotherapies. Thus, rather than referring to a specific form, "psychotherapies are viewed as a class of treatments defined by overlapping techniques, mechanisms, and proposed outcomes" (Castonguay et al., 2013; p. 87). Although the present work also includes sections in which specific treatments are more closely discussed, these should rather be viewed as exemplary applications that can be easily expanded to other psychotherapeutic approaches.

This generic approach to psychotherapy very well reflects a major paradigmatic shift that has taken place in the field of psychotherapy research over the last years. In short, this shift can be described as a change in the focus of most of the research endeavors from "treatment-focused" to "patient-focused": For quite a time, the pivotal goal of psychotherapy research was to prove that therapy works. Eysenck (1952) stimulated lots of research with a literature review providing evidence for the equivalency of psychodynamic therapy to spontaneous remission. Notwithstanding the intensive discussions about the drawbacks of this review, it should be regarded as crucial in terms of the research efforts it entailed. In the following "legitimation phase" (Grawe, 1992, 1997) numerous studies and meta-anlyses of these studies (e.g., Smith, Glass, Miller, 1971) showed the *efficacy* of psychotherapy in randomized controlled trials (e.g., Elkin et al, 1989), with a focus on the internal validity, as well as the *effectiveness* of psychotherapy under naturalistic conditions (e.g., Seligman, 1995), with a focus on the external validity. Consequently, today psychotherapy is regarded as an effective treatment for psychological disorders that helps most of the patients in overcoming their psychological distress and as such is a crucial element in health care systems around the globe (Lambert, 2013). Another debate of "treatment-focused" psychotherapy research is connected to the question: which form of psychotherapy is the most effective ("competition phase"; Grawe, 1997). Basically, this debate divides researchers and practitioners into two camps: Those who consider the approach in which they are trained, or to which they have some other kind of allegiance, as the most effective form and those who are convinced that there are no meaningful

differences in the effects of the different approaches (e.g., Wampold & Imel, 2015). While members of the first camp propose differences in the effectivity of the different treatments and explain them with differences in the specific techniques inherent to the approaches (e.g., cognitive restructuring in cognitive therapy or interpretation of dreams in psychodynamic psychotherapy), members of the latter camp propose the uniformity of outcomes and explain them with the importance of common factors (e.g., the therapeutic alliance or the healing context), which are shared by all forms of psychotherapy. Both camps continuously provide evidence for their way of thinking. Lately, a special issue on this debate, which has its origins in a study from Rosenzweig (1936), was published in *Psychotherapy* (2014 Volume 51, issue 4). Although, the articles in this current special issue again show the controversy of this debate, the authors from both camps recognize the importance of specific as well as common factors (e.g., Hofmann & Barlow, 2014).

These studies on the absolute and relative effects of psychotherapies contributed to the establishment and professionalization of these interventions. However, three central limitations of classical, treatment-focused psychotherapy research were repeatedly noted (e.g., Lutz, 2002): First, practitioners complain about the missing relevance of research results for their everyday clinical practice (e.g., Castonguay et al., 2013): A rigorous treatment focus is especially doubtful in terms of practical utility taking into account that between one half and two thirds of the practicing therapists do not use just one approach, but report to integrate techniques from different therapy schools (e.g., Norcross, Karg, & Prochaska, 1997). Having that in mind, knowledge on the effects of a specific sequence of techniques from within one approach has not much to offer for most of the clinicians.

Second, although it could have been shown in numerous studies and meta-analyses that psychotherapy is a psychological treatment that produces large average effect sizes, there is still considerable room for improvement. On an individual level, there are many patients who do not profit from psychotherapy or even deteriorate (e.g., Lambert, 2013).

Third, the question of how psychotherapy leads to change is rarely investigated, resulting in a lack of knowledge on the mediators and mechanisms that are active in successful psychotherapies (e.g., Kazdin, 2014).

The described shift from treatment to patient-focused research can be viewed as a direct reaction to these three criticisms. Core to patient-focused research is the continuous monitoring of outcome variables over the course of the treatment and the immediate use of the so gained information for ongoing treatments via feedback to therapists (Howard et al, 1996; Castonguay

et al., 2013; Lambert, Hansen, & Finch, 2001; Lutz, 2002; Lutz, De Jong, & Rubel, in press). As such, continuous outcome monitoring with psychometric questionnaires provides researchers with large databases on treatments, which are actually conducted in routine care. These large datasets widen the possibilities of psychotherapy researchers to address the criticism described above. Two applications of data from patient-focused research are central for the questions raised in the three studies summarized in this umbrella and thus are described in more detail below: Feedback on individual change courses and the identification of typical patterns of change to psychotherapy.

2.2 Psychometric Feedback on individual change courses

As stated above, it has been repeatedly shown that therapists often overestimate the progress their patients have made until a certain point during treatment and will make until the end of the treatment (e.g., Breslin, Sobell, Sobell, Buchan, & Cunningham, 1997; Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010; Walfish, McAlister, O'Donnel, & Lambert, 2012). Therefore, to increase therapists' awareness of negative developments, a continuous monitoring of patients' progress and the immediate feedback of the psychometric information to therapists has been proposed. Research on such kind of feedback interventions repeatedly showed an enhanced treatment effect for conditions in which therapists received psychometric feedback (e.g., Lambert & Shimokawa, 2011). These effects were especially pronounced for those patients showing negative developments (not on track patients; NOT) early in treatment (e.g., Lambert, 2007; Lambert & Shimokawa, 2011; Shimokawa, Lambert, & Smart, 2010). The most recent review on the effects of progress monitoring and feedback reports significantly positive effects of feedback vs. no feedback in 17 of 25 studies (Krägeloh et al., 2015). In terms of effect sizes, the most recent meta-analysis reports for NOT patients a superiority of feedback compared to treatment as usual (TAU) in the range of $g = .28$ ($p < .001$) in intention-to-treat (ITT) analyses and $g = .53$ ($p < .001$) in efficacy analyses[1] (Shimokawa et al., 2010). For on-track (OT) patients the advantage of the feedback condition was smaller, ranging from $g = .12$ ($p < .001$) in ITT to $g = .30$ ($p < .001$) in efficacy analyses. The effects of feedback were even more pronounced when therapists were provided with clinical support tools (CSTs) for NOT cases: $g = .44$ in ITT and $g = .70$ in efficacy analyses.

[1] In the efficacy analyses of the NOT/OT cases only patients were included who met the following conditions: 1) Patients attended at least four/two treatment sessions; 2) Patients filled out the outcome measure in at least three/two sessions; 3) At least one score in the outcome measure comes from a session after the patient was identified as NOT (only for NOT cases).

In addition, in early studies the use of continuous feedback seemed to optimize resource allocation. With the use of course feedback, NOT cases on average remained longer in treatment, while well developing cases (OT) terminated therapy earlier (e.g., Lambert et al., 2003). However, the results regarding the relations between feedback and treatment length are more mixed than those regarding treatment outcome (e.g., Shimokawa et al., 2010). A recent review found in only 6 out of 17 studies a significant association between feedback and treatment length. Accordingly, Shimokawa et al. (2010) did neither find longer treatments for NOT cases in the feedback condition compared to the NOT cases in the TAU condition nor did they find shorter treatments for OT cases in the feedback condition compared to OT cases in the TAU condition (Shimokawa, Lambert, & Smart, 2010).

The early feedback studies were restricted to moderately impaired patients in university counseling settings (e.g., Lambert et al., 2003). By now, the positive effects of feedback has been shown for different treatment modalities (e.g., group treatments; Davies, Burlingname, Johnson, Gleave, & Barlow, 2008), patient populations (e.g., youth; Bickman, Kelley, Breda, de Andrada, & Riemer, 2011), treatment settings (e.g., inpatient settings; Probst et al., 2013; Byrne, Hooke, Newnham, & Page, 2012), and diagnostic groups (e.g., eating disorders; Simon et al., 2013).

However, recent research suggests that feedback is not uniformly effective for every patient and therapist. Only 50% of the therapists in a study by Simon et al. (2012) were actually able to use the feedback to substantially improve their clients' outcomes. Similarly, Lutz and colleagues (2015) showed that therapist differences are an under-investigated factor in feedback studies and that therapists satisfaction with the feedback system and their use of the feedback information to modify the ongoing treatment are important variables that significantly predict treatment outcome.

Likewise, De Jong et al. (2012) found substantial differences between therapists in the degree to which they profited from a feedback intervention. Female therapists and those with a higher commitment to use the feedback information showed a higher probability to actually use the feedback. The use of feedback in turn was associated with being more effective for NOT patients. Furthermore, a higher commitment to use the feedback at the beginning of the study was related to more rapidly progressing patients. Following that, there seems to be a differential ability of therapists in using feedback information for the goods of their patients.

One of the tools that might help therapists to effectively use the feedback are additional benchmarks for the evaluation of their patients' progress. Right from the start of feedback research, it was recognized that graphical depictions of patients change courses are not

enough. Therapists need additional information to contextualize and evaluate the actual change course of their patients (Finch et al., 2001). Therefore, decision rules that provide therapists with additional information are an integral part of most feedback systems and facilitate the interpretation of individual change courses. These decision rules can be based on rational considerations of experts who use their knowledge to define them (rational decision rules; e.g., Lutz, Böhnke, Köck, 2011) or on empirical data which are used to infer individual predictions and typical change patterns (empirical decision rules; e.g., Stulz et al., 2007). The latter are focused in the next section.

2.3 Typical change patterns as benchmarks to evaluate response to psychotherapy

Repeated assessment of patients' distress or well-being on a session by session basis allows differentiated investigations of typical change patterns in psychotherapy. These typical change patterns are pivotal for psychometric feedback procedures. Without additional knowledge of what should be typically expected and which patterns are connected to good or bad treatment outcome, therapists can hardly evaluate the observed change course of their patients, which is fed back to them.

Different methods have been applied to identify typical change patterns (e.g., Tang & DeRubeis, 1999; Lutz et al., 2013; Stulz et al., 2007). Recently, especially growth mixture modeling (GMM; e.g., Muthen, 2006; see also chapter 4 for more details) was the method of choice to identify groups of patients with similar change trends over time. GMM clusters patients into subgroups with similar change trajectories on the basis of a latent categorical variable. It thus is a form of group-based trajectory modeling that extracts groups of change curves, which develop similar over time. This kind of clustering methods already stimulated much research in clinical psychology (cf. Nagin & Odgers, 2010). However, even within the endeavors using GMM, there is large heterogeneity with regard to the investigated treatment segment (whole treatment vs. specific treatment phases). For a variety of diagnostic groups, settings and psychological treatments, the use of GMM revealed clinically informative subgroups of patients with similar change trajectories over the whole treatment course:

Two of these studies found two response patterns, a group of "responders" and a group of "low responders" (Gildengers et al., 2005; Sunderland, Wong, Hilvert-Bruce, & Andrews, 2012). In these two studies, depressed patients, elderly depressed patients, and patients with generalized anxiety disorder (GAD) were treated with internet-based cognitive behavioral treatments or different combinations of antidepressants combined with interpersonal therapy in inpatient and outpatient settings. Two other studies report four typical response patterns for depressed

patients treated with CBT, CBT + medication, CBT + clinical management, or treatment as usual (Cuijpers et al., 2005; Keller & Hautzinger, 2007). The "responder" group found in the two above described studies splits in a "steep response" and a "less steep response" group; the "low responder" group splits in a "now response" and a "low response" group.

Knowledge on typical change patterns has also been shown to help differentiate the effects of different treatments. Stulz, Gallop, Lutz, Wrenn and Crits-Christoph (2010) used two-dimensional GMM to reanalyze data of a sample of 346 cocaine dependent patients treated with four psychosocial interventions in the multi-site randomized National Institute on Drug Abuse (NIDA) Collaborative Cocaine Treatment Study (CCTS). Besides a large group of "common and fast responders" (52%) that experienced a very rapid reduction of cocaine and overall drug use already in the first month of treatment, two groups with a more gradually response were detected. One of these two groups was characterized by an initially high overall drug and cocaine use ("severe users", 13%) and the other group started treatment averagely impaired ("moderate responders", 35%). Through this finer grained investigation Stulz et al. (2010) could show that the previously found superiority of one of the interventions (individual drug counseling; IDC) was mainly attributable to better outcomes of this intervention for one of the subgroups ("moderate responders"). This highlights the additionally gained information, which would have remained undetected without the use of modern statistical methods such as GMM.

All the applications described so far classified patients on the basis of shared response curves over the whole treatment. However, it is important to differentiate between applications investigating response patterns over the whole treatment course and those which separately model specific treatment phases. Most often patterns were investigated over the whole course and a concentration on patterns in the early treatment phase has been rare. However, such categorizations on the basis of shared response patterns over the whole treatment can easily miss important early developments. With every further session included in the clustering model, the importance of the early phase for the categorization is lessened. Consequently, the longer the period that builds the basis for the clustering, the less detailed is the representation of the early treatment phase. One reason to have a separate look on the early phase of therapy is the finding that patients' end state scores in an outcome measure are regularly highly positively correlated with initial scores in that measure (e.g., Lutz et al., 1999). That is, patients with higher levels of impairment at the beginning of a treatment also tend to be more distressed at the end of the treatment. To further improve the predictions of treatment outcome, patients' developments early in treatment has been considered as an additional predictor for later developments and ultimate treatment outcome (e.g., Haas et al., 2001). Mostly, research into

early change patterns was concerned with the identification of patients showing rapid early symptom reductions (e.g., Renaud et al., 1998). However, there is large variety in the definition of early response: First, the period supposed as "early" differs considerably between studies; this variability is a natural consequence of the large differences between countries with respect to the average treatment length provided. Second, several different methodological approaches have been applied to identify positive changes early in treatment. Some of the studies divided patients in early responders and early non-responders depending on whether or not an a priori, rationally defined criterion was met. For example sudden gains, which are defined as substantial symptom reductions from one session to the next, has been shown to predominantly be present in an early phase of the treatment (e.g., Lutz et al., 2013). Other studies used ROC curves to identify the minimum amount of change that accurately (in terms of specificity[2] and sensitivity[3]) predict remission at the end of the treatment. For example Gunlicks-Stoessel and Mufson (2011) reanalyzed data of a randomized controlled trial which tested the effectiveness of an interpersonal psychotherapy for depressed adolescents (IPT-A). For the rather small sample of 63 adolescents a minimum symptom reduction in the first 4 weeks of 16.2% (IPT-A) and 24.4% (TAU) respectively, was found to predict remission status at week 16 specifically and sensitively (Gunlicks-Stoessel & Mufson, 2011). Haas, Hill, Lambert and Morell (2002) mapped treatment response on a weekly basis and found that quick positive treatment response within the first three sessions was related to positive outcome at termination and follow-up. Furthermore, early responders needed less sessions of therapy to achieve these stable improvements and thus were more likely to terminate treatment earlier (Haas et al., 2002). Early response in this study was defined on the basis of expected treatment response (ETR) curves generated by hierarchical linear models. Those 25% of the patients, whose observed response curves were the most positively different from their expected response curves, were defined as early responders.

Despite these differences in terms of definition, early response consistently has been shown to be present and to be a powerful incremental predictor of therapy outcome in naturalistic studies (e.g., Flückiger, Holtforth, Znoj, Caspar, & Wampold, 2012; Haas et al., 2002; Lewis, Simons, & Kim, 2012; Lutz, Stulz, Smart, & Lambert, 2007; Nordberg et al., 2014; Stulz et al., 2007) and controlled trials (e.g., Lutz, Stulz, & Köck, 2009; Tadić et al., 2010; Uher et al., 2010) for psychological (e.g., Haas et al., 2002; Stulz et al., 2007), pharmacological (e.g., Rynn, Khalid-

[2] Specificity refers to the proportion of patients not identified as early positive responders from the ultimately not improved patients.
[3] Sensitivity refers to the proportion of patients identified as early positive responders from the ultimately improved patients.

Khan, Garcia-Espana, Etemad, & Rickels, 2006) and combined treatments (e.g., van Calker et al., 2009) across different samples of minor, mild, mild to moderate and severely depressed adolescents, adults and elder patients (e.g., Gildengers et al., 2005; Tadić et al., 2010; van Calker et al., 2009), patients with anxiety disorders (e.g., Hofmann, Schulz, Meuret, Moscovitch, & Suvak, 2006; Rynn et al., 2006) and patients with eating disorders (e.g., Grilo et al., 2006; Hilbert et al., 2007).

Research into early change patterns was mainly concerned with the identification of patients showing rapid early improvements. This restricted perspective has the consequence that other clinically important classes could have remained undetected. Group-based trajectory models like GMM are not restricted to a specific response pattern and thus are an appropriate approach to identify several typical change curves. Only three studies so far investigated specifically the early treatment phase with GMM (Stulz et al., 2007; Lutz et al., 2007; Lutz, Stulz, & Köck, 2009). Two of these three studies were conducted with data from routine care (Stulz et al., 2007; Lutz et al., 2007) and the other study used data from the National Institute of Mental Health (NIMH) Treatment for Depression Collaborative Research program (TDCRP; Elkin et al., 1990; Lutz et al., 2009). However, due to the different characteristics of the studies and the slightly different approaches applied, it is difficult to extract convergences in the results. Stulz et al. (2007) identified five shared response trajectories over the first six sessions using the whole patient sample. Lutz et al. (2007) took into account that GMM is often liable to intake differences resulting in classes which are mainly different in the level of intake distress and not that much in the rate of change. Thus in a first step, the authors divided the sample on the basis of their intake impairment into three groups. Afterwards, typical patterns were identified over the first five sessions for each of these three groups separately. Consequently, instead of five trajectories (Stulz et al., 2007), Lutz et al. (2007) identified 12 different response patterns within this early phase. Besides these differences, both studies report a subgroup of early improving patients with a high probability for a successful treatment outcome. While Stulz et al. (2007) report a subgroup of 11% that started treatments severely impaired and showed a rapid substantial improvement over the first six sessions, Lutz et al. (2007) observed this pattern for only 3.8% of the patients. However, it is unclear whether these differences are due to the different time frames that have been investigated (five vs six sessions), the different outcome measure used, the different composition or size of the samples, or the different modeling approach.

In the only study that used data from a randomized controlled trial (RCT) to determine early change patterns, about 61% of the N = 162 depressed patients were identified as early

responders over the first eight weeks of treatment. Two other classes were identified showing moderate early improvements but starting on different levels of pre-treatment depression. While the one started treatment moderately to severely impaired, the other group showed mild to moderate levels of pre-treatment depression. Due to the multiple differences between this study and the two other studies on patterns of early treatment response, the much larger share of early improving patients in this study must be interpreted cautionary and could not be attributed to any single study characteristic.

Summarizing the evidence on patterns of change to psychotherapy the following limitations should be noted, which encourage further research: So far, only one study investigated typical change patterns in a homogenous patient sample treated under controlled conditions with a structured intervention (i.e., Lutz et al., 2009). Further studies are needed that investigate specific diagnostic groups from randomized controlled trials. Especially other diagnostic groups than depression should be examined more closely.

Another limitation is related to the time units, which are considered with pattern recognition approaches. Past endeavors tried to identify typical change patterns either over the complete treatment course or specifically for the early treatment phase. However, both approaches are prone to miss important developments, which take place in later phases of the treatment. While the complete treatment approach is not sensitive enough to discover these developments in specific phases, the early phase approach is limited to early developments.

Furthermore, one should bear in mind that GMM is a rather complex statistical method, which is not easy to implement in routine care. Other methods like clinically significant change criteria can be easily calculated by clinicians and thus have some advantages over GMM in terms of dissemination. Therefore, GMM needs to legitimate its application in routine clinical decision making in view of the increased effort, which would be connected to its implementation in routine care.

3 RESEARCH QUESTIONS

Taking into account past research on patterns of change in psychotherapy, the following research questions are addressed in the three studies presented in this umbrella:

3.1 Study I

1. Which early change patterns can be identified in a homogeneous sample of panic disorder patients treated with a manualized CBT treatment?
2. How predictive are these early change patterns for treatment outcome and treatment length?
3. Which variables predict membership in one of the early change classes?
4. An explorative aim of this study is to add knowledge on differences and similarities of change patterns between different disorders and research designs.

3.2 Study II

1. Which change patterns in a progress measure can be identified in different phases of outpatient psychotherapy?
2. Which variables do predict class membership in different phases?
3. How predictive are these class memberships for the membership in classes of later phases, treatment outcome and treatment length?

3.3 Study III

1. How similar or distinct are classifications of early positive treatment response based on different methodological approaches: Growth mixture modeling and clinically significant change criteria.
2. Which intake variables are able to differentiate the differentially identified patient subgroups?
3. Can the computationally intensive method of GMM improve outcome prediction models over easier clinically significant change criteria?

4 METHODOLOGICAL ASPECTS

All three summarized studies share a common procedure for the identification of typical change patterns, Growth Mixture Modeling (GMM; e.g., Muthen, 2006). Therefore, this approach is shortly described in general and with an emphasis on the specific model selecting procedure applied in the three applications.

4.1 Growth Mixture Modeling

GMM is a latent variable method that expands traditional latent growth models. Latent growth models capture individual change over time by the use of latent intercept and slope mean factors and random variation around these means. It is assumed that all individuals stem from a common population for which these factor means and variances apply. By introducing an additional latent categorical variable, GMM relaxes this assumption. This categorical variable enables the estimation of several intercept and slope factor means as well as their respective variances. The number of estimated classes must be defined a priori. Since the number of classes in the population is unknown, a series of models are tested. On the basis of multiple model fit indices it is decided how many classes should be specified. Besides determining the mean and variance of the intercept and slope factors for each class, GMM also estimates for each individual the probability to origin from the respective classes.

GMMs vary with regard to the specification of the variance parameters. In the most conservative GMM family it is assumed that all individuals from the same class have exactly the same change trajectory. That means, that no variation is assumed within the classes and consequently all within-class variances are fixed to zero (Nagin & Tremblay, 2001). These models are also referred to as latent class growth models (e.g., Pickles & Croudace, 2010). Latent class growth models capture all the differences between the individual growth curves by the intercept and slope mean factors of the different classes. Within the classes individuals are assumed to have exactly the same intercept and slope. Less conservative GMMs not only capture change by class specific intercept and slope means but additionally allow individuals to have varying intercepts and slopes within the classes (e.g., Muthen, 2006). In all the three studies which are part of this dissertation, a hybrid model has been chosen. While the variances of the slope factors were fixed to zero, the variances in the intercept factors were freely estimated but held constant between the classes. This approach was used because it increased the stability of the models and matched the theoretical emphasis on differences in change courses instead of intercepts. Given the fact that the individual intercepts are allowed to vary within the classes, the extraction procedure was more sensible to differences in the individual

slopes and less sensitive to intercept differences. *Figure 1* shows a simplified version of the path model for this approach. Latent variables are displayed in circles and manifest variables are displayed in rectangles. The path loadings and error terms are not depicted to improve the clarity of the illustration. All six manifest variables load on the intercept factor with 1 (i.e., the actual loadings would be: 1, 1, 1, 1, 1, 1). A logarithmic relation between number of treatment sessions and patients distress has been repeatedly shown to be the most adequate model of change in psychotherapy (e.g., Stulz et al., 2013). To model logarithmic growth over the six therapy sessions of this example, the loadings on the slope factor start with the natural logarithm (or the logarithm to the base 10) of one for the first session score and ends with the natural logarithm of six for the sixth session score (i.e., the actual loadings would be: log(1), log(2), log(3), log(4), log(5), log(6)). Since the log(1) is zero the intercept represents the estimated score in the first session. The single arrow directed on the intercept factor represents the variation which is allowed to be freely estimated for this factor.

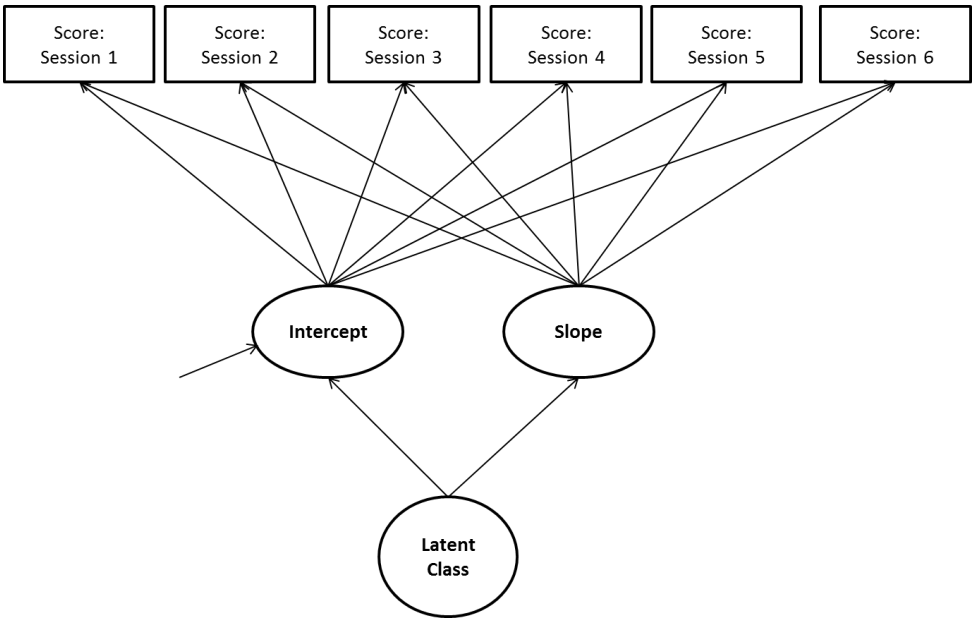

Figure 1: Simplified path model of the applied GMM for the determination of latent change patterns over the first six treatment sessions.

The modeling procedure for all three studies was divided in two steps: In a first step, the overall shape of change (no change, linear change, log-linear change) was determined applying relative model fit indices (BIC; Schwartz, 1978) for the comparison of the different models. In a second step, separate models with an sequentially increasing number of classes were fitted and compared via the BIC and likelihood ratio tests (Lo, Mendel, & Rubin, 2001; Nylund, Asparouho, Muthen, 2008). This model has been considered as best solution which had the lowest BIC value and at the same time showed to be superior than a model with one class less

in a likelihood ratio test. A combination of these two fit indices has been applied since each of these statistics alone has shown to be rather to liberally extracting classes in simulation studies (e.g., Nylund et al., 2008).

For each individual GMM estimates a probability to belong to each of the identified classes. However, in all three applications patients are categorized after estimation, in the class for which they have the highest probability. This approach is typical for probabilistic pattern recognition procedures like GMM (e.g., Pickles & Croudace, 2010).

5 STUDY I: EARLY CHANGE PATTERNS IN A RANDOMIZED CONTROLLED TRIAL FOR PATIENTS WITH PANIC DISORDER

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*(2), 287-297.

5.1 Introduction

Most research on early change patterns has been conducted within naturalistic datasets (e.g., Stulz et al., 2007; Lutz et al., 2007). These datasets are characterized by heterogeneous diagnostic samples as well as unstructured, not manualized treatments. However, to provide therapists with as much empirical information as possible on the expected treatment response of their patients, it is important to collect data from different settings and samples. Besides data from big naturalistic studies with heterogeneous patient samples and treatments, analyses from more homogeneous samples with specific disorders and treatments are needed to derive and validate reliable benchmarks for treatment response in these settings. Recently, few studies dealt with treatment response patterns of depressed patients treated with pharmacotherapy (Gueorguieva, Mallinckrodt, & Krystal, 2011; Uher et al., 2010). The only study investigating early change patterns in patients treated with psychological treatments also dealt with depressed patients (Lutz et al., 2009). Other diagnostic groups than depression as well as response to psychological treatments have hardly been investigated with regard to early change.

Although there are several reports utilizing GMM to examine early symptom changes (Lutz et al, 2007; 2009; Stulz et al., 2007), none has specifically focused on panic disorder patients treated with CBT. In accordance with the approach applied by Lutz et al. (2009), in study I data of a controlled trial is reanalyzed in which panic disorder patients were treated with a specialized CBT treatment (Aaronson et al., 2008; Barlow & Craske, 2007; White et al., 2010).

5.2 Methods

Sample

The sample of study I consists of N = 326 patients who attended at least three sessions of an 11-session manualized CBT treatment for panic disorder (Barlow & Craske, 2007). Data was collected as part of a multisite clinical trial at four sites in the US (Aronson et al., 2008; White et al., 2010; White et al., 2013).

Progress Monitoring

As central measure of patient symptom change over the course of the treatment, the Panic Disorder Severity Scale-Self Report version (PDSS-SR; Sear et al., 1997, 2001) was administered before each treatment session.

Patient intake characteristics

Information on patients age, initial clinician-rated depression (Hamilton Rating Scale for Depression, HAM-D; Hamilton, 1960), initial clinician-rated anxiety (Hamilton Anxiety Rating Scale, HAM-A; Hamilton, 1959), symptoms of separation anxiety in adulthood (Adult Separation Anxiety-Checklist, ASAC; Manicavasagar, Silove, Wagner, & Hadzi-Pavlovic, 1999), and the degree of restrictions caused by symptoms in 5 life domains (Work and Social Adjustment Scale, WSAS-SR; Mundt, Marks, Shear, & Greist, 2002; Hafner & Marks, 1976) were available and subsequently used as predictors of class membership, treatment outcome and treatment length.

Assessment of treatment outcome

Treatment outcome was assessed by the means of pre to post effect sizes (d) in the PDSS-SR and number of reliably improved patients from pre to post treatment.

Data Analytic Strategy

In a first step patients are classified based on their first five scores in the PDSS-SR using the hybrid GMM approach described in 4.1. Subsequently, patients' intake characteristics are tested as predictors of class membership using ANAOVAs and multinomial logistic regression models. Differences between the change classes regarding treatment outcome and length are analyzed using ANOVA and Chi-square tests. Finally, the incremental predictive power of class membership for treatment outcome and treatment length is evaluated with hierarchical multiple regression analyses.

5.3 Results

GMM analyses revealed four typical patterns of early symptom change in this sample of panic disorder patients. *Figure 2* shows the mean latent growth curves for these four classes as well as the number of patients and their percentage in relation to the whole sample.

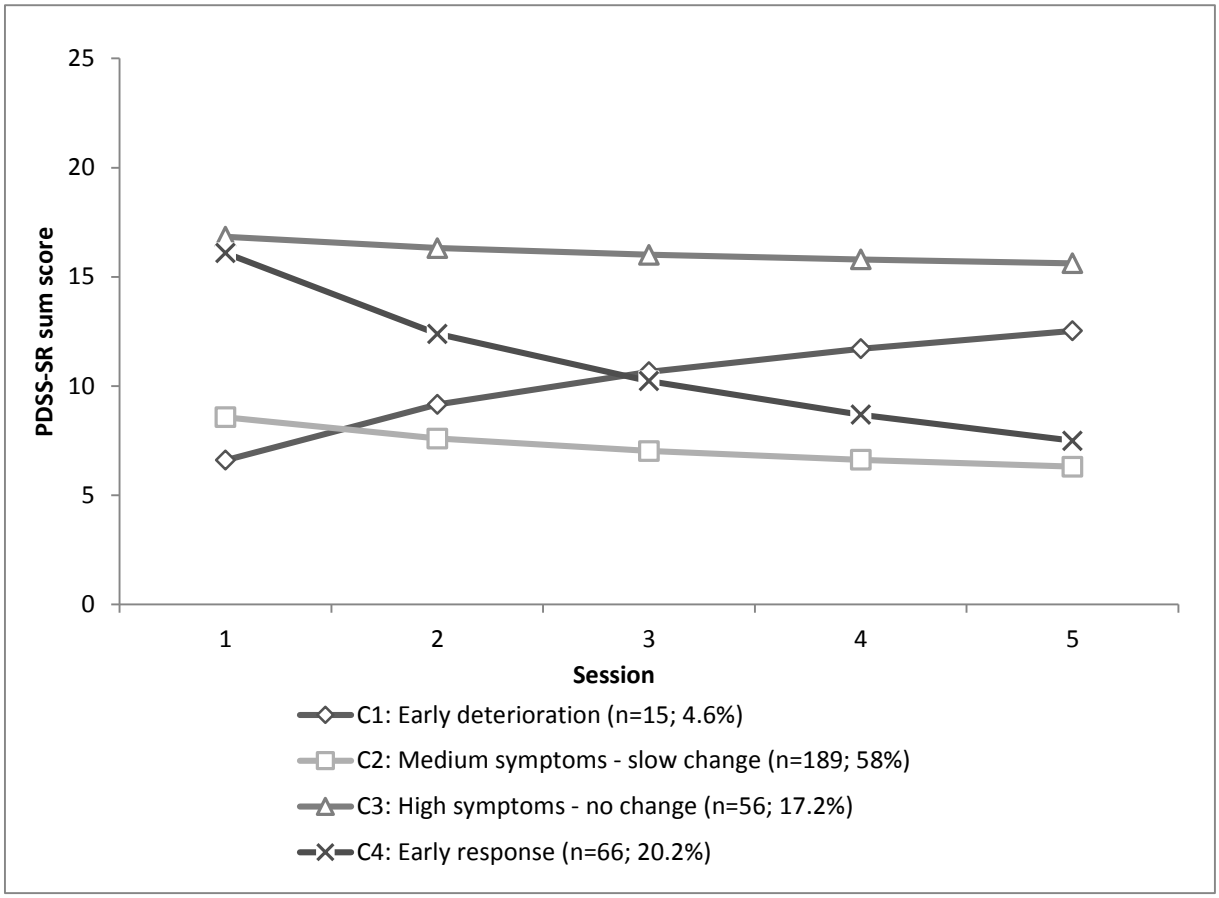

Figure 2. Mean latent growth curves (early change classes; C1-C4) within the first five sessions of patients with panic disorder treated with CBT.

All intake characteristics with the exception of patient age were significantly different between the four classes. Generally, patients in classes one and two showed significantly lower initial depression (HAM-D), anxiety (HAM-A) and adult separation anxiety (ASAC) scores than patients in classes three and four. This was similar for patients functioning (WSAS-SR) scores with the exception that classes one and four were not significantly different.

Early change classes also showed to be predictive for treatment outcome and treatment length. Beyond the first PDSS-SR score and the other patient intake variables, membership in one of the four classes explained 16% additional variation in treatment outcome and 3% in treatment length. Patients in class four showed significantly higher pre-post effect sizes (d = 2.11) and rates of reliably improved patients (93.3%) than all other classes. In contrast, patients with early deteriorations (C1) showed negative pre-post effects sizes (d = -0.49) and none of these patients improved reliably. With regard to treatment length there was a significantly higher number of patients from class 3 that prematurely ended treatment in the first three to five sessions. Whereas this group of early non-responding patients had the lowest average

treatment length (*M* = 9.02*)*, patients in the early response group had the highest average treatment length (*M* = 10.29*).*

5.4 Conclusions

Study I revealed similar typical patterns of early change for panic disorder patients treated with a manualized 11-session CBT as those found in previous investigations for other disorders, treatments and samples. However, the rates of patients showing positive early treatment response differ quite a bit between these studies. While in the present study about 20.2% of the patients showed positive early response, 61.1% of depressed patients treated with psychological therapies (Lutz et al., 2009), and 25% of depressed patients treated with pharmacotherapy (Uher et al., 2010) were identified as early responders in previous studies. Thus, this study seems to provide preliminary evidence for the assumption that early positive treatment response is more prevalent in depression than in panic disorder. In comparison with investigations in naturalistic samples, the share of patients showing early response seems to be somewhat higher (20%, Lutz et al., 2007; 12.5%, Stulz et al., 2007). In contrast to other studies investigating the early phase of clinical trial treatments for depression, in this panic disorder sample, also a small group of patients with early deteriorations was identified. However, the conclusion that early deteriorations are more typical for panic disorder than for depression should also be drawn with caution due to the between study comparison.

Adult separation anxiety as well as social/work functioning seem to be important additional predictors for the early treatment course, which should be more regularly considered for panic disorder patients in particular and maybe also for other patient groups.

Again, as has been repeatedly shown in several other studies (Flückiger, grosse Holtforth, Znoi, Caspar, & Wampold, 2013; Tadic et al., 2010), early treatment response showed to be an informative predictor for treatment outcome and length. Interestingly, early positive treatment response was connected to rather long treatments in the current study, while patients with high symptoms and slow initial changes showed an increased probability for premature treatment termination. This finding is in contrast to what is known from research on early response in naturalistic datasets in which early positive response is connected to rather short treatments. This might be caused by the different structure of these different study designs. While naturalistic therapy is open ended in nature, treatments in the context of RCTs are time limited with a fixed number of treatment sessions. Therefore, for patients participating in clinical trials early positive treatment response might enhance their compliance to the treatment and they rather complete the beforehand known fixed number of sessions. In contrast, for patients treated

under naturalistic settings, early positive response might lead to a faster achievement of their individual good-enough level (Barkahm et al., 2006) resulting in a consensual termination of the treatment.

6 STUDY II: PATTERNS OF CHANGE IN DIFFERENT PHASES OF OUTPATIENT PSYCHOTHERAPY

Rubel J., Lutz W. and Schulte D. (2015), Patterns of change in different phases of outpatient psychotherapy: A stage-sequential pattern analysis of change in session reports. *Clinical Psychology & Psychotherapy*, *22*(1), 1–14.

6.1 Introduction

Study II extends research on change patterns in outpatient psychotherapy in two ways. First, the examination is not restricted to the early phase of the treatment: Additionally to the first six sessions (first phase), two additional phases (sessions seven to twelve and session thirteen to eighteen) are investigated. Thus, typical change patterns are examined in each of these phases as well as typical transition patterns between these. Second, for the first time, patterns of change were identified with a progress instead of a symptom measure. Past research was limited to typical patterns of change in symptom measures (e.g., Stulz et al., 2007).

6.2 Methods

Sample

The sample of Study II comprised N = 1229 patients treated with individual CBT for heterogeneous disorders in two German outpatient psychotherapy clinics.

Progress Monitoring

As central measure of patient progress a subscale of the Session Report (SR; Flückiger, Regli, Zwahlen, Hostettler, & Caspar, 2010; Schulte & Eifert, 2002), measuring progress with regard to mastery and clarification experiences, was administered after each treatment session.

Patient intake characteristics and assessment of treatment outcome

Information on patients overall symptom distress (Brief Symptom Inventory, BSI; Franke, 2000; German translation of Derogatis, 1975) and interpersonal distress (German version of the Inventory of Interpersonal Problems, IIP-D; Horowitz, Strauss, & Kordy, 2000) were assessed before and after the treatment. Pre and post scores as well as change scores (d) were analyzed with regard to their associations with class membership.

Data Analytic Strategy

In a first step, patients are classified based on their first six scores (phase 1), second six scores (sessions 7-12, phase 2), and third six scores (sessions 13-18, phase 2) in the SR using the

hybrid GMM approach described in 4.1. Subsequently, separate AN(C)OVAs for each variable and phase were calculated to compare the classes regarding 1) patients' intake and post treatment scores in the BSI and the IIP-D, 2) change from pre- to post treatment (d) in these instruments as well as 3) treatment length.

6.3 Results

GMM analyses revealed five typical patterns of patient progress in the first five sessions and only three and two patterns in the following phases, respectively. *Figure 3* depicts the mean latent growth curves in each of the three phases as well as the absolute and relative number of patients categorized to the classes.

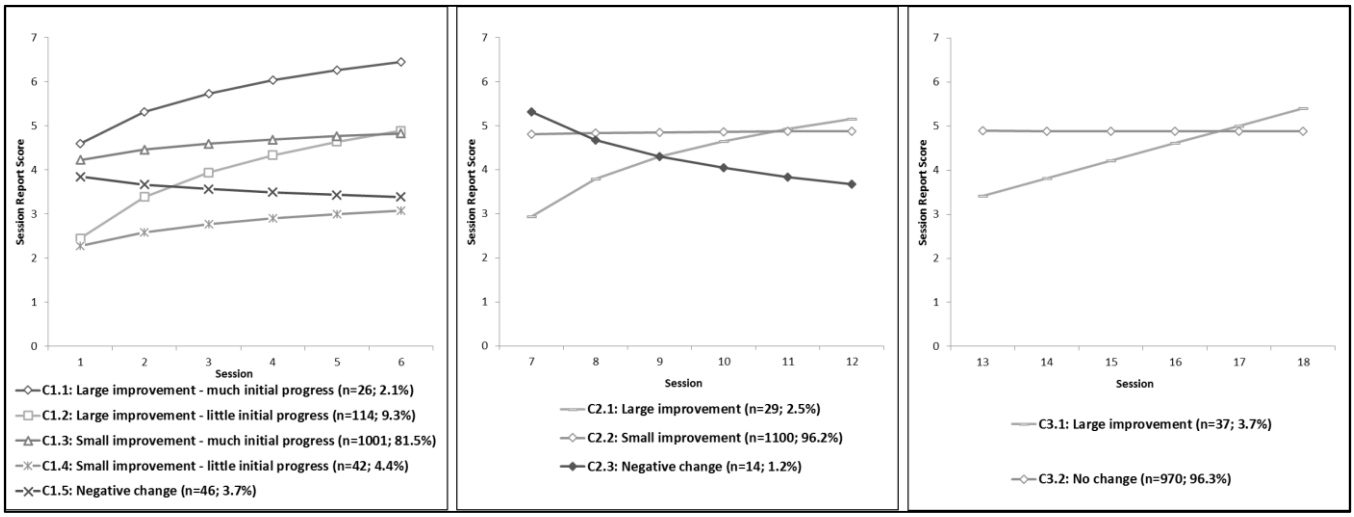

Figure 3. Typical change patterns in patients' subjectively perceived mastery and clarification experiences in three phases (sessions 1-6, 7-12, and 13-18) of outpatient psychotherapy.

In each of the phases the biggest class of patients starts with rather much progress in the first session of the respective phase and does not change much over the following five sessions (C1.3, C2.2, C3.2). The number of patients within this biggest class seems to increase slightly over time. Also consistently prevalent in each phase is a class of patients who show little initial progress and improve significantly over the next five sessions (C1.2, C2.1, C3.1). The number of patients showing such a pattern of progress seems to decrease from phase one to phase two and remains at about that level in phase three. Similarly, the number of patients showing a pattern of negative change decreases from phase one (C1.5) to phase two (C2.3) and cannot be observed in phase three. Additionally to the group of patients showing large improvements in the early phase and showing little initial progress, in phase one an additional small group of patients who already experience much progress in the first session and showing large improvements over the next five sessions could be observed (C1.1).

IIP-D pre scores were significantly related to class membership in the first and the third phase but not in the second phase. Patients of the large C1.1 had significantly less interpersonal distress than patients of C1.4. Similarly, in phase three patients from C3.1 were less distress with regard to interpersonal problems than patients from C3.2.

Class membership in phase one was significantly related to treatment outcome and length. Patients in C1.4 and C1.5 had significantly higher BSI_{post} scores than patients from all other three groups. Interpersonal distress after treatment was significantly lower for patients from C1.1 than for all other patient groups. Patients from C1.1 had significantly shorter treatments than patients in C1.3 and C1.5.

In phase two only symptom distress after treatment was related to class membership. Patients in C2.3 were significantly more impaired after treatment than patients from the other groups. Membership in one of the two classes in phase three was not significantly related to treatment outcome or length.

Due to the predominance of one class in each of the phases, the most prevalent transition patterns are connected with these classes. Consequently, more than two thirds of the patients who were still in treatment in all three phases showed a pattern of C1.3-C2.2-C3.2. The only noteworthy deviation from this transition pattern is observed for patients in C1.4. These patients who experience little initial progress and show small improvements in the following five sessions (C1.4) seem to have an increased chance to experience large improvements in phase 2. However, given the small classes in phases two and three, inferential conclusions via Chi-square tests cannot be derived.

6.4 Conclusions

Study II showed that it is possible to identify different patterns of early change in a session report and that these resemble those patterns identified in previous studies with symptom measures (e.g., Stulz et al., 2007). The group of early positively responding patients, which could have been shown in previous studies with symptom measures might be split up in session report classes C1.1 and C1.2. Although it is also possible to identify typical patterns of change in later phases, the number of observed subgroups decreases over time. Thus, response patterns in patients' session reports seem to be more differentiated in earlier phases than in later phases. This finding of reduced number of typical change patterns is most likely due to a more diverse use of the instrument by the patients in the beginning of the treatment. In later sessions, variation in patients responses in the session reports seem to be reduced and level out on a relatively high level. To test this hypothesis, the average deviations from the individual growth curves in these

session reports were compared between the three phases. In accordance with this hypothesis, the root mean square error (RMSE)[4] decreased considerably from the first to the second phase and again slightly from the second to the third phase. Thus, patients responses in the applied progress measure seem to be clinically most informative in earlier treatment sessions, while in later treatment sessions there seem to be only little variation from session to session.

Overall, the most promising pattern with regard to treatment outcome and length seems to be if patients perceive much progress from the first session on and remain to score on that level or even increase further over time (C1.1). Patients showing such a pattern in the first phase of the treatment showed the best outcomes, while simultaneously having the shortest treatments. The findings of this study underline the importance of the early phase of the treatment as an invaluable source of information for clinicians. However, it reveals also helpful information on what could be typically expected in patients' session reports in later phases of the treatment.

[4] The RMSE is calculated as the square root of the averaged squared deviations of the observed scores from the expected scores estimated based on the individual growth curve.

7 STUDY III: DEFINING EARLY POSITIVE RESPONSE TO PSYCHOTHERAPY

Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2014). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. *Psychological Assessment, 27*(2), 478-488.

7.1 Introduction

Studies I and II as well as several other studies investigating patterns of early change repeatedly identified a subgroup of patients showing high initial impairment (or small initial progress; compare Study II) and rapid early improvements with the use of GMM. As described, GMM is a rather computationally intensive method that is not applicable without the help of special statistical software. However, for the definition of early positive treatment response simpler methods have been discussed. One of these are clinically significant change criteria (Jacobsen & Truax, 1991). In accordance with the concept of clinically significant change, a patient improved clinically meaningfully from one time point to another, if he fulfills two criteria: First, the amount of change must be greater than a certain, instrument specific change score that takes into account measurement error (reliable change index; *RCI*). Second, a patient's score must change from above to below a certain predefined cut-off score that distinguishes an impaired from a healthy sample. Knowing these two indices, the respective RCI and cut-off score for an instrument, practitioners can easily evaluate change in their patients from one time point to another. Study III is a comparison between early positive responder groups, which are differentially identified with GMM or with clinically significant change criteria.

7.2 Methods

Sample

The sample of Study III comprised N = 5.484 patients treated for a minimum of four sessions by one of 240 therapists in one of 20 college counseling centers, four primary care medical centers, or two private mental health centers.

Progress Monitoring

As central measure of patient change the Behavioral Health Measure-20 (BHM-20; Kopta & Lowry, 2002), was administered prior to each treatment session. The overall mean is a measure of Global Mental Health (GMH) with high scores indicating good psychological functioning.

Patient intake characteristics and assessment of treatment outcome

The first and the last available GMH scores were used as measures of pre- and post-treatment psychological functioning respectively.

Data Analytic Strategy

In a first step, patients were classified based on their first three GMH scores using the hybrid GMM approach described in 4.1. It is expected that this classification identifies a subgroup of early positively responding patients. Additionally, patients first and third GMH scores were evaluated applying clinically significant change criteria. These classifications should result in three potentially overlapping early positively responding subgroups: One identified with GMM, one identified with the criterion of reliable change, and one identified with the criteria of clinically significant change. Subsequently, these subgroups were compared with regard to the following dimensions: Initial psychological functioning, pre to post functioning change, stability of change, and predictive accuracy. Predictive accuracy of the groups was evaluated in terms of specificity and sensitivity for the prediction of reliable and clinical significant change from pre to post treatment.

7.3 Results

GMM revealed four typical change patterns in the first three GMH scores. As expected, the GMM approach identified one patient subgroup starting treatment on a relatively low level of psychological functioning and substantially improving from the first to the third session. This subgroup comprised 396 (7.2%) of all patients. A larger number of 1341 (24.5%) patients improved reliably from intake to session three and 2407 (43.9%) crossed the cutoff score that separates a clinical from a non-clinical sample in the direction of the latter. 778 (16.3%) of these patients met both criteria necessary to be classified as clinically significant improved (see *Figure 4*). As displayed in *Figure 4* this “GMM early response” class is largely overlapping with the positive response groups identified via clinically significant change criteria. All patients in the *`GMM - early positive change´* group also improved reliably from intake to session three. Clinically significant improved were 253 (63.9%) of these `GMM - early positive change´ patients. Generally, the GMM categorized much less patients as “early positive responders” than the rationale methods.

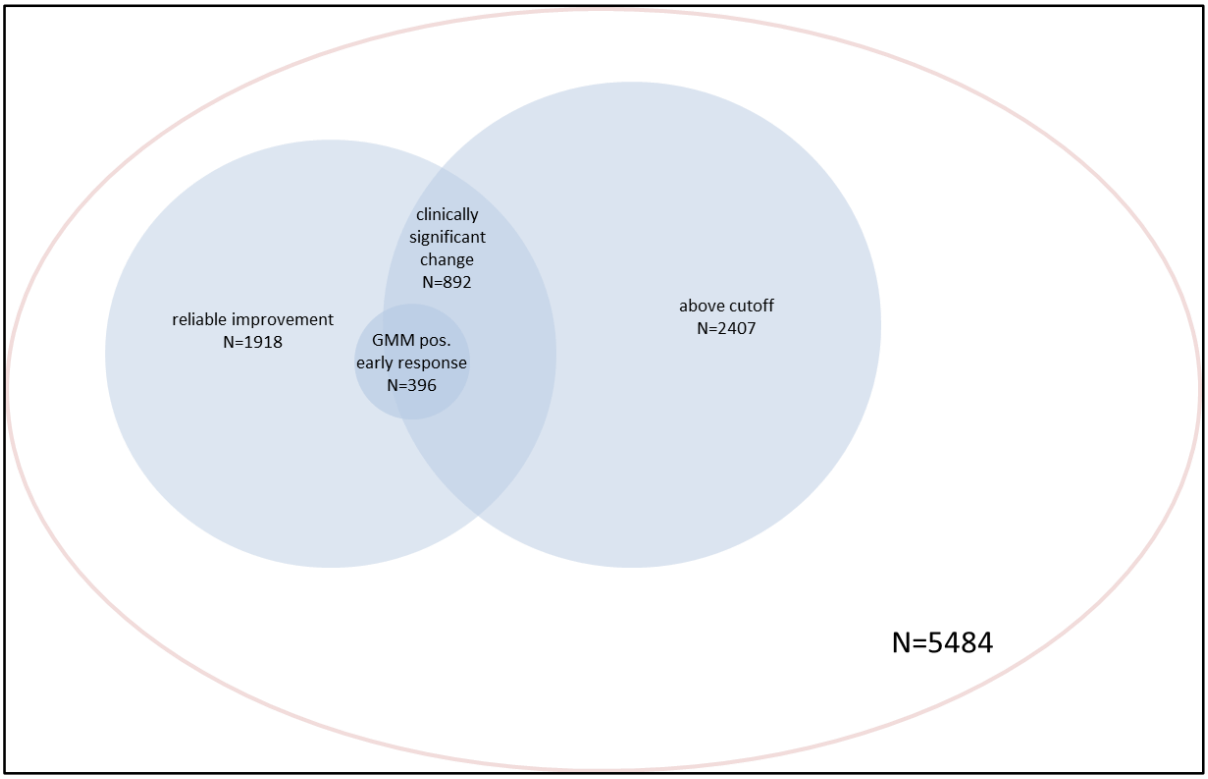

Figure 4. Venn diagram showing the number of patients in and the overlap between the three differentially identified early positive response groups.

While there were no differences between the three groups with regard to treatment length, the subgroup identified via GMM started treatment with significantly lower levels of psychological functioning and achieved significantly larger pre to post effect sizes (d) than the subgroups identified with clinically significant change criteria. These differences were stable over the course of treatment and independent of the number of sessions a patient received. Overall, patients from all three early positive response groups had better outcomes than the average patient. With regard to predictive accuracy, GMM provided the most specific predictions for reliable and clinically significant improvement from pre to post treatment, while at the same time being the most insensitive. Both clinically significant change methods are a little less specific than GMM but much more sensitive for the prediction of ultimate reliable change/clinically significant change.

7.4 Conclusions

GMM is a much more conservative method for the identification of early positive treatment response than clinically significant change criteria and mainly identifies patients starting with high levels of impairment. Although all of the patients identified via GMM were also identified via the reliable change method and most of the patients were identified via the clinically

significant change method (64%), the group of GMM early positive response patients showed to be highly specific for the prediction of treatment outcome.

Given that many patients, who show positive ultimate treatment outcome, are missed by the GMM method, it cannot be recommended as sole method for clinicians who want to evaluate the stability of early treatment responses of their patients. Rather, a stepwise approach using simpler clinically significant change criteria first as a screening and after that, the information provided by GMM should be proposed to clinicians. This stepwise approach allows an increasingly specific prediction of treatment outcome.

8 GENERAL DISCUSSION

The studies summarized in the current umbrella address three important limitations of past research on typical patterns of early change to psychotherapy:

Study I expands research on change patterns with regard to the investigated diagnostic group, the study design, and the examined predictors of early change patterns: For the first time, different patterns of change could be identified in a homogeneous sample of panic disorder patients treated with a manualized cognitive-behavioral treatment. The only other study investigating early change patterns in a randomized controlled psychotherapy trial focused on depressed patients (Lutz et al., 2009).

Study II expands prior research with regard to the instrument that was used as a measure of patient progress and with regard to the treatment stages, which were examined. For the first time, not a symptom measure, but a patient session report has been the basis for the identification of typical change patterns. Additionally, not only the early phase of the treatment but also two later stages have been separately analyzed.

Study III compares the characteristics of early response patients identified with GMM with similar groups identified with less computationally intensive methods (clinically significant change criteria).

In the following, some general conclusions that can be drawn from the three studies are summarized and their limitations discussed.

8.1 General conclusions:

Taken the three presented studies together, some general conclusions about early change patterns can be deduced in the context of past research:

Generally, the shapes of change, which can be observed in the early treatment phase, are relatively consistent between different studies. Regularly, two or three classes are identified whose patients show not much change and differ mainly with regard to initial impairment. Additionally, one class is consistently recognized, which shows rapid positive changes and in some studies also a class of deteriorating patients is found. Although the average change patterns do not vary much between studies, the rate of patients in the respective classes differ considerably.

Especially, some further knowledge could be gained with regard to the rate of patients that show substantial improvements in the first few sessions. Comparing the rates of early responding

patients in the current three studies with those from past research, there is considerable variability. Besides the different methodological criteria to define early improvement, several other factors could have contributed to this variation: Differences in the study design (RCT vs. naturalistic), the composition of the sample with regard to the predominant diagnostic group, the interval which is considered as early, and the instrument used to track patients change.

Focusing on differences in diagnostic groups, it could be noted that compared to Study I in which about 20% of the patients with panic disorders improved early, about 60% of the depressed patients in the study of Lutz et al. (2009) were classified as early responders. This large difference seems to suggest that the phenomenon of early response might more likely occur in patients with affective disorders than in patients with anxiety disorders. However, since these two studies were not only different with regard to the investigated diagnostic group, but also regarding the number of weeks considered as early, the treatment approach, and the specific modeling approach, conclusions must be drawn very cautionary.

Regarding the design of the study, results suggest a higher rate of early responding patients in controlled than in naturalistic settings. Again, the three studies do not allow a stringent test of this statement since study design was not systematically varied. However, Study I which used data from a controlled trial identified about 20%, while Study III which used data from routine care identified only about 7% early improving patients. Other investigations using data from naturalistic settings also report lower rates of early responding patients (12.5%; Stulz et al., 2007; 4.4%; Lutz et al., 2007). Notwithstanding the difference in the used instrument, Study II in which also data from naturalistic settings were used, similarly report a lower rate of 11.4% of patients who showed large positive changes in their session report scores in the first six treatment sessions.

With regard to the association of early change patterns and treatment outcome, results seem to be consistent. Subgroups with early positive changes were consistently associated with the most successful treatment outcomes. Especially those subgroups with early deteriorations were at risk of treatment failure (Lutz et al., 2007; Study I). Study I estimated that about 16% of the proportion of variation in patients outcomes could be explained by the early change pattern, beyond patients' intake information. In comparison, Lutz et al. (2009) found a slightly lower share of 8%.

Regarding the association between early change patterns and treatment length an interesting pattern is observed. While early improving patients tend to have shorter treatments than the other classes in naturalistic conditions (Study II; Stulz et. al., 2007; Lutz et al., 2007), they tend

to stay longer in treatment within RCTs (Study I). This differential effect might be connected to the differences between controlled trials and routine care with regard to the time limit of the treatments. While naturalistic psychotherapy often has an unlimited character, treatments in controlled trials are a priori limited to a specific number of sessions. Consequently, patients who improve early in routine care might need not as many sessions to reach their personal good-enough level (Barkham et al., 2006) and thus quit treatments earlier than other patients. In controlled trials early improving patients might be especially committed to adhere to the agreed upon number of sessions and thus stay in treatment for the complete trial rather than dropping out prematurely.

Concerning the differences between early positively responding patients defined via different methods, Study III suggests that the chosen method is an important factor influencing the number and kind of patients who are identified as early positive responders. All other determinants held constant, these different methods lead to very different patients which are identified as early positive responders. GMM showed to be a more conservative method which mainly identifies highly impaired patients. In practice, clinicians and researchers must decide in advance what they expect from the applied method. If viable, a stepwise approach combining different methods should be recommended. If sensitivity is crucial, easy and more liberal methods like reliable and clinically significant change are recommended. For future research it would also be interesting to compare the characteristics of these different methods for the identification of patients with early non-response.

Despite the strength of the current studies in addressing some limitations of past research, they also have some shortcomings, which are discussed in the following section.

8.2 General Limitations and future research

As already mentioned above, most of the drawn general conclusions result from comparisons between different studies. However, these studies do not only differ with regard to one specific aspect, which would allow a stringent conclusion with regard to this difference. Rather these aspects are confounded with other differences hampering clear-cut conclusions (e.g., Shapiro, 1985). Consequently, comparisons between different studies should always be exercised with caution. Future research needs to address these issues with studies that systematically manipulate different aspects regarding early change patterns independently from each other. Considering for example the different rate of early improving patients that suffer from panic disorder (Study I) and depressed patients (Lutz et al., 2009): To draw reliable conclusions about differences in the early response rates, which are connected to the psychopathology of the

patients, a study would be needed in which the same analyses are applied to comparably distressed patient samples who differ in their psychological disorder (e.g., depression vs. panic disorder), were treated with a comparable treatment, and whose treatment response were tracked on a comparable outcome dimension. Besides replication, only such a study could strengthen the preliminary evidence drawn from comparisons between different studies.

Another limitation of the current studies as well as of all the applications that have dealt with early patterns of change to psychotherapy is connected to the complete neglection of therapist differences. Therapist effects have been discussed repeatedly as an important factor explaining the differences in patients' outcomes. The most recent meta-analysis estimates that about 5-7% of the differences in outcomes are due to therapist differences (Baldwin & Imel, 2013). Given the stability of early changes and the resulting high correlations between early changes and treatment outcome, it must be assumed, that therapist differences might also manifest in early change patterns. Future research should consider this factor. One possibility to take the therapist variable into account might be multilevel mixture modeling (Asparouhov & Muthén, 2008). Interesting questions to investigate would be if there are systematic differences between therapists in the probability with which their patients show early improvements and if those therapists who have a high probability to induce early response are also more successful at the end of the treatment.

Another limitation is connected to the observational character of the current studies. Although predictors of class membership and in turn relations of class membership with treatment outcome and length were investigated, nothing can be said about the causes of the different change curves. Due to the fact that no control group of patients who did not receive any treatment was analyzed, we even could not rule out that the observed patterns simply reflect different paths of spontaneous recovery. However, everyday clinical practice is neither controlled for patients who do not improve because of the psychotherapeutic treatment itself but because of some kind of self-healing mechanism collapsing by coincidence with the early treatment phase. Future research should have a closer look at the mechanisms underlying the phenomena of early response and non-response.

Generally, it should be noted that outcome monitoring and feedback systems including decision rules are not supposed to substitute the clinical decision of the therapist, which is based on their clinical impression. Rather, it should be considered as a helpful support tool for clinicians in their decision making process. As such, these tools are just one of several aspects and are not designed to substitute the clinical impression of the therapists. In contrary, McAleavy,

Nordberg, Kraus, and Castonguay (2014) emphasize potential limitations of feedback and sensitize clinicians not to accept the trajectories and predictions as a reality, which is deterministically connected with a certain outcome. Consequently, it seems to be important to train therapists in the integration of feedback information in their everyday clinical practice. However, McAleavy et al. (2014) also conclude that once these caveats are managed, feedback systems and their extensions are one of the most promising ways to bring research results into clinical practice and further improve the effects of psychotherapeutic interventions.

8.3 Concluding remarks

Despite these limitations, the presented three studies expand the knowledge on early change patterns, which could be provided to therapists as additional information within outcome monitoring and feedback systems. As such, findings from this research could support clinicians' decision-making at different junctions.

Already before the actual treatment therapists can use the identified predictors of treatment response to more precisely estimate the probability with which a specific patient is about to show a respective change pattern (Study I). Although these predictors were mainly identified for patients with panic disorders, past research suggests that intake predictors are relatively stable across patients' diagnoses and thus could be generalized to other diagnoses (Lueger, Howard, Martinovich, Lutz, Anderson, & Grissom, 2001).

Within the early phase of the treatment therapists could use simple clinically significant change criteria to identify early response in their patients. Since Study III showed how reliable early changes identified via this easily applied method predict treatment outcome, this information allows clinicians to better foresee the future course of their patients.

Especially in an early treatment phase also the change trajectories of patients progress measures could provide therapists with further information on the future treatment course (Study II). Although less informative, also typical trajectories in a session report measure from later phases of the treatment are provided with which therapists can compare the scores of their patients.

The integration of empirical findings on patients change courses, and their relations to potential predictors and treatment outcome within computerized feedback systems is one way to build a closer link between research and practice. Both, practitioners and researchers, would profit a great deal from an extension of the use of psychometric outcome monitoring in routine care (e.g., Lutz & Rubel, 2015). Gradually, this kind of research could be an integral part of everyday clinical practice, while at the same time data from routine care is

continuously generated for future research. At the end of this development, the often complained gap between research and practice could in this way be transformed into a symbiotic self-reinforcing system.

9 REFERENCES

Aaronson, C. J., Shear, M. K., Goetz, R. R., Allen, L. B., Barlow, D. H., White, K. S., ... & Gorman, J. M. (2008). Predictors and time course of response among panic disorder patients treated with cognitive-behavioral therapy. *Journal of Clinical Psychiatry, 69*(3), 418-424.

Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(1), 93-101.

Atzil-Slonim, D., Bar-Kalifa, E., Rafaeli, E., Lutz, W., Rubel, J., Schiefele, A. K., & Peri, T. (2015). Therapeutic Bond Judgments: Congruence and Incongruence. *Journal of Consulting and Clinical Psychology*. Advance online publication. doi: 10.1037/ccp0000015

Baldwin, S. A. & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield´s Handbook of Psychotherapy and Behavior Change* (6th ed.), (pp. 258-297). New York: John Wiley & Sons.

Barkham, M., Connell, J., Stiles, W. B., Miles, J. N., Margison, F., Evans, C., & Mellor-Clark, J. (2006). Dose-effect relations and responsive regulation of treatment duration: the good enough level. *Journal of Consulting and Clinical Psychology, 74*(1), 160.

Barlow, D. H. & Craske, M. G. (2007). Mastery of your anxiety and panic. Oxford: Oxford University Press.

Bickman, L., Kelley, S. D., Breda, C., de Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services, 62*(12), 1423-1429.

Breslin, F. C., Sobell, M. B., Sobell, L. C., Buchan, G., & Cunningham, J. A. (1997). Toward a stepped care approach to treating problem drinkers: the predictive utility of within-treatment variables and therapist prognostic ratings. Addiction, 92(11), 1479-1489.

Castonguay, L. G., Barkham, M., Lutz, W., & McAleavey, A. 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: John Wiley & Sons.

Cinar, S., Voshaar, R. O., Janzing, J. G. E., Birkenhäger, T. K., Buitelaar, J. K., & van den Broek, W. W. (2010). The course of depressive symptoms in unipolar depressive disorder during electroconvulsive therapy: A latent class analysis. Journal of Affective Disorders, 124(1), 141-147.

Cuijpers, P., van Lier, P. A., 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*(1), 137-146.

Davies, D. R., Burlingame, G. M., Johnson, J. E., Gleave, R. L., & Barlow, S. H. (2008). The effects of a feedback intervention on group process and outcome. *Group Dynamics: Theory, Research, and Practice, 12*(2), 141-154.

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

Eysenck, H. J. (1952). The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology, 16*(5), 319-324.

Flückiger, C., Grosse Holtforth, M., Znoj, H. J., Caspar, F., & Wampold, B. E. (2013). Is the relation between early post-session reports and treatment outcome an epiphenomenon of intake distress and early response? A multi-predictor analysis in outpatient psychotherapy. *Psychotherapy Research, 23*(1), 1-13.

Gildengers, A. G., Houck, P. R., Mulsant, B. H., Dew, M. A., Aizenstein, H. J., Jones, B. L., ... & Reynolds III, C. F. (2005). Trajectories of treatment response in late-life depression: psychosocial and clinical correlates. *Journal of Clinical Psychopharmacology, 25*(4), S8-S13.

Grawe, K. (1992). Psychotherapieforschung zu Beginn der neunziger Jahre [Psychotherapy research at the beginning of the nineties]. *Psychologische Rundschau, 43*(3), 132-162.

Grawe, K. (1997). Research-informed psychotherapy. *Psychotherapy Research, 7*(1), 1-19.

Grilo, C. M., Masheb, R. M., & Wilson, G. T. (2006). Rapid response to treatment for binge eating disorder. *Journal of Consulting and Clinical Psychology, 74*(3), 602-613.

Haas, E., Hill, R. D., Lambert, M. J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? *Journal of Clinical Psychology, 58*(9), 1157-1172.

Hafner, J. & Marks, I. M. (1976). Exposure in vivo of agoraphobics: Contributions of diazepam, group exposure and anxiety evocation. *Psychological Medicine, 6*(1), 71– 88.

Hamilton, M. (1959). The assessments of anxiety states by rating. *British Journal of Medical Psychology, 32*(1), 50-55.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23*(1), 56-62.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*(2), 155-163.

Hatfield, D., McCullough, L., Frantz, S. H. B., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists´ability to detect negative client change. *Clinical Psychology and Psychotherapy, 17*(1), 25-32.

Hilbert, A., Saelens, B. E., Stein, R. I., Mockus, D. S., Welch, R. R., Matt, G. E., & Wilfley, D. E. (2007). Pretreatment and process predictors of outcome in interpersonal and cognitive behavioral psychotherapy for binge eating disorder. *Journal of Consulting and Clinical Psychology, 75*(4), 645-651.

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51*(10), 1059-1064.

Jacobson, N. S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19.

Kazdin, A. E. (1986). Comparative outcome studies of psychotherapy: methodological issues and strategies. *Journal of Consulting and Clinical Psychology, 54*(1), 95-105.

Kazdin, A. E. (2014). Moderators, mediators, and mechanisms of change in psychotherapy. In W. Lutz & S. Knox (ed.). *Quantitative and Qualitative Methods in Psychotherapy* (pp. 87-101). New York, NY: Routledge.

Keller, F., & Hautzinger, M. (2007). Klassifikation von Verlaufskurven in der Depressionsbehandlung [Classification of change trajectories in the treatment of depression]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 36*(2), 83-92.

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*(4), 413–426.

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: John Wiley & Sons.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69*(2), 159-172.

Lambert, M.J., Rubel, J., & Lutz, W. (in press). Der Outcome-Questionnaire (OQ-45.2) [The outcome-questionnaire (OQ-45)]. In *Diagnostische Verfahren in der Psychotherapie*. Göttingen: Hogrefe.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it Time for Clinicians to Routinely Track Patient Outcome? A Meta-Analysis. *Clinical Psychology: Science and Practice*, *10*(3), 288-301.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767–778.

Lueger, R. J., Howard, K. I., Martinovich, Z., Lutz, W., Anderson, E. E., & Grissom, G. (2001). Assessing treatment progress of individual patients using expected treatment response models. *Journal of Consulting and Clinical Psychology, 69*(2), 150-158.

Lutz, W. (2002). Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirically based clinical practice. *Psychotherapy Research, 12*(3), 251-272.

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*(3), 311-317.

Lutz, W., De Jong, K., & Rubel, J. (2015). Patient-focused and feedback research in psychotherapy: Where are we and where do we want to go? Psychotherapy Research. Advance online publication. doi: 10.1080/10503307.2015.1079661

Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M. A., Jorasz, C., ... & Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research, 23*(1), 14-24.

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*(2), 287-297.

Lutz, W., Martinovich, Z., & Howard, K. I. (1999). Patient profiling: An application of random coefficient regression models to depicting the response of a patient to outpatient psychotherapy. *Journal of Consulting and Clinical Psychology, 67*(4), 571.

Lutz, W. & Rubel, J. (2015). Warum die Psychotherapieforschung in Deutschland ausgebaut werden sollte/könnte/müsste [Why psychotherapy research should and can be expanded in Germany]. *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie, 65*(5), 161-162.

Lutz, W. & Rubel, J. (2013a). Psychotherapieforschung [Psychotherapy research]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie,* 17. Aufl. (S. 1356). Bern: Hans Huber Verlag.

Lutz, W. & Rubel, J. (2013b). Psychotherapieforschung [Psychotherapy research]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie,* 16. Aufl. (S. 1268). Bern: Hans Huber Verlag.

Lutz, W., Rubel, J., & Böhnke, J. R. (in press). Der Fragebogen zur Evaluation von Psychotherapieverläufen (FEP) [The questionnaire for the evaluation of psychotherpay (QEP)]. In *Diagnostische Verfahren in der Psychotherapie*. Göttingen: Hogrefe.

Lutz, W., Rubel, J., Schiefele, A.-K., Zimmermann, D. Böhnke, J., & Wittman, W. W. (2015). Feedback and Therapist effects in the context of treatment outcome and treatment length. *Psychotherapy Research*. Advance online publication. doi: 10.1080/10503307.2015.1053553

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*(1), 60-68.

Lutz, W., Stulz, N., Smart, D. W., & Lambert, M. J. (2007). Die Identifikation früher Veränderungsmuster in der ambulanten Psychotherapie [The identification of early change patterns in outpatient psychotherapy]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 36*(2), 93-104.

Lutz, W., Wittmann, W. W., Böhnke, J. R., Rubel, J., & Steffanowski, A. (2012). Zu den Ergebnissen des Modellprojektes der Techniker-Krankenkasse zum Qualitätsmonitoring in

der ambulanten Psychotherapie aus Sicht des wissenschaftlichen Evaluationsteams [Results from the pilot project of the Techniker Krankenkasse (TK) 'quality monitoring in outpatient psychotherapy': the evaluators' perspective]. *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie, 62*(11), 413-417.

Lutz, W., Wittmann, W., Böhnke, J., Rubel, J., & Steffanowski, A. (2013). Das Modellprojekt der TK zum Qualitätsmonitoring in der ambulanten Psychotherapie aus Sicht der Evaluatoren–Ein Plädoyer für mehr Psychotherapieforschung in Deutschland [The TK-Project Quality Monitoring in Outpatient Psychotherapy from the Perspective of the Evaluation Team-A Plea for More Psychotherapy Research in Germany]. *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie, 63*(6), 225-228.

Manicavasagar V., Silove D., Wagner R., & Hadzi-Pavlovic D. (1999). Parental representations associated with adult separation anxiety and panic disorder agoraphobia. *Australian and New Zealand Journal of Psychiatry, 33*(3), 422– 428.

McAleavey, A. A., Nordberg, S. S., Kraus, D., & Castonguay, L. G. (2012). Errors in treatment outcome monitoring: Implications for real-world psychotherapy. *Canadian Psychology, 53*(2), 105-114.

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry, 180*(5), 461–464.

Muthén, B. (2006). The potential of growth mixture modelling. *Infant and Child Development, 15*(6), 623–625.

Muthén, B. & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, & G. Molenberghs (Eds.), *Longitudinal Data Analysis*, (pp. 143-165). Boca Raton, FL: Taylor & Francis.

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

Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: a group-based method. *Psychological Methods, 6*(1), 18-34.

Norcross, J. C., Karg, R., & Prochaska, J. (1990). 0.(1997). Clinical psychologists in the 1990's, II. *The Clinical Psychologist, 50*(3), 4-11.

Nordberg, S. S., Castonguay, L. G., Fisher, A. J., Boswell, J. F., & Kraus, D. (2014). Validating the rapid responder construct within a practice research network. *Journal of Clinical Psychology, 70*(9), 886-903.

Nylund, K.L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and Growth Mixture Modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535–569.

Pickles, A. & Croudace, T. (2010). Latent mixture models for multivariate and longitudinal outcomes. *Statistical Methods in Medical Research, 19*(3), 271-289.

Pöhlmann, P. D. K., Altmann, U., Butter, A., Israel, M., Keller, A., Schellong, J., ... & Joraschky, P. (2008). Differenzielle Therapieverläufe in der stationären Psychotherapie: Eine störungsübergreifende clusteranalytische Differenzierung von Verlaufstypen auf der Basis des Merkmals Depressivität [Differential treatment courses in inpatient psychotherapy: A transdiagnostic clusteranalytic differentiation of change types on the basis of depressive symptoms]. In *Prävention und Versorgungsforschung* (pp. 195-216). Berlin: Springer.

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 & Adolescent Psychiatry, 37*(11), 1184-1190.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry, 6*(3), 412-415.

Seligman, M. E. (1995). The effectiveness of psychotherapy: the Consumer Reports study. *American Psychologist, 50*(12), 965-974.

Smith, M. L. & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*(9), 752-760.

Rubel, J., Ehrlich, T., & Lutz, W. (in press). Feedback-Ansätze zur Reduktion des Misserfolgsrisikos und zur Optimierung von Psychotherapie [Feedback appraoches for the reduction of treatment failure and the enhancement of psychotherapy]. *Psychotherapie im Dialog – PiD.*

Rubel, J. & Lutz, W. (2013a). Wachstumsanalyse [Growth Analysis]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie (*17th edition), (p. 1775). Bern: Hans Huber Verlag.

Rubel, J. & Lutz, W. (2013b). Phasenmodell psychotherapeutischer Veränderung [Phase model of psychotherapeutic change]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie (*17th edition), (p. 1262). Bern: Hans Huber Verlag.

Rubel, J. & Lutz, W. (2013c). Wachstumsanalyse [Growth Analysis]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie (*16th edition), (p. 1653). Bern: Hans Huber Verlag.

Rubel, J. & Lutz, W. (2013d). Phasenmodell psychotherapeutischer Veränderung [Phase Model of Psychotherapeutic Change]. In M.A. Wirtz (Ed.), *Dorsch. Lexikon der Psychologie (*16th edition), (p. 1187). Bern: Hans Huber Verlag.

Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2014). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. *Psychological Assessment, 27*(2), 478-488.

Rubel J., Lutz W., & Schulte D. (2015). Patterns of change in different phases of outpatient psychotherapy: A stage-sequential pattern analysis of change in session reports. *Clinical Psychology & Psychotherapy, 22*(1), 1–14.

Rynn, M., Khalid- Khan, S., Garcia- Espana, J. F., Etemad, B., & Rickels, K. (2006). Early response and 8week treatment outcome in GAD. *Depression and Anxiety, 23*(8), 461-465.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464.

Shapiro, D. A. (1985). Recent applications of meta-analysis in clinical research. Clinical *Psychology Review, 5*(1), 13-34.

Stulz, N., Gallop, R., Lutz, W., Wrenn, G. L., & Crits-Christoph, P. (2010). Examining differential effects of psychosocial treatments for cocaine dependence: An application of latent trajectory analyses. *Drug and Alcohol Dependence, 106*(2), 164-172.

Stulz, N. & Lutz, W. (2007). Multidimensional patterns of change in outpatient psychotherapy: The phase model revisited. *Journal of Clinical Psychology, 63*(9), 817-833.

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*(4), 593-600.

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*(6), 864-874.

Strauss, B. M., Lutz, W., Steffanowski, A., Wittmann, W. W., Boehnke, J. R., Rubel, J., ... & Kirchmann, H. (2015). Benefits and challenges in practice-oriented psychotherapy research in Germany: The TK and the QS-PSY-BAY projects of quality assurance in outpatient psychotherapy. *Psychotherapy Research, 25*(1), 32-51.

Sunderland, M., Wong, N., Hilvert-Bruce, Z., & Andrews, G. (2012). Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. *Behaviour Research and Therapy, 50*(6), 374-380.

Tadić, A., Helmreich, I., Mergl, R., Hautzinger, M., Kohnen, R., Henkel, V., & Hegerl, U. (2010). Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. *Journal of Affective Disorders, 120*(1), 86-93.

Tang, T. Z. & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(6), 894-904.

Uher, R., Muthén, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., ... & McGuffin, P. (2010). Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine, 40*(08), 1367-1377.

van Calker, D., Zobel, I., Dykierek, P., Deimel, C. M., Kech, S., Lieb, K., ... & Schramm, E. (2009). Time course of response to antidepressants: predictive value of early improvement and effect of additional psychotherapy. *Journal of Affective Disorders, 114*(1), 243-253.

Walfish, S., McAlister, B., O'Donnell, P. A. U. L., & Lambert, M. J. (2012). An Investigation of Self-Assessment Bias in Mental Health Providers. *Psychological Reports*, *110*(2), 639-644.

Wampold, B. E. & Imel, Z. E. (2015). The great psychotherapy debate: The evidence for what makes psychotherapy work (2nd ed.). New York: Routledge.

White, K. S., Allen, L. B., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2010). Attrition in a multicenter clinical trial for panic disorder. *The Journal of Nervous and Mental Disease, 198*(9), 665–671.

White, K. S., Payne, L. A., Gorman, J. M., Shear, M. K., Woods, S. W., Saska, J. R., & Barlow, D. H. (2013). Does maintenance CBT contribute to long-term response of panic disorder? A randomized controlled clinical trial. *Journal of Consulting and Clinical Psychology, 81*(1), 47-57.

10 FIGURES

Figure 1: Simplified path model of the applied GMM for the determination of latent change patterns over the first six treatment sessions…….21

Figure 2: Mean latent growth curves (early change classes; C1-C4) within the first five sessions of patients with panic disorder treated with CBT……...25

Figure 3: Typical change patterns in patients‘ subjectively perceived mastery and clarification experiences in three phases (sessions 1-6, 7-12, and 13-18) of outpatient psychotherapy…...29

Figure 4: Venn diagram showing the number of patients in and the overlap between the three differentially identified early positive response groups……34

11 ORIGINAL PUBLICATIONS

Article 1

Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M., 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*****(2), 287-297.**

Journal of Consulting and Clinical Psychology
2014, Vol. 82, No. 2, 287–297

0022-006X/14/$12.00 DOI: 10.1037/a0035535

Patterns of Early Change and Their Relationship to Outcome and Early Treatment Termination in Patients With Panic Disorder

Wolfgang Lutz
University of Trier

Stefan G. Hofmann
Boston University

Julian Rubel
University of Trier

James F. Boswell
University at Albany, State University of New York

M. Katherine Shear
Columbia University

Jack M. Gorman
Franklin Behavioral Health Consultants

Scott W. Woods
Yale University

David H. Barlow
Boston University

Objective: Recently, innovative statistical tools have been used to model patterns of change in psychological treatments. These tools can detect patterns of change in patient progress early in treatment and allow for the prediction of treatment outcomes and treatment length. ***Method:*** We used growth mixture modeling to identify different latent classes of early change in patients with panic disorder ($N = 326$) who underwent a manualized cognitive-behavioral treatment. ***Results:*** Four latent subgroups were identified, showing clusters of change trajectories over the first 5 sessions. One of the subgroups consisted of patients whose symptoms rapidly decreased and also showed the best outcomes. This information improved treatment prediction by 16.1% over patient intake characteristics. Early change patterns also significantly predicted patients' early treatment termination. Patient intake characteristics that significantly predicted class membership included functional impairment and separation anxiety. ***Conclusions:*** These findings suggest that early treatment changes are uniquely predictive of treatment outcome.

Keywords: patterns of early change, panic disorder, patient-focused research, randomized controlled trial, growth mixture modeling

Although cognitive-behavioral therapies (CBTs) for the treatment of panic disorder with or without additional agoraphobia are effective (e.g., Aaronson et al., 2008; Barlow & Craske, 2007; Hofmann & Smits, 2008), there is still considerable room for improvement. Further knowledge about individual differences in patterns of change for specific subgroups of patients might enable researchers and clinicians to maximize treatment outcome in individual patients (Barlow, 2010; Lambert, 2007; Lutz, 2002).

The investigation of patterns of change in psychological treatments has recently emerged as a topic in the research literature. Different methods have been applied to isolate subgroups of patients with similar treatment response. A common strategy for the identification of patient subgroups has been the classification on the basis of pre- to posttreatment comparisons (e.g., Aaronson et al., 2008). This approach defines change patterns on the basis of the assumption that those patients who experience positive outcome at the end of the treatment have followed a positive response pattern. However, those who experience negative outcome at the end of the treatment have followed a negative path. As a result, groups are combined that reach similar treatment outcome criteria, despite potentially very distinct treatment courses (Morral, Iguchi, Belding, & Lamb, 1997). The pathways to improvement could be very diverse, and this diversity might be clinically meaningful and might be a useful predictor of treatment outcome.

This article was published Online First January 20, 2014.

Wolfgang Lutz, Department of Psychology, University of Trier, Trier, Germany; Stefan G. Hofmann, Department of Psychology, Boston University; Julian Rubel, Department of Psychology, University of Trier, Trier, Germany; James F. Boswell, Department of Psychology, University at Albany, State University of New York; M. Katherine Shear, Department of Psychology, Columbia University; Jack M. Gorman, Franklin Behavioral Health Consultants, Bronx, New York; Scott W. Woods, Department of Psychology, Yale University; David H. Barlow, Department of Psychology, Boston University.

This work was supported by National Institute of Mental Health Grants R01 MH45963 (Jack M. Gorman, principal investigator [PI]), MH45964 (M. Katherine Shear, PI), MH45965 (David H. Barlow, PI), and MH45966 (Scott W. Woods, PI). Wolfgang Lutz is supported by grants from the Humboldt Foundation (TransCoop Program, TCVERL-DEU/1133562). Stefan G. Hofmann is supported by NIH Grant AT007257.

Correspondence concerning this article should be addressed to Wolfgang Lutz, Clinical Psychology and Psychotherapy, Department of Psychology, University of Trier, D-54286 Trier, Germany. E-mail: wolfgang.lutz@uni-trier.de

These individual differences in treatment change might reflect different change mechanisms and processes (Kazdin, 2007). Such pattern recognition based on the similarities of the change trajectories shared by a group of patients can be identified via growth mixture modeling (GMM) (e.g., Muthén, 2006; Nagin & Odgers, 2010). GMM is an advanced cluster analytic method that allows categorizing individuals into subgroups following similar change trajectories over a defined time period. It isolates groups of patients with similar treatment response patterns or profiles over time. Although GMM is still a relatively new method, it has already stimulated much research in clinical psychology and other areas of the social sciences (Nagin & Odgers, 2010). For example, Cuijpers, van Lier, van Straten, and Donker (2005) compared CBT with a treatment-as-usual control group in a sample of depressed patients using GMM. The average patient trajectories in the two conditions were statistically equivalent. However, large differences between the two treatments were detected for subgroups who were identified with GMM. For two highly impaired subgroups, CBT was substantially more effective than treatment as usual ($d = 0.75$ and 0.86). Additional studies have used a similar approach in different settings and disorders (e.g., Stulz, Gallop, Lutz, Wrenn, & Crits-Christoph, 2010). However, these and other studies have only classified patients on the basis of shared response curves over the entire treatment period. The objective of the present study was to predict outcome for specific subgroups on the basis of patients' change patterns during the early stage of treatment using GMM.

Studies on early response in psychological treatments have tracked treatment response on a weekly basis and revealed that quick, positive treatment response within the first three sessions related to positive outcome at termination and follow-up (e.g., Haas, Hill, Lambert, & Morrell, 2002). Furthermore, Haas et al. (2002) found that early responders needed fewer sessions to achieve stable improvement and were, therefore, more likely to terminate treatment earlier. Early response has been shown to be a powerful predictor of outcome in different samples (e.g., adolescents, adults, and elderly patients; e.g., Gunlicks-Stoessel & Mufson, 2011), modalities (psychological and pharmacological treatments; e.g., Hofmann, Schulz, Meuret, Moscovitch, & Suvak, 2006; Uher et al., 2010; van Calker et al., 2009), and diagnostic groups (depression, anxiety, and eating disorders; e.g., Aderka, Nickerson, Bøe, & Hofmann, 2012; Gunlicks-Stoessel & Mufson, 2011; Lutz, Stulz, & Köck, 2009). This literature has been primarily concerned with identifying patients who show rapid early improvement. Other change trajectories have not been the primary focus of research attention, although they might also have important clinical implications (see Lutz et al., 2013). Moreover, relatively few studies have used GMM in this context.

In one of the first studies investigating early change patterns with GMM, Stulz, Lutz, Leach, Lucock, and Barkham (2007) clustered 192 naturalistic outpatient psychotherapy patients on the basis of shared response trajectories over the first six sessions. Five distinct trajectory classes best described the individual early change courses; 11% of the patients showed a pattern that was characterized by high initial impairment and a rapid substantial improvement over the first six sessions. Over 90% of these early responders still remained substantially improved at the end of treatment. A second cluster consisting of 23.1% of patients was characterized by *high impairment* with little or no early change. Another cluster (14.6%) comprising little or no change over the first six sessions began treatment with a relatively low level of impairment. The remaining clusters were two moderately impaired groups with similar average growth curves, but very different individual treatment courses around the group mean trajectories. Whereas the patients in one group showed relatively continuous individual change trajectories (27.6%), patients' change courses in the other subgroup (23.6%) were characterized by higher levels of discontinuity, with higher variation in session-to-session assessments.

The first study in which a specific diagnostic group, major depression, was used was conducted by Lutz and colleagues (2009), using data from the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (TDCRP). The results of the completer data ($N = 162$) revealed three typical patterns of early change over the first 8 weeks of treatment, irrespective of the type of treatment protocol provided: (a) moderate to severe depression with moderate early improvement, (b) moderate to severe depression with rapid early improvement, and (c) mild to moderate depression with moderate early improvement. These differential patterns of early response (together with overall pretreatment symptom severity) predicted outcome (depression severity) at treatment termination and over the 1.5-year follow-up period.

In this study, we analyzed data from a multisite clinical trial examining long-term strategies in the treatment of panic disorder with and without agoraphobia. In the initial study phase, all patients were treated with CBT (Aaronson et al., 2008; White et al., 2010). On the basis of the rating of response status by a trained independent evaluator, patients were then triaged into two clinical trials. Responders were randomized to 9 months of monthly booster sessions or no booster sessions, and then followed for 1 year (White et al., 2013). Nonresponders were randomized to either 3 months of continued CBT or to 3 months of paroxetine, and then followed for an additional 9 months (Payne et al., 2012). In this study, data were analyzed from the initial treatment phase only where all patients received CBT. In the present study, the Panic Disorder Severity Scale–Self-Report version (PDSS-SR; Shear et al., 1997, 2001) was used to identify patterns of change over the course of treatment. On the basis of the existing literature, we predicted the existence of distinct patient response clusters. We further hypothesized that these clusters would predict treatment response and treatment length. Moreover, we examined patient intake characteristics as predictors of class membership.

Method

Participants and Treatment

The study was conducted at four sites (Center for Anxiety and Related Disorders at Boston University, Hillside/Long Island Jewish Hospital in New York, Western Psychiatric Institute and Clinic in Pittsburgh, and Yale University in New Haven) between November 1999 and July 2002. A total of 454 individuals were screened for inclusion and exclusion criteria. The first treatment session was attended by 381 patients who were screened positive for panic disorder and negative for active medical illness, substance abuse or dependence, bipolar disorder, or psychosis.

In this analysis, 326 patients were included who met the following inclusion criteria: attended the 11-session CBT treatment for at least three sessions and had symptom severity ratings (PDSS-SR) prior to the first session. The average age at the beginning of treatment was 37.0 years (SD = 11.9, range = 18–67), and 210 (64.4%) participants were women. Patients were predominantly White/Caucasian (87.7%), and about 5% each identified as African American (4.9%) or Asian American (5.5%). Further, six (1.8%) patients provided no information on their ethnicity. Treatment duration ranged from three to 11 sessions (M = 9.9, SD = 2.3).

All patients received a modified version of panic control treatment with the following elements (Barlow & Craske, 2007): education about the nature of anxiety and panic, identification and correction of maladaptive thoughts about anxiety and its consequences, interoceptive exposure, and graded exposure to avoided situations and activities. Treatment comprised up to 11 sessions, which were administered on a weekly (biweekly for the last two sessions) basis and over a maximum period of 18 weeks (due to missed visits). Primary outcomes from this trial (e.g., number of responders and nonresponders, attrition/dropout) were reported in Aaronson et al. (2008) and White et al. (2010), respectively (see Grilo et al., 1998; Hofmann et al., 1998, for similar reports on previous trials examining this treatment approach for panic disorder). Additionally, White et al. (2013) recently reported the primary outcomes for the study of acute treatment responders who were randomized to receive maintenance CBT or no maintenance. In the present investigation, we focused on session-by-session PDSS-SR ratings collected at each visit during the acute open-treatment phase of this trial.

Measures

The PDSS-SR (Shear et al., 1997). Patients completed the PDSS-SR at each treatment session. The seven PDSS-SR items measure the severity of panic disorder symptoms (frequency of panic attacks, distress during panic attacks, anticipatory anxiety, agoraphobic fear and avoidance, interoceptive fear and avoidance, impairment of or interference in work, and social functioning) on a 5-point Likert scale (0 = *no symptoms of panic*; 4 = *high symptoms of panic*). In the present study, the sum of these items was used. The test–retest reliability for the PDSS-SR is r_{tt} = .81, and internal consistency is α = .92 (Houck, Spiegel, Shear, & Rucci, 2002). The reliable change index (RCI)[1] for the PDSS-SR was calculated using the equation in Footnote 1, with the standard deviation and the internal consistency scores from an independent outpatient reference sample (Houck et al., 2002). The RCI score for the PDSS-SR was 5.17 total points. In the sample of N = 326 patients, four (1.2%) showed a reliable deterioration from the first to the last treatment session; 136 (50%) experienced no reliable improvement within the trial, and 159 (48.8%) patients reached an amount of improvement in PDSS-SR scores that was classified as reliable.

In the present study, the following measures, administered at intake and previously shown to contribute significantly to outcome (Aronson et al., 2008), were used as potential predictors of membership in one of the latent trajectory classes, and examined as predictors of treatment outcome and duration.

Hamilton Anxiety Rating Scale (HAM-A; Hamilton, 1959). The HAM-A is a clinician-administered scale with 14 items to measure severity of anxiety symptoms. Each of these items is scored from 0 (*not present*) to 4 (*severe*) to reflect the intensity of the particular symptom. The interrater reliability, measured by the intraclass coefficient for the HAM-A total scores of the two raters, is ICC = .74 (Maier, Buller, Phillip, & Heuser, 1988).

Hamilton Rating Scale for Depression (HAM-D; Hamilton, 1960). The HAM-D is a clinician-administered symptom severity measure with 17-items, each representing a depressive symptom. The interrater reliability, measured by the intraclass coefficient for the HAM-D total scores of the two raters, is r = .90 (Hamilton, 1960).

Adult Separation Anxiety-Checklist (ASAC; Manicavasagar, Silove, Wagner, & Hadzi Pavlovic, 1999). The ASAC measures symptoms of separation anxiety in adulthood, with 16 items on a 4-point scale ranging from 0 (*this has never happened*) to 3 (*this happens very often*). Item scores are summed to derive a total score. The ASAC has good internal consistency (Cronbach's α = 0.93; White et al., 2010).

Work and Social Adjustment Scale–Self-Report version (WSAS-SR; Mundt, Marks, Shear, & Greist, 2002). The WSAS-SR is a five-item self-report measure that assesses the degree of functional interference caused by symptoms in five life domains, on a scale ranging from 0 (*no interference*) to 8 (*very severe interference*); it is a modified version of the scale created by Hafner and Marks (1976). This measure demonstrated good internal consistency in the complete sample of the primary trial (Cronbach's α = 0.80; White et al., 2010).

Data Analytic Strategy

Latent patterns of early change in panic symptoms, measured with the PDSS-SR, over the first five sessions of CBT treatment were identified using GMM. There is no consensual definition in the literature on the number of sessions that can be regarded as "early" in psychological treatments. In the present investigation, the first five sessions were defined as the early treatment phase because all active ingredients of the treatment protocol had been delivered by the end of Session 5 (also consistent with Otto et al., 2012). In other words, no new additional treatment techniques were introduced after Session 5.

Subsequently, the identified latent change patterns were used to predict outcome at the end of treatment and early treatment termination, controlling for initial impairment in the outcome measure (PDSS-SR) and several predictor variables (HAM-A, HAM-D, ASAC, and the WSAS-SR, as well as age and gender) in multiple regression analysis. Therefore, the additional predictive information of early change in panic symptoms over intake vari-

[1] The RCI is defined as the pretreatment–posttreatment difference ΔRC that is large enough to be not attributable to measurement error and is calculated as:

$$\Delta_{RC} = z_{RC} \cdot \sqrt{2\left(SD\sqrt{1-r}\right)^2} = 1.96 \cdot \sqrt{2\left(6.6 \cdot \sqrt{1-.92}\right)^2} = 5.17,$$

where z_{RC} is the reliability of the instrument (i.e., 95% two-tailed), SD is the standard deviation of the PDSS-SR score in a sample of psychiatric outpatients, and r is the reliability (internal consistency) of the PDSS-SR (Houck et al., 2002).

ables could be evaluated. Because the purpose of the study was to evaluate the impact of early change on overall treatment response, we examined the effect of early change patterns on change from pre- to posttreatment.

Reliable change criteria as described above were used at the posttreatment assessment to classify patients into groups of *reliably improved* (pre- to postimprovement larger than the RCI of the PDSS-SR), *reliably deteriorated* (pre- to postdeterioration larger than the RCI), and *not reliably changed* (pre- to postchange that is not different from the RCI). Associations between these change categories and early change patterns were investigated with chi-square tests. To identify predictors of latent class membership, separate analyses of variance (ANOVAs) and multinomial logistic regression analyses were conducted. Additionally, the pretreatment variables described above were tested as potential predictors of treatment outcome and length of treatment.

GMM. GMM was used to identify latent subpopulations of patients with similar change trajectories within the first five sessions. Whereas in conventional latent growth models (e.g., hierarchical linear modeling) individual differences in intake scores (intercepts) and change parameters (slopes) are captured by the means of random coefficients (i.e., variation around a global mean intercept and slope), in GMM an additional categorical variable is included that accounts for sample heterogeneity (Duncan, Duncan, & Strycker, 2010; Muthén, 2006). Instead of assuming that there is only one underlying population with a single change pattern (intercept and slope), GMM allows investigators to test for an a priori unknown number of latent subpopulations that can differ in their mean intercepts and slopes (in the case of a linear model) as well as class-specific variations around these parameters.

In the final analysis, we fixed the variances around the class-specific slopes to zero, whereas intercept variances were freely estimated but constrained to be constant between classes.[2] Therefore, heterogeneity in change had to be captured completely by the difference in mean slopes of different latent classes; thus, we forced the estimation procedure to be more sensitive to patterns of change over time than to differences in initial level of impairment, which was the main interest. These models can be seen as a hybrid of models in which the variances of all parameters are fixed to zero (latent class growth models; Nagin, 1999), and models allowing all parameters to be freely estimated[3] (for similar approaches, see also Colder, Campbell, Ruel, Richardson, & Flay, 2002; Hunter, Muthén, Cook, & Leuchter, 2010; Uher et al., 2010).

Several fit criteria have been discussed to determine the optimal number of latent trajectory classes (e.g., Tofighi & Enders, 2007). For the present study, we used the Bayesian information criterion (BIC; Schwarz, 1978) and the bootstrapped likelihood ratio test (BLRT), following recommendations from simulation studies (e.g., Nylund, Asparouhov, & Muthén, 2007). Thus, the model determination process was twofold. In a first step, the model with the lowest value in the BIC was identified by starting to estimate a one-class model and successively adding one more class in each subsequent run. Once the BIC value did not decrease further from a model with k-classes to a model with k + 1 classes, this solution was then tested with the BLRT against a solution with k − 1 classes. If the BLRT revealed a significant p value ($p < .05$), the model was chosen as the best solution. If, however, the BLRT was not significant, the model was rejected and the solution with one class less (k − 1) was tested against a model with two classes less (k − 2). This procedure was repeated until the BLRT resulted in a significant p value.

In accordance with prior research on change trajectories in psychotherapy, a log-linear (base 10) transformation of the time scale was used for these analyses (Lutz et al., 2009; Stulz et al., 2007). Dose-effectiveness research has shown a consistent pattern in most analyses of rapid response early in therapy (e.g., Lambert, 2007). This consistent curvilinear pattern is parsimoniously approximated by a log-linear transformation of session number, as in the present case, and widely used in this area of research (e.g., Gibbons et al., 1993). However, an average log-linear shape of change does not imply that each patient has already reached his or her maximum change after five sessions, and no further change can be achieved after that.

An intercept-only linear model and a more complex quadratic model were compared with the chosen log-linear model using the BIC (intercept only: 8974.54, quadratic: 8776.69, log-linear: 8775.45), but evidenced a less good fit. The Mplus software package (Version 6.11; Muthén & Muthén, 2010) was used to estimate the GMMs.

Results

Early Change Patterns

Model fit indices are presented in Table 1. The BIC suggested a four-class solution, by steadily decreasing from one- through four-class solutions, and increasing from the four- to the fifth-class solution (see Table 1). The BLRT of this four-class solution against a three-class solution resulted in a significant p value ($p < .001$), indicating superiority of the model with four latent classes. Consequently, the four-class solution was used for further analyses.[4] Each participant was treated as being categorized in the latent class for which he or she showed the highest membership probability. Average membership probabilities ranged from 0.768 for patients categorized in Class 1 (C1) to 0.853 for patients categorized in Class 2 (C2) (see Table 2).

Figure 1 shows the estimated mean latent growth curves for the four classes of the above described solution within the first five sessions, and observed mean scores in these classes for Sessions 6–11. As shown in Figure 1, two classes were identified that were characterized by high panic symptom severity at intake on the PDSS-SR (Class 3 [C3]: $M = 16.84$; Class 4 [C4]: $M = 16.09$), clearly above the mean for psychiatric outpatients ($M = 9.0$;

[2] Models with varying intercept variances between classes were also tested and consistently resulted in unstable results with negative intercept variances. Therefore, the commonly applied assumption of equal intercept variances between classes was ultimately adopted.

[3] We also tested the more restrictive (latent class growth analysis) and more liberal (intercept and slope variances equal across all classes) models. Applying the same model identification criteria as the final model, the first resulted in eight and the latter in two latent change patterns. Both model solutions resulted in less clinically meaningful classes with lower predictive power than the final model.

[4] We reran the analyses on four randomly chosen subgroups of the complete sample (these subgroups contained 90%, 80%, 70%, 60%, and 50% of the patients in the complete sample) to test the reliability of the four-class solution. In each of these runs, the four-class solution was the best when applying the above-described procedures.

Table 1
Information Criteria, Entropy, and p Value in Bootstrapped Likelihood Ratio Test for up to Eight Latent Classes

# classes	BIC	SABIC	AIC	CAIC	Entropy	BLRT *p* value	LRT *p* value
1	8814.224	8788.848	8783.929	8784.383	—	—	—
2	8795.430	8760.539	8753.774	8754.615	0.521	.0000	.0073
3	8788.727	8744.319	8735.710	8737.060	0.595	.0000	.2195
4	8779.074	8725.151	8714.697	8716.684	0.686	.0000	.0103
5	8787.684	8724.245	8711.946	8714.700	0.721	.2667	.4320
6	8790.180	8717.226	8703.081	8706.737	0.733	.0128	.0584
7	8795.022	8712.551	8696.562	8701.258	0.749	.0300	.3556
8	8807.364	8715.378	8697.544	8703.422	0.740	.6667	.2209

Note. BIC = Bayesian information criterion; SABIC = sample-size-adjusted BIC; AIC = Akaike's information criterion; CAIC = corrected AIC; BLRT = bootstrapped likelihood ratio test; LRT = Lo-Mendel-Rubin likelihood ratio test. Dashes indicate that for single class models no scores are available.

Houck et al., 2002). The two other classes started treatment at low to medium panic symptom severity (C1: $M = 6.61$; C2: $M = 8.57$). All four mean class intercepts differed significantly from zero ($p < .001$). The latent patient subgroups differed with regard to their change trajectories over the first five sessions.

The first and smallest patient subgroup (C1: $n = 15$; 4.6%) was characterized by an increase in panic symptom severity between Sessions 1–5. The estimated mean slope (B) for patients within this class was significantly positive (B_{C1}: 8.48; $p = .001$), indicating that within each of the five sessions, an average increase in PDSS-SR scores was expected for these patients. The early change effect size in this group was negative ($d = -1.09$), reflecting early deterioration. Thus, this subgroup was labeled *early deterioration.* Nevertheless, after Session 5, the participants in this subgroup who did not drop out showed a continuous decrease in panic symptoms, yet only reached their original status of impairment by the end of treatment (see Figure 1).

The second subgroup comprised more than half of all study participants (C2: $n = 189$; 58%) and was characterized by moderate initial panic symptom severity and a slow but steady decrease within the first five sessions (B_{C2}: −3.24; $p < .001$). The early response effect size within this latent class was of medium magnitude ($d = .42$). This class was labeled *medium symptoms/slow change*.

A third subgroup of patients (C3: $n = 56$; 17.2%) started therapy at a high level of symptom severity and did not demonstrate significant change in severity within the first five sessions (B_{C3}: −1.74; $p = .103$). The average early response effect size of patients within this class was very low ($d = 0.22$). This class was named *high symptoms/no change*. However, a substantial decrease in PDSS-SR scores occurred between Sessions 6 and 11 (see Figure 1). The fourth subgroup (C4: $n = 66$; 20.2%) showed an early decrease in initially high panic symptom severity through the first five sessions (B_{C4}: −12.30; $p = .001$). This class showed the highest effect size in the first five sessions ($d = 1.58$) and was labeled *early response*. This improvement was maintained and extended during the second half of the treatment (see Figure 1).

Table 2
Average Latent Class Probabilities for Most Likely Latent Class Membership (Row) by Latent Class (Column)

Most likely latent class	Mean probability to belong to one of the latent classes			
	Class 1	Class 2	Class 3	Class 4
Class 1	0.768	0.077	0.131	0.025
Class 2	0.052	0.794	0.154	0.000
Class 3	0.043	0.071	0.853	0.033
Class 4	0.023	0.000	0.181	0.796

Prediction of Early Change Based on Patient-Intake Characteristics

Next, we investigated the relationships between membership in one of the four classes and patients' intake characteristics (age, HAM-D, HAM-A, ASAC, and WSAS-SR). First, differences in intake variables across the early change groups were investigated via separate ANOVAs. Using Bonferroni-corrected p values, each of these variables, with the exception of age, $F(3, 322) = 3.50$, $p = .080$, showed significant relationships with class membership. Intake values on the HAM-D, $F(3, 322) = 16.74$, $p < .001$; HAM-A, $F(3, 322) = 22.94$, $p < .001$; and ASAC, $F(3, 322) = 20.35$, $p < .001$, were significantly lower for C1 and C2 compared with C3 and C4. A similar pattern was observed for intake scores on the WSAS-SR, whereas C4 and C1 did not differ significantly, $F(3, 322) = 33.43$, $p < .001$.

When adding these significant variables in multinomial logistic regressions, only impairment in social functioning (WSAS-SR; $\chi^2 = 40.4$, $p < .001$) and separation anxiety (ASAC; $\chi^2 = 10.24$, $p = .017$) demonstrated specific predictive power for class membership, as suggested by significant chi-square values in likelihood ratio tests. Intake ASAC and WSAS-SR values accounted for an amount of explained variance in the differences in class membership of $R^2_{\text{Nagelkerke}} = 28.6\%$. Results of multinomial logistic regression analyses with WSAS-SR and ASAC as predictors of class membership are presented in Table 3. Intake WSAS-SR scores significantly discriminated C1 and C2 from C3 and C4. Higher WSAS-SR scores were associated with higher probabilities for membership in C3 and C4 compared with C1 and C2. With every unit increase in baseline WSAS-SR score, the probability of belonging to C3 rather than C1 increased by a factor of 2.25, and by a factor of 1.78 compared with C2. The probability of belonging to C4 rather than C1 increased by a factor of 2.14 with every unit increase in WSAS-SR and by a factor of 1.69 compared with C2. Intake values for separation anxiety only differentiated between C2 and C3. For every unit increase, patients' probability increased slightly by a factor of 1.06 to belong to C3 rather than to C2.

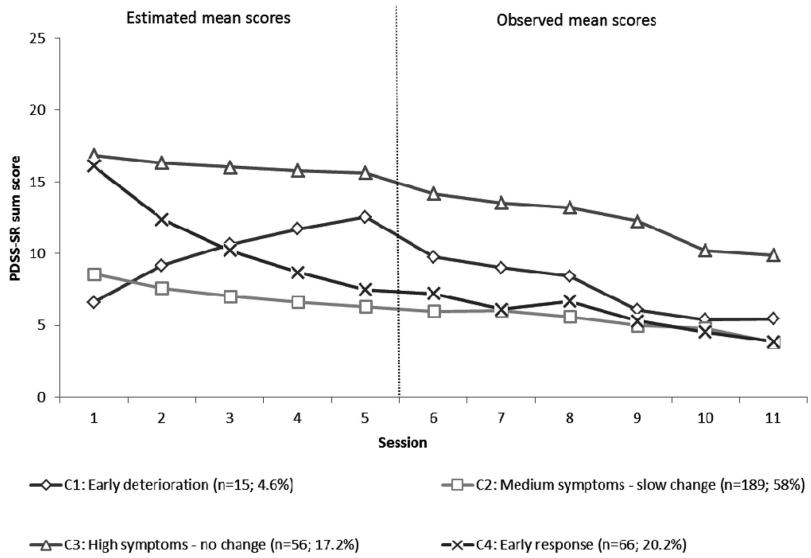

Figure 1. Mean latent growth curves for growth mixture modeling solution with four latent classes within the first five sessions, and observed mean scores in the respective classes during Sessions 6–11. PDSS-SR = Panic Disorder Severity Scale—Self-Report version; C1, C2, C3, and C4 = Class 1, Class 2, Class 3, and Class 4.

Early Change Patterns, Treatment Outcome, and Early Treatment Termination

Table 4 shows the relative frequencies of the PDSS-SR RCI *classes*, pre- to posteffect sizes of change in PDSS-SR (*d*), number of treatment sessions attended (three to five sessions, six to 10 sessions, and 11 sessions), and average treatment length for the different latent change classes. Each of these variables showed a significant association with class membership. The relationships between treatment response based on self-report measures, $\chi^2(3) = 77.79$, and treatment length, $\chi^2(6) = 16.80$, were analyzed with chi-square tests. One-way ANOVAs revealed significant associations between pre–post PDSS-SR change, $F(3, 322) = 69.05$, and mean treatment length, $F(3, 322) = 4.15$.

With regard to self-rated treatment outcome, the *early responder* group (C4) showed the highest rates of improvement (93.3%; standardized residual[5] = 5.3) and the largest amount of change in PDSS-SR scores from pre- to posttreatment ($d = 2.11$), twice the average effect size of $d = 1.02$. By contrast, none of the participants in the early deterioration group (C1) improved reliably on the PDSS-SR during treatment (standardized residual = −2.7), and these participants deteriorated from pre- to posttreatment ($d = -0.49$). In the largest class (C2), a minority of participants reported reliable improvements during treatment (37.6%; standardized residual = −2.2) with a mean pre- to posteffect size ($d = 0.74$) below the average effect size of the full sample. C3 showed high symptoms at intake and no change during the first five sessions, but demonstrated average improvement rates by the end of treatment (46.4%; standardized residual = −0.3) with the pre–post effect size approximately the same as the sample average ($d = 1.0$). These participants showed the biggest improvement in the second half of the treatment compared with all other classes.

Examining frequencies in the completer and early termination categories, most of the patients in our sample (76.7%) attended all of the scheduled 11 sessions, whereas 10.1% left treatment after three to five sessions, and further 13.2% left treatment prior to Session 11. The classes of participants showing *early deterioration* or no change (C1 and C3) were more likely to drop out of treatment early. In C1, 20% and in C3, 19.6% of the patients ended treatment prematurely between Sessions 3 and 5. These two classes also showed the highest rates of patients who ended treatment between the sixth and the 10th session (see Table 4). Approximately 60% of the patients in those two classes stayed in treatment until Session 11. The highest rate of treatment completion was found in the *early response* class (C4; 86.4%). Additionally, one-way ANOVA revealed significant associations between mean number of sessions and class membership. Tukey's post hoc tests revealed that C2 and C4 attended more treatment sessions than the *no-change* class (C3).

The predictive power of early change patterns for posttreatment change in PDSS-SR and treatment attendance was estimated via two separate hierarchical multiple regression analyses: one analysis for treatment outcome and one for treatment length. Each of these analyses was composed of three stages. In the first step, only the PDSS-SR score of the first session was included as a predictor. Second, all other available patient-intake characteristics that showed significant bivariate associations with outcome or duration

[5] The standardized residual represents the deviation of the observed frequency in a cell of a contingency table from the expected frequency:

$$standardized\ \ residual = \frac{\text{observed}_{ij} - \text{model}_{ij}}{\sqrt{model_{ij}}},$$

with observed_{ij} being the observed frequency in row i, column j and model_{ij} being the expected cell frequency. Standardized residuals are *z*-distributed and, thus, indicate significant deviations for values ± 1.96 ($p < .05$). A significant deviation indicates frequencies in that specific class that are higher than would be expected under the assumption of equal distributions and, thus, contribute to the significance of the overall chi-square test and the relevance of the detected class.

Table 3
Prediction of Group Membership by Patient-Intake Characteristics via Multinomial Logistic Regression Analyses

Variable	*B* (*SE*)	*p*	95% confidence interval for odds ratio		
			Lower	Odds ratio	Upper
Class 1 vs. Class 2					
Intercept	2.51 (0.58)	<.001			
ASAC	−0.03 (0.22)	.320	0.97	1.03	1.10
WSAS-SR	0.24 (0.03)	.279	0.52	0.79	1.21
Class 1 vs. Class 3					
Intercept	−1.69 (0.73)	.200			
ASAC	0.02 (0.04)	.507	0.96	1.02	1.10
WSAS-SR	0.81 (0.24)	.001	1.41	2.25	3.60
Class 1 vs. Class 4					
Intercept	−0.75 (0.67)	.262			
ASAC	0.00 (0.04)	.982	0.93	1.00	1.07
WSAS-SR	0.76 (0.23)	.001	1.35	2.14	3.38
Class 2 vs. Class 3					
Intercept	−4.20 (0.49)	<.001			
ASAC	0.06 (0.02)	.003	1.02	1.06	1.10
WSAS-SR	0.57 (0.12)	<.001	1.40	1.78	2.26
Class 2 vs. Class 4					
Intercept	−3.26 (0.40)	<.001			
ASAC	0.03 (0.02)	.053	1.00	1.03	1.07
WSAS-SR	0.52 (0.11)	<.001	1.36	1.69	2.10
Class 3 vs. Class 4					
Intercept	0.94 (0.54)	.082			
ASAC	−0.02 (0.02)	.239	0.94	0.98	1.02
WSAS-SR	−0.05 (0.13)	.679	0.74	0.95	1.21

Note. R^2 = .25 (Cox & Snell, 1989), .29 (Nagelkerke, 1991). Model $\chi^2(6) = 94.31$. For each comparison, the first mentioned class is used as the reference class in multinomial logistic regression. ASAC = Adult Separation Anxiety Checklist; WSAS-SR = Work and Social Adjustment Scale–Self-Report version.

were added in the model (HAM-D, HAM-A, ASAC, and WSAS-SR). After controlling for the influence of variables included in the first two steps, the dummy-coded class membership variables were added to the model in the last stage of the analysis.

As shown in Table 5, the first PDSS-SR score explained 31.2% of the variance in PDSS-SR treatment outcome. A higher PDSS-SR score at intake was associated with higher standardized pre- to posttreatment differences in this scale. A one standard deviation (SD = 5.45) higher initial PDSS-SR score corresponded to an increase in d of 0.56. An additional 3.7% of the variance in treatment outcome was explained by patient-intake characteristics included in the second step of the regression analysis. Besides the PDSS-SR score, the ASAC and the WSAS-SR were also significant pretreatment predictors of treatment outcome. For each of these two predictors, a higher score at intake was associated with lower effect sizes from pre- to posttreatment when controlling for the other variables included in Step 2 (see Table 5).

Adding the dummy-coded variables for class membership in Step 3 of the regression analysis resulted in a further increase of 16.4% explained variance in treatment outcome. Thus, a total amount of 51.1% of the variability in PDSS-SR change during treatment was explained by the model that contained all intake as well as early change predictors. Controlling for early change patterns (Step 3), higher PDSS-SR and lower WSAS-SR scores were still significantly related to greater treatment effects. A one standard deviation above the mean initial PDSS-SR score corresponded to an increase of 0.61 in d, a one standard deviation above the mean WSAS-SR score was related to an increase in effect size (d) of 0.14. Only the initial ASAC score, which was a significant predictor in Step 2, became nonsignificant in Step 3 of these analyses.

Results of the regression analysis for log-transformed (base 10) treatment duration as the dependent variable are also presented in Table 5. In the first step, no variance in treatment duration was explained by the PDSS-SR score at Session 1 as the only predictor variable. Adding patient-intake characteristics in the second step also did not lead to significantly explained variance in treatment length. Only the information on patterns of early change accounted for a small but significant amount of additionally explained variability (2.7%) in treatment length (Step 3).[6]

Discussion

The purpose of this study was to explore patterns of treatment response in the first five sessions of an 11-session treatment for panic disorder. We were able to characterize and predict typical patterns of change over the first five sessions in a sample of patients with panic disorder being treated with a standardized CBT protocol. Using GMM, four patterns of early treatment response were identified: a rapidly improving group, an initially highly symptomatic and slowly improving group, an initially low-symptom and slowly improving group, and an early deteriorating group. These early response profiles were predictive of treatment outcome and, to a lesser degree, the number of sessions attended. Initial level of patient-rated panic symptoms (PDSS-SR) and patient-intake characteristics predicted group membership.

Comparing these four average change trajectories with those reported in previous studies, the patterns resemble those described in other settings and samples: The two slowly changing subgroups, which differed predominantly by the level of initial impairment (C2 and C3), as well as the subgroup of early responders (C4), have been found consistently in prior studies investigating the early treatment phase with GMM (Lutz et al., 2009; Stulz et al., 2007). However, the size of the subgroups differed somewhat between studies. The subgroup of early improving patients (C4; 20.2%) was slightly larger than the one found in a naturalistic sample (12.5%; Stulz et al., 2007), and somewhat smaller than in a sample of depressed patients treated with medication (25%; Uher et al., 2010). This early improvement group was also smaller than the one found in a sample of depressed patients treated in the NIMH TDCRP study, independent of treatment condition (61.1%) (Lutz et al., 2009). One possible explanation for this could be the difference in how "early" treatment was defined. Although in the present study the first five sessions were examined, the first 8

[6] To evaluate the consequences of using patients with at least three to five sessions in the models, we reran all the analyses presented in Table 5 with only those individuals who completed six or more sessions and obtained consistent results.

Table 4
Relative Frequencies of Improvement Categories at Final Treatment Outcome, Reliable Change During Treatment on the PDSS-SR (Effect Sizes), and Treatment Length in Patient Groups of Early Change

		Final treatment outcome		Treatment completion status (number of sessions attended)			
Variable	*n*	Reliable improvement (%)	ES change in PDSS-SR during treatment (*d*) [95% CI]	3–5 (%)	6–10 (%)	11 (%)	Mean number
All patients	326	48.8	1.02 [0.85, 1.19]	10.1	13.2	76.7	9.87
Class 1	15	0*	−0.49 [−1.22, 0.26]	20	20	60	9.2
Class 2	189	37.6*	0.73 [0.51, 0.94]	6.9	14.3	78.8	10.04
Class 3	56	46.4	1.00 [0.58, 1.41]	19.6*	17.9	62.5	9.02
Class 4	66	93.3*	2.11 [1.61, 2.60]	9.1	4.5	86.4	10.29
p		<.001[a]	<.001[b]		<.001[a]		.007[b]

Note. PDSS-SR = Panic Disorder Severity Scale–Self-Report version; ES = effect size; CI = confidence interval.
[a] Chi-square tests were performed testing the association between class membership and categorized treatment outcome and length. [b] One-way analyses of variance were performed testing the association between class membership and mean pre- to postchange and mean treatment length.
* standardized residual > 1.96.

weeks were modeled in the NIMH TDCRP data set. Thus, in the latter study, patients had more time to achieve substantial improvement and be categorized as early improvers.

At this point, little is known about the reasons for rapid symptom reductions early in treatment. An interesting variable to investigate in the context of early treatment response might be readiness to change. Boswell et al. (2012), for example, found that readiness to change moderated the relationship between pre- and posttreatment levels of depression and anxiety. The highly negative association between levels of pretreatment impairment and pre-to-post change was reversed for patients with higher levels of readiness to change. All in all, the construct of readiness is a promising potential predictor of early treatment response and is worthy of further investigation.

Table 5
Hierarchical Multiple Regression Analyses Predicting Final Treatment Outcome (Standardized Pre- to Postdifferences) and Treatment Length (Log-transformed; Base 10) by Patient-Intake Characteristics and Patterns of Early Change

	Change during treatment (effect sizes)			Treatment length		
Predictor	ΔR^2	β	*p*	ΔR^2	β	*p*
Step 1	.31		<.001	.00		.362
First PDSS-SR		.56	<.001		−.05	.362
Step 2	.04		.001	.02		.121
First PDSS-SR		.71	<.001		.05	.536
HAM-D		.07	.448		−.09	.377
HAM-A		−.03	.765		−.01	.944
ASAC		−.15	.011		−.12	.087
WSAS-SR		−.15	.022		.01	.948
Step 3	.16		<.001	.03		<.05
First PDSS-SR		.61	<.001		.06	.554
HAM-D		−.01	.911		−.13	.234
HAM-A		.03	.689		.02	.835
ASAC		−.10	.063		−.10	.164
WSAS-SR		−.14	.016		.01	.926
C1 dummy		−.19	<.001		−.07	.215
C3 dummy		.20	<.001		−.14	.086
C4 dummy		.25	<.001		.05	.540
Total R^2	.51		<.001	.05		.026
n	324			324		

Note. C1 dummy, C3 dummy, and C4 dummy represent the dummy-coded class membership variable with Class 2 (*low symptoms − slow change*) used as reference class. PDSS-SR = Panic Disorder Severity Scale–Self-Report version; HAM-D = Hamilton Rating Scale for Depression; HAM-A = Hamilton Anxiety Rating Scale; ASAC = Adult Separation Anxiety Checklist; WSAS-SR = Work and Social Adjustment Scale–Self-Report version.

The *high symptoms/slow change* group was similar in size to that found in previous studies (17.2% compared with 13%–25% in previous studies). Conversely, the group of low-impaired, slowly changing patients was relatively large (58% compared with 10%–30% in previous studies). This might be due to the fact that different instruments were used in different studies.

Only two other studies in which change patterns with a cluster analytic method were investigated revealed a class of early deteriorating patients (Lutz et al., 2005; Morral et al., 1997), which we found in the present study. A possible explanation for the lack of early deteriorating groups in several other studies within the early response literature might be the fact that most previous research only analyzed treatment completer samples. Early deteriorating patients, however, might constitute a subgroup with increased risk for treatment dropout. As a consequence, a completer analysis could miss this clinically relevant subgroup. Highlighting its clinical importance, the subgroup of early deteriorating patients clearly showed the worst treatment outcomes. None of the 15 patients classified as early deteriorators reached reliable improvement at the end of the treatment, and their average pre-to-post effect size was also negative. This information could provide clinicians with clinically relevant information early on in treatment.

Membership in one of the four change groups was predicted by patient-rated impairment in social functioning (WSAS-SR) and separation anxiety (ASAC). Patients with higher levels of functional impairment were more likely to be categorized in one of the two symptomatically impaired groups. Of note, however, this could be partly explained by the similarity between Items 6 and 7 on the PDSS-SR and items on the WSAS. High initial levels of separation anxiety only heightened the probability of being in the *high symptoms – slow change* group. A probable explanation for the lack of predictive power of other intake variables might be the

administration of the instruments. Like the PDSS-SR, the WSAS-SR, and the ASAC are self-report measures, whereas the HAM-D and HAM-A are clinician ratings. Patients' self-report symptom assessments often load highly on a shared general factor that represents overall psychological distress.

The results further provide evidence that membership in one of the four latent trajectory classes could serve as a helpful predictor of ultimate treatment outcome and also, especially in the case of early deterioration, early treatment termination. In accordance with findings from previous studies, patients with early positive change were likely to be reliably improved at the end of the treatment. Notably, the mean effect size for early improving patients was more than twice as high as the average overall effect size. The initial PDSS-SR score and patient-rated functional impairment (WSAS-SR) remained significant predictors of overall PDSS-SR change during treatment, when early change patterns had been added to the model. Patient-rated functional impairment seemed to tap an additional specific aspect of the differences in treatment outcomes, which could not be explained by initial level of symptom severity and early change.

In line with previous findings (Lutz et al., 2009), patients in the low symptoms − slow change group showed a low probability of being reliably improved on the PDSS-SR at the end of the treatment (37.6%). This may be due to characteristics of the instrument (e.g., floor effect). Patients in this subgroup started treatment at a relatively low level of impairment on the PDSS-SR and, thus, had less room for improvement.

Participants with early symptom deterioration as well as those with *high symptoms and slow initial change* had lower treatment completion rates than other participants. By contrast, those with early PDSS-SR improvement were more likely to complete all 11 sessions of the protocol. Considering that in the literature early response is often associated with shorter treatment length (e.g., Haas et al., 2002), this finding might be a surprise. This could be a result of the structured nature of this treatment, which was designed for a relatively short period of 11 sessions. In naturalistic settings with a nonfixed number of sessions, patients might rather tend to leave treatment early if they experience early improvement. However, in settings with more structured and controlled characteristics, early positive changes might increase the compliance to stay in treatment until the regularly planned termination.

In summary, our findings could provide some guidance to clinicians. Identifying these four patterns of early change could have implications for treatment outcome and treatment completion rates. Session-by-session monitoring and feedback of this information to clinicians could increase awareness of these early change patterns (Lutz et al., 2013).

The following limitations of this study are noteworthy. First, only patients with at least three treatment sessions and a first-session PDSS-SR score were included in the present analyses. Therefore, results cannot be generalized to patients terminating treatment after one or two sessions and/or those reporting no initial symptoms of panic on this self-report measure. However, previous studies have shown substantial symptom reductions within the first two sessions (Haas et al., 2002). Consequently, some patients might have terminated treatment after one or two sessions due to very early improvements and, thus, were excluded from the present investigation.

Another limitation is the reliance on a self-report measure, the PDSS-SR, to assess treatment progress. We used the PDSS-SR because it was administered at every session. This enabled us to identify and examine individual patterns of patients' change during the first five sessions. Prior publications, which reported outcome data from the same trial, used independent evaluators to determine response rates (Aaronson et al., 2008; White et al., 2013). On the basis of the assessments of independent evaluators and the completer data set, these studies reported slightly higher response rates (50.3% instead of 48.8%) than those based on the RCI used in the present investigation. The evaluator-rated response status proved to be quite stable among responders over an ensuing 9-month period (White et al., 2013). Unfortunately, these independent evaluator assessments were only available for the completer sample and, thus, not applicable to the present study.

Another shortcoming of the present study relates to the interpretation of the early positive response pattern. It cannot be ruled out that early response was also partially due to factors such as regression to the mean or placebo effects (e.g., Stewart et al., 1998). To eliminate these alternative explanations, it would be necessary to investigate additional change patterns among an untreated group of patients with panic disorder, and compare the patients within early positive change classes. We conducted additional analyses of the percentages of reliable change using a concept introduced by Speer (1992) to control for regression to the mean, and compared those percentages with the percentages in Table 4. We did not find any differences regarding the percentages of reliable change, or any change in the four classes, which indicates that regression to the mean is unlikely to explain the phenomenon of early response alone. Furthermore, the present study was somewhat limited with regard to the number of intake variables that could be used as predictors of early change profiles and treatment outcome.

Concerning the application of GMM, the choice of the optimal number of groups and the potential of reifying groups have been topics of much discussion and debate (Bauer, 2007, 2011; Nagin & Odgers, 2010; Nagin & Tremblay, 2005). Simulation studies have shown that modest specification errors could result in an overextraction of trajectory classes through GMM (Bauer, 2007). Taking this tendency to extract too many classes into account, the most conservative model selection criteria were used in this study and revealed a parsimonious description of the individual change trajectories (Nylund et al., 2007). Due to model stability issues, the restrictive criteria of intercept variances being equal across classes and slope variances being fixed to zero were applied. Problems with model stability can arise because of limited or small sample sizes. In particular, the assumption of equal intercept variances might be relaxed, and the results might be tested in comparison to our results in future research with larger samples.

Furthermore, the identified subgroups should not be interpreted as literally distinct entities, but rather as approximations of a more complex reality (Bauer, 2011; Nagin & Odgers, 2010). Thus, according to the recommendations of several methodologists, the clusters explored in the present study were interpreted as groups of individuals following similar change trajectories (e.g., Bauer, 2011; Nagin & Odgers, 2010; Nagin & Tremblay, 2005). It should also be noted that class categorization is not a deterministic but a probabilistic procedure. Generally, each individual has a certain probability for belonging to each one of the classes that has been

identified via GMM. However, for the sake of reduced complexity, we treated these individuals as being categorized to the class for which they demonstrated the highest class membership probability. GMM was used as an exploratory tool that reduces the complexity of individual change trajectories to provide clinicians with a vivid representation of typical early response patterns. However, valid representations of patients' change patterns can only emerge through the accumulation of findings over several studies.

Despite these limitations and cautionary remarks, the present study illustrates the encouraging potential of investigating early change patterns. More studies focusing specifically on early treatment profiles for different homogenous and heterogeneous diagnostic groups and treatment settings are needed in order to develop decision rules that reliably support clinicians' decision-making processes and provide options to offer more cost-effective and tailored treatments. With knowledge of the actual early treatment course of a specific patient, these kinds of empirically derived decision rules could inform therapists of what to expect in later stages of the treatment with this patient. Ultimately, this information could be used to develop new treatment protocols and feedback tools for special subgroups of patients.

References

Aaronson, C. J., Shear, M. K., Goetz, R. R., Allen, L. B., Barlow, D. H., White, K. S., . . . Gorman, J. M. (2008). Predictors and time course of response among panic disorder patients treated with cognitive-behavioral therapy. *The Journal of Clinical Psychiatry, 69,* 418–424. doi:10.4088/JCP.v69n0312

Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80,* 93–101. doi:10.1037/a0026455

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

Barlow, D. H., & Craske, M. G. (2007). *Mastery of your anxiety and panic*. Oxford, England: Oxford University Press.

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42,* 757–786. doi:10.1080/00273170701710338

Bauer, D. J. (2011). Evaluating individual differences in psychological processes. *Current Directions in Psychological Science, 20,* 115–118. doi:10.1177/0963721411402670

Boswell, J. F., Sauer-Zavala, S. E., Gallagher, M. W., Delgado, N. K., & Barlow, D. H. (2012). Readiness to change as a moderator of outcome in transdiagnostic treatment. *Psychotherapy Research, 22,* 570–578. doi:10.1080/10503307.2012.688884

Colder, C. R., Campbell, R. T., Ruel, E., Richardson, J. L., & Flay, B. R. (2002). A finite mixture model of growth trajectories of adolescent alcohol use: Predictors and consequences. *Journal of Consulting and Clinical Psychology, 70,* 976–985. doi:10.1037/0022-006X.70.4.976

Cox, D. R., & Snell, E. J. (1989). *The analysis of binary data* (2nd ed.). London, United Kingdom: Chapman & Hall.

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. doi:10.1016/j.jad.2005.09.001

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2010). *An introduction to latent variable growth curve modeling (2nd ed.)*. New York, NY: Psychology Press.

Gibbons, R. D., Hedeker, D., Elkin, I., Waterneaux, C., Kraemer, H., Greenhouse, J., . . . Watkins, J. T. (1993). Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Archives of General Psychiatry, 50,* 739–750. doi:10.1001/archpsyc.1993.01820210073009

Grilo, C. M., Money, R., Barlow, D. H., Goddard, A. W., Gorman, J. M., Hofmann, S. G., . . . Woods, S. W. (1998). Pretreatment patient factors predicting attrition from a multicenter randomized controlled treatment study for panic disorder. *Comprehensive Psychiatry, 39,* 323–332. doi:10.1016/S0010-440X(98)90043-8

Gunlicks-Stoessel, M., & Mufson, L. (2011). Early patterns of symptom change signal remission with interpersonal psychotherapy for depressed adolescents. *Depression and Anxiety, 28,* 525–531. doi:10.1002/da.20849

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. doi:10.1002/jclp.10044

Hafner, J., & Marks, I. M. (1976). Exposure in vivo of agoraphobics: Contributions of diazepam, group exposure and anxiety evocation. *Psychological Medicine, 6,* 71–88. doi:10.1017/S0033291700007510

Hamilton, M. (1959). The assessments of anxiety states by rating. *British Journal of Medical Psychology, 32,* 50–55. doi:10.1111/j.2044-8341.1959.tb00467.x

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23,* 56–62. doi:10.1136/jnnp.23.1.56

Hofmann, S. G., Barlow, D. H., Papp, L. A., Detweiler, M., Ray, S., Shear, M. K., . . . Gorman, J. M. (1998). Pretreatment attrition in a comparative treatment outcome study on panic disorder. *American Journal of Psychiatry, 155,* 43–47.

Hofmann, S. G., Schulz, S. M., Meuret, A. E., Moscovitch, D. A., & Suvak, M. (2006). Sudden gains during therapy of social phobia. *Journal of Consulting and Clinical Psychology, 74,* 687–697. doi:10.1037/0022-006X.74.4.687

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *The Journal of Clinical Psychiatry, 69,* 621–632. doi:10.4088/JCP.v69n0415

Houck, P. R., Spiegel, D. A., Shear, M. K., & Rucci, P. (2002). Reliability of the self-report version of the Panic Disorder Severity Scale. *Depression and Anxiety, 15,* 183–185. doi:10.1002/da.10049

Hunter, A. M., Muthén, B. O., Cook, I. A., & Leuchter, A. F. (2010). Antidepressant response trajectories and quantitative electroencephalography (QEEG) biomarkers in major depressive disorder. *Journal of Psychiatric Research, 44,* 90–98. doi:10.1016/j.jpsychires.2009.06.006

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

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. doi:10.1080/10503300601032506

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

Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.-A., Jorasz, C., . . . Tschitsatz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research, 23,* 14–24. doi:10.1080/10503307.2012.693837

Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., . . . Iveson, S. (2005). Predicting rate and shape of change for individual clients receiving psychological therapy: Using growth curve modeling and nearest neighbor technologies. *Journal of Consulting and Clinical Psychology, 73,* 904–913. doi: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 de-

pressive disorders. *Journal of Affective Disorders, 118,* 60–68. doi:10.1016/j.jad.2009.01.019

Maier, W., Buller, R., Philipp, M., & Heuser, I. (1988). The Hamilton Anxiety Scale: Reliability, validity and sensitivity to change in anxiety and depressive disorders. *Journal of Affective Disorders, 14,* 61–68. doi:10.1016/0165-0327(88)90072-9

Manicavasagar, V., Silove, D., Wagner, R., & Hadzi Pavlovic, D. (1999). Parental representation associated with adult separation anxiety and panic disorder agoraphobia. *Australian and New Zealand Journal of Psychiatry, 33,* 422–428. doi:10.1046/j.1440-1614.1999.00566.x

Morral, A. R., Iguchi, M. Y., Belding, M. A., & Lamb, R. J. (1997). Natural classes of treatment response. *Journal of Consulting and Clinical Psychology, 65,* 673–685. doi:10.1037/0022-006X.65.4.673

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry, 180,* 461–464. doi:10.1192/bjp.180.5.461

Muthén, B. (2006). The potential of growth mixture modelling. *Infant and Child Development, 15,* 623–625. doi:10.1002/icd.482

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

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78,* 691–692.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4,* 139–157. doi:10.1037/1082-989X.4.2.139

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

Nagin, D. S., & Tremblay, R. E. (2005). Developmental trajectory groups: Fact or a useful statistical fiction? *Criminology, 43,* 873–904. doi:10.1111/j.1745-9125.2005.00026.x

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14,* 535–569. doi:10.1080/10705510701575396

Otto, M. W., Tolin, D. F., Nations, K. R., Utschig, A. C., Rothbaum, B. O., Hofmann, S. G., & Smits, J. A. (2012). Five sessions and counting: Considering ultra-brief treatment for panic disorder. *Depression and Anxiety, 29,* 465–470. doi:10.1002/da.21910

Payne, L. A., White, K. S., Gallagher, M. W., Woods, S. W., Shear, M. K., Gorman, J. M., . . . Barlow, D. H. (2012). *Second-stage treatments for relative non-responders to cognitive-behavioral therapy (CBT) for panic disorder with or without agoraphobia: Continued CBT versus paroxetine*. Manuscript submitted for publication.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464. doi:10.1214/aos/1176344136

Shear, M. K., Brown, T. A., Barlow, D. H., Money, R., Sholomskas, D. E., Woods, S. W., . . . Papp, L. A. (1997). Multicenter collaborative Panic Disorder Severity Scale. *American Journal of Psychiatry, 154,* 1571–1575.

Shear, M. K., Rucci, P., Williams, J., Frank, E., Grochocinski, V., Vander Bilt, J., . . . Wang, T. (2001). Reliability and validity of the panic disorder severity scale: Replication and extension. *Journal of Psychiatric Research, 35,* 293–296. doi:10.1016/S0022-3956(01)00028-0

Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60,* 402–408. doi: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. doi:10.1001/archpsyc.55.4.334

Stulz, N., Gallop, R., Lutz, W., Wrenn, G. L., & Crits-Christoph, P. (2010). Examining differential effects of psychosocial treatments for cocaine dependence: An application of latent trajectory analyses. *Drug and Alcohol Dependence, 106,* 164–172. doi:10.1016/j.drugalcdep.2009.08.009

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. doi:10.1037/0022-006X.75.6.864

Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in a growth mixture model. In G. R. Hancock (Ed.), *Mixture models in latent variable research* (pp. 317–341). Greenwich, CT: Information Age.

Uher, R., Muthén, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., . . . McGuffin, P. (2010). Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine, 40,* 1367–1377. doi:10.1017/S0033291709991528

van Calker, D., Zobel, I., Dykierek, P., Deimel, C. M., Kech, S., Lieb, K., . . . Schramm, E. (2009). Time course of response to antidepressants: Predictive value of early improvement and effect of additional psychotherapy. *Journal of Affective Disorders, 114,* 243–253. doi:10.1016/j.jad.2008.07.023

White, K. S., Allen, L. B., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2010). Attrition in a multicenter clinical trial for panic disorder. *The Journal of Nervous and Mental Disease, 198,* 665–671. doi:10.1097/NMD.0b013e3181ef3627

White, K. S., Payne, L. A., Gorman, J. M., Shear, M. K., Woods, S. W., Saska, J. R., & Barlow, D. H. (2013). Does maintenance CBT contribute to long-term response of panic disorder? A randomized controlled clinical trial. *Journal of Consulting and Clinical Psychology, 81,* 47–57. doi:10.1037/a0030666

Received December 21, 2012
Revision received October 22, 2013
Accepted November 22, 2013 ■

Article 2

Rubel J., Lutz W. and Schulte D. (2015). Patterns of change in different phases of outpatient psychotherapy: A stage-sequential pattern analysis of change in session reports. *Clinical Psychology & Psychotherapy, 22*(1), 1–14.

Clinical Psychology and Psychotherapy
Clin. Psychol. Psychother. **22**, 1–14 (2015)
Published online 2 October 2013 in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/cpp.1868

Patterns of Change in Different Phases of Outpatient Psychotherapy: A Stage-Sequential Pattern Analysis of Change in Session Reports

Julian Rubel,[1]* Wolfgang Lutz[1] and Dietmar Schulte[2]
[1] *Department of Clinical Psychology & Psychotherapy, University of Trier, Trier, Germany*
[2] *Department of Psychology, Ruhr-University, Bochum, Germany*

Knowledge about typical change patterns of psychotherapy patients can help to improve treatment guidelines for psychological disorders. Recent studies showed that it is possible to identify several patient subgroups with regard to their early change pattern. However, although focusing on the early phase of treatment, change patterns in later stages have hardly been investigated yet. In this study, Growth Mixture Modelling was used to identify latent change classes in different phases of therapy in a naturalistic sample of 1229 psychotherapy outpatients. Furthermore, this paper inquired into the relation between the change patterns in different phases as well as their predictive power for therapy length and outcome. Results revealed different change patterns for the three investigated phases. While in an early treatment phase, (sessions one–six) five different change patterns could be identified: the number of change classes decreased considerably over time, resulting in three patterns in the second (sessions 7–12) and two in the third phase (sessions 13–18). In each phase, by far, the biggest class showed a pattern of good progress with small/no further improvements. Copyright © 2013 John Wiley & Sons, Ltd.

Key Practitioner Messages:

- Most change in patients' progress estimates took place in an early phase of the treatment and levelled out on a relatively high level in later phases of the treatment.
- Substantial improvements were still present in later phases of the treatment but occurred less frequent than in early stages.
- Continuous outcome monitoring and feedback systems should integrate progress measures to monitor patients progress especially in the early phase of the treatment and feed the so gained information back to therapists.

Keywords: Patient-focused Research, Differential Change Patterns, Growth Mixture Modelling, Session Reports, Transition Analysis

INTRODUCTION

Recent developments in psychotherapy research emphasized the importance of an in depth investigation of patients' individual change trajectories over the course of treatment (Krause, Howard, & Lutz, 1998). The application of an empirically based treatment alone, which on average has shown to be efficacious and effective in randomized controlled trials and naturalistic studies, does not guarantee that the respective treatment is the right one for a specific patient (Howard, Moras, Brill, Martinovich, & Lutz, 1996). As has been shown repeatedly, the monitoring of patients' progress and the feedback of this information to therapists is a promising tool, which helps therapists to identify patients at risk for treatment failure early on and thus reduces the number of patients with negative outcome and helps to allocate more resources (in terms of sessions) to the patients in need (Lambert, 2007; Shimokawa, Lambert, & Smart, 2010). Several large scale studies in the context of the new emerging field of patient-focused psychotherapy research produced large datasets of patients' change courses in naturalistic settings (Castonguay, Barkham, Lutz, & McAleavey, 2013). Using different statistical methods, researchers analyzed these datasets to derive empirical decision rules that help therapists to guide their clinical decision making. One empirical approach, that arised recently in clinical research and already stimulated some research, is group based trajectory modelling such as growth mixture modelling (GMM; e.g., Muthén, 2006). GMM is a kind of advanced cluster analytic method that categorizes individuals into subgroups with similar treatment response trajectories in a change variable over a

*Correspondence to: Julian Rubel, Department of Clinical Psychology & Psychotherapy, University of Trier, Trier, Germany.
E-mail: rubel@uni-trier.de

Clinical Psychology & Psychotherapy

defined period. The resulting subgroups represent typical change patterns or profiles that could be expected as response to psychotherapy. Although much research was concerned with the differentiation of psychopathological entities and specific implications for these disorder subgroups, less is known about typical change patterns within therapy (Lutz, Stulz, Martinovich, Leon, & Saunders, 2009). However, knowledge on typical response profiles would offer several advantages for researchers and therapists (Morral, Iguchi, Belding, & Lamb, 1997): (a) subgroups of shared treatment response patterns might reveal differential treatment effects for certain patient groups (Cuijpers, van Lier, van Straten, & Donker, 2005; Lutz, Stulz, & Köck, 2009; Stulz, Gallop, Lutz, Wrenn, & Crits-Christoph, 2010); (b) early treatment response patterns could be an additional predictor of treatment outcome (Lutz, Stulz, & Köck, 2009; Stulz, Lutz, Leach, Lucock, & Barkham, 2007); (c) therapist could use typical response trajectories as a benchmark with which they can compare the progress of their actual client (Lutz, Stulz, Martinovich, et al., 2009); and (d) given a certain configuration of intake variables, patients most likely change pattern for a specific treatment could be identified, and thus, the most promising treatment could be determined before the treatment actually started (Lutz et al., 2006). Only recently in clinical research, pattern analyses were used to identify clusters of individuals following similar change trajectories on an outcome variable over a period (Nagin & Odgers, 2010). These studies investigated patterns of change in heterogeneous, naturalistic samples (e.g., Stulz et al., 2007) as well as in different disorder specific subgroups (e.g., Cuijpers et al., 2005). Whereas some focused specifically on change patterns in an early phase of the treatment (Lutz, Stulz, & Köck, 2009), others clustered change trajectories over the whole therapy (Stulz et al., 2010).

However, the reviewed pattern analyses of patients' change courses are restricted with regard to the applied outcome variables and the investigated period: these investigations were markedly limited to symptom measures as outcome variable and to a single phase approach to psychotherapy (whether an early phase or the complete treatment). Accordingly, the current paper expands on the existing literature with regard to these shortcomings: (1) for the first time the classification of patterns of change is based on a post-session progress measure and (2) multiple phases of the treatment are separately analyzed using stage-sequential GMM.

Patterns of change in patients' session reports have been recently examined by Lutz et al. (2012). They applied a different method of pattern analysis, namely the concept of sudden gains and sudden losses (Stiles et al., 2003; Tang & DeRubeis, 1999). The authors found that about 40% of positive clear-cut changes between two subsequent sessions in patients' post session reports took place in an early phase of the treatment, thus emphasizing the importance of this early stage. Nevertheless, this also implies that substantial changes are present in later stages as well. As a consequence, a mere concentration on early treatment stages alone would miss these later changes in patients' subjectively perceived progress (Lutz et al., 2012). Similarly, conceptualizing the complete course of psychotherapy as one meaningful period could mask important aspects of the nature of the change process. It seems to be more appropriate to see the complex course of psychotherapy as a period that is composed of several different phases with different focus areas and possibly different change mechanisms (Kazdin, 2007). In the present paper, a method for the analysis of psychotherapy progress data is demonstrated that allows for a multiphase and thus non-linear approach to change in psychotherapy.

Different models have been proposed to analyze multiphase longitudinal data (Kim & Kim, 2012). In the current investigation, stage-sequential GMM was chosen. These models allow classifications that are solely based on change trajectories within the respective phases—thus, the classification within a certain phase is not influenced by group membership within one of the other two phases.

Considering the above described research on patterns of change in psychotherapy, this study focuses on the following research questions: first, which patterns of change in patients' session reports can be identified in an early phase of psychotherapy (sessions one–six) and how are they related to therapy outcome and treatment length? Second, which patterns could be identified in two additional later treatment phases (sessions 7–12 and sessions 13–18) and what are their specific properties? Third, do the identified change groups differ with regard to patients' intake characteristics and/or initial symptomatic and interpersonal impairment? Fourth, how are the latent change groups of the different phases connected; i.e., are there more and/or less probable transitions from one treatment phase to the next?

With respect to the first question, we expected to find similar early change patterns to those reported in the literature for other samples and instruments. Especially a group of patients showing large positive early change in session reports at an early stage was expected and anticipated to show better final outcome while having relatively short treatments. This early change group was not studied with session reports so far and was the main focus of the first research question.

Furthermore, given previous research on change patterns and dose-effectiveness research (e.g., Lambert, Hansen, & Finch, 2001; Lutz et al., 2012), we expected more variance in changes in the session reports and therefore more patterns of change in the early phase of treatment as well as overall more positive change than in later treatment phases. However, taking the lack of prior research on patterns of change in later treatment phases and their relation to early phases and outcome into account, this research questions are more explorative in nature. Finally,

change patterns were expected to be predictive for treatment outcome and length of treatment as well as predictable through patients' intake characteristics (symptoms and interpersonal problems).

METHODS

Patients and Procedure

The sample comprised 1229 patients receiving cognitive–behavioural therapy (CBT) treatments in German outpatient psychotherapy clinics. Patients had an average age of 36.3 years (standard deviation [SD] = 11.1, range 15–74), and 56% were female. All patients who participated in the assessment system and attended at least five therapy sessions were included. The mean therapy duration was 32.0 sessions (SD = 17.3, range: 5–150). Diagnoses were based on the *Structured Clinical Interview for Axis I Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) Disorders–Patient Edition* (First, Spitzer, Gibbon & Williams, 1995). The diagnostic interview was conducted by another clinician than the one who provided the actual treatment. For 48 patients (3.9%), no axis I or II diagnoses was given, whereas 443 (36.0%) had a single diagnosis, 133 (10.8%) had two, 544 (44.3%) had three and 61 (4.9%) had four or more diagnoses. The distribution of patient diagnosis was anxiety disorders (55.9%; $n = 687$), affective disorders[1] (18.1%; $n = 223$), other disorders (10.0%; $n = 123$), comorbid anxiety and affective disorders (7%; $n = 86$) or other comorbid disorders (5.0%; $n = 62$). 3.3% patients ($n = 40$) out of the 1229 had at least one additional axis II personality disorder.[2] Patients were treated by 129 therapists in clinical training on different stages during their 3 to 5 year clinical training programme. About one percent of patients were treated by therapists who just finished their licence between 1 and 3 months ago. Therapists had an average age of 32.8 years (SD = 7.0, range 25–56) and 67% were female.

Instruments and Data Collection

Session Report (SR)

Patients assessed their perceived progress by completing the SR (Flückiger, Regli, Zwahlen, Hostettler, & Caspar, 2010; Schulte & Eifert, 2002) at the end of each session. A six-item subscale[3] ranging from 1 ('*not at al*') to 7 ('*yes, exactly*'),[4] was created out of the complete 12-item version, resulting in a valid measure of patients' progress (Flückiger, Grosse Holtforth, Znoj, Caspar, & Wampold, 2013; Lutz et al., 2012). Average scores over the first five sessions as well as positive and negative fluctuations in SR early in treatment showed to be significant predictors of pre-post symptom change (Flückiger et al., 2013; Lutz et al., 2012). The internal consistency of the six-item progress scale was $\alpha = 0.86$ (Lutz et al., 2012). On average patients rated their progress on the seven-point scale with $M = 4.80$ (SD = 1.01) indicating generally high perceived progress.

Brief Symptom Inventory (BSI)

To asses overall change in symptoms, the *BSI* (Franke, 2000; German translation of Derogatis, 1975) was administered pre-treatment and post-treatment. The *BSI* is a 53-item short form of the Derogatis' Symptom Checklist that assesses the general symptomatic status of patients in nine subscales on a five-point Likert scale. In this study, only the overall general symptom index was used. The internal consistency of the BSI is $\alpha = 0.92$ and the retest-reliability $r_{tt} = 0.90$ (Franke, 2000).

Inventory of Interpersonal Problems (IIP-D)

Changes in interpersonal problems from pre-therapy to post-therapy were assessed with the *IIP-D* (Horowitz, Strauss & Kordy, 2000). For the present project, the 64-items form of the IIP-D was used, which measures interpersonal problems regarding behaviour, thoughts and emotions. In this study, only the overall score was used, which has an internal consistency of $\alpha = 0.94$ and a retest-reliability (10 weeks) of $r_{tt} = 0.98$ (Horowitz et al., 2000).

Data Analysis Strategy

In a first step, the first 18 sessions of treatment were divided in three phases with six sessions each (sessions 1–6, 7–12 and 13–18). The size of the phases was determined in accordance to prior research on early change patterns (e.g., Stulz et al., 2007). For each of these phases, a separate growth model was fitted to determine the specific average rate and shape of change in patients' session reports within these treatment periods for available patients in the phases

[1]The following DSM-IV diagnoses were subsumed in the anxiety disorders cluster: 300.2X; 300.0X; the following DSM-IV diagnoses were subsumed in the affective disorders cluster: 296.X; 296.XX; 300.4; 301.13.

[2]The represented personality disorders (PD) in the sample were paranoid PD, borderline PD, histrionic PD, narcissistic PD, avoidant PD, dependent PD, obsessive compulsive PD and not otherwise specified PD.

[3]This subscale comprises of six items: (1) 'I have the feeling that I better understand myself and my problems'; (2) 'Today we got closer to the core of my problems.'; (3) 'Today I became aware why I react towards some people in a certain way and not in a different way.'; (4) 'I am now better able to cope with problematic situations, were I was not able to cope with before treatment.'; (5) 'Now I am more confident to solve my problems by myself.'; and (6) 'Now I know better what I want.'

[4]The other anchors were 2 'no', 3 'rather not', 4 'neither nor', 5 'rather yes' and 6 'yes'.

(some patients terminated after each phase). The best fitting model in each phase was determined using the Bayesian Information Criterion (BIC, Schwartz, 1978).

In a next step, the fitted growth models for each phase were expanded with GMM. GMM is a type of cluster analysis over time within a latent variable framework (Muthén, 2006). This method reveals subgroups of patients with similar change trajectories within the respective treatment phases by adding a categorical latent variable to a conventional growth model. These conventional growth models rely on the assumption of a continuously distributed patient population and thus capture individual differences in intake scores (intercepts) and change parameters (slopes) by the means of random coefficients (i.e., variation around an average intercept and slope). In GMM, a categorical variable is implemented that additionally accounts for sample heterogeneity (Duncan, Duncan, & Strycker, 2006). Compared with conventional cluster analytic methods, GMM also takes the uncertainty of class assignment into account. For each individual, the probability of belonging to each of the groups is estimated, instead of clustering each individual deterministically into one of the classes. For all analyses, individuals were assigned to the group with the largest posterior probability estimate. This is the group that best fits to the individuals' change in session reports in the respective phase.

Due to stability of the models, variances around the class specific slopes were fixed at zero, whereas intercept variances were freely estimated but constrained to be constant between classes. Given that, heterogeneity in change has to be captured completely by the differences in mean slopes of different latent classes. These models can be seen as a hybrid between models where the variances of all parameters are fixed to zero, so called Latent Class Growth Models (Nagin, 1999) and models that allow all parameters to be freely estimated.

The optimal number of latent classes was determined by the use of the *BIC* (Schwartz, 1978) and the *Bootstrapped Likelihood Ratio Test* (*BLRT*). These tests have been recommended in simulation studies (e.g., Nylund, Asparouhov, & Muthen, 2008). Thus, the model determination process for each of the three phases was two-fold: First, the model with the lowest value in the BIC was identified by starting to estimate a one-class model and successively adding one more class. In a second step, the k-class solution was tested against a solution with k-1 classes using the BLRT. A significant *p*-value ($p < 0.05$) indicates superiority of the k-class over the k-1 class solution. If the BLRT was not significant, the k-class solution had been rejected, and the k-1 class model was tested against a model with two classes less (k-2). This procedure was repeated until the BLRT had a significant *p*-value. All growth models were estimated using the Mplus software package (Version 6.11, Muthén & Muthén, 2010).

In a further step, connections of the three phases were investigated by calculating transition probabilities. For each group of the first treatment phase, the average probability of its members to belong to one of the groups of the second phase or to terminate treatment prematurely is calculated. Equivalently, membership in one of the second phase groups is connected to membership in the third phase groups. Accumulations of transition patterns have been analyzed using Chi-square tests.

To identify predictors of latent class membership, multinomial logistic regression analyses were conducted. As potential predictors of the latent change classes, only the intake scores in the three outcome measures (BSI and IIP) as well as age and sex of the patients were available.

In a last step, analyses of variance were used to compare therapy outcomes (standardized pre-to-post differences in the three outcome measures) and durations[5] between the latent trajectory classes in each treatment phase.

RESULTS

As best description of change in the first two treatment phases (sessions 1–6 and sessions 7–12), growth models indicated a negatively accelerating (log-linear) trend over time, whereas the third phase (sessions 13–18) was best described by a linear shape.[6]

Overall, the stage-sequential investigation of change patterns revealed a decreasing number of latent trajectory classes over time (Figure 1). Five different change groups (C1.1–C1.5) were identified in the first treatment phase (sessions one–six), there were three (C2.1–C2.3) in the second phase (sessions 7–12) and only two (C3.1–C3.2) in the third phase of the treatment (session 13–18). In all three phases, one big group could be differentiated from other relatively small groups (Figure 1). The proportion of patients belonging to the biggest group of the respective phases increased considerably from the first to the second phase (C1.3: 81.5% versus C2.2: 96.2%) and remained relatively constant from the second to the third phase (C3.2: 96.3%).

Phase 1: Patterns of Change in Sessions 1–6

By using the described model determination procedure, a five class solution revealed the best model fit for the GMM in the first phase of the treatment. The average change trajectories of the five groups are depicted in the phase 1

[5]In all analysis, a logarithmic transformation of therapy length to the base *e* has been used. This is a common practice to normalize skewed variables such as the therapy duration (Tabachnick & Fidell, 2009).

[6]BIC values for phase 1 model: intercept-only: 17 821, linear: 17 011, log-linear: 16 867, quadratic: 16 889BIC values for phase 2 model: intercept-only: 15 362, linear: 15 342, log-linear: 15 337, quadratic: 15 350 BIC values for phase 3 model: intercept-only: 13 924, linear: 13 877, log-linear: 13 892, quadratic: 13 883

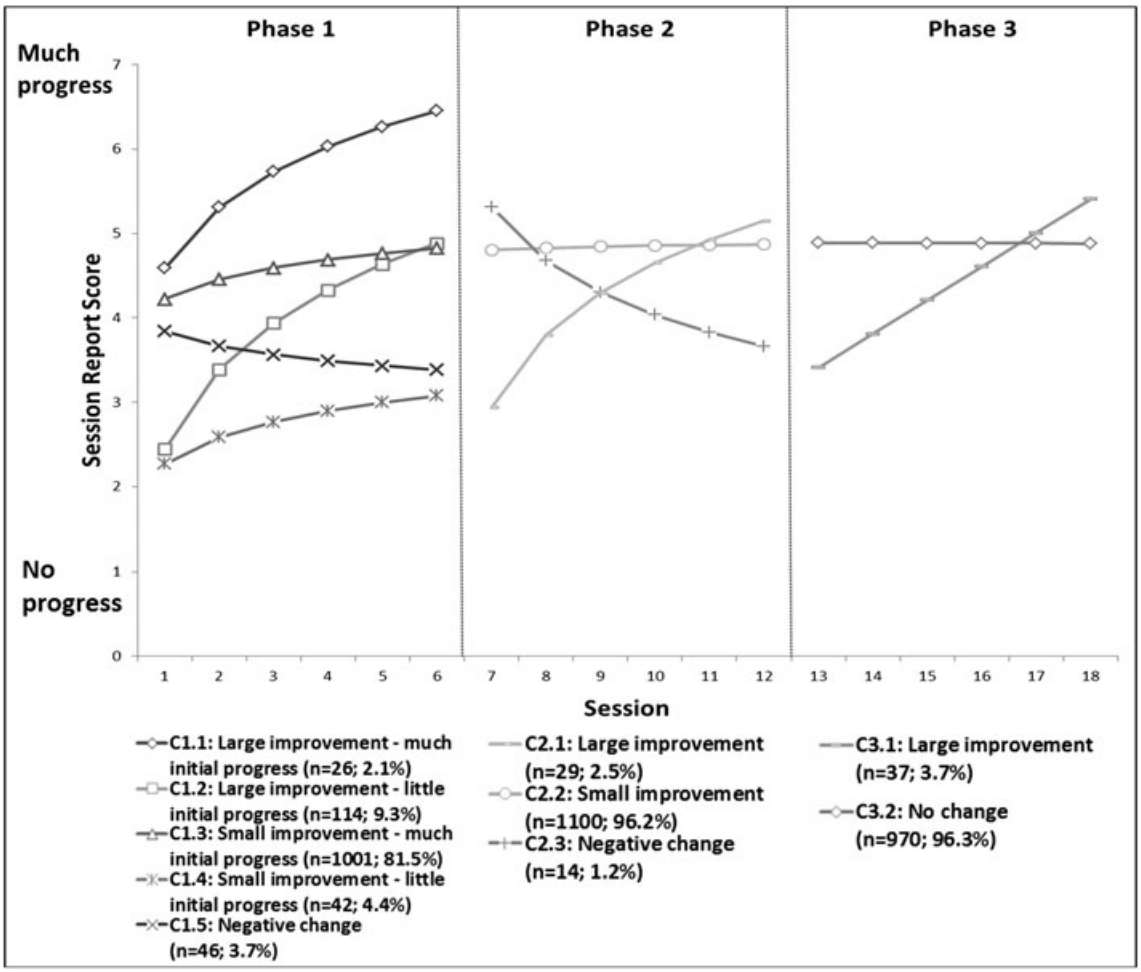

Figure 1. Typical patterns of change (latent growth trajectories revealed through Growth Mixture Modeling) for patients' session reports in three phases of psychotherapy (sessions 1–6; 7–12; 13–18)

part of Figure 1. The first class (C1.1) was labelled 'large improvement–much initial progress' and comprised 26 (2.1%) patients: While patients within this class showed the comparatively highest amount of progress after the first session ($SR_{1C1.1} = 4.6$), their average scores considerably increased further over the next five sessions ($slope_{C1.1} = 2.4$; $p < 0.05$). Patients in the second class (C1.2) also showed large average improvement in their estimates ($slope_{C1.2} = 3.1$; $p < 0.05$) but assessed their progress after the first session ($SR_{1C1.2} = 2.4$) much less positive. Consequently, this group was labelled 'large improvement–little initial progress'. Class three (C1.3), by far the biggest class, consisted of 1001 (81.5%) patients: these assessed the first session on a relatively high level ($SR_{1C1.3} = 4.2$) but showed only small, yet significant, improvement within the first treatment phase ($slope_{C1.3} = 0.8$; $p < 0.05$). Given these characteristics, the third class was named 'small improvement–much initial progress'.

Patients within class four (C1.4) showed the lowest average progress estimate after the first treatment session of all groups ($SR_{1C1.4} = 2.3$). Also the average change trajectory of these 42 (4.4%) patients showed a relatively small increase in the early treatment phase ($slope_{C1.4} = 1.0$; $p < 0.05$) and is thus labelled 'small improvement–little initial progress'.

The fifth class (C1.5) showed similar assessments after the first treatment session ($SR_{1C1.5} = 3.8$) as classes one and three but a completely different trend over the first six sessions. Although classes one and three showed a positive trend in this phase, the average perceived progress of patients within class five ($n = 46$; 3.7%) decreased significantly ($slope_{C1.5} = -0.6$; $p < 0.05$). Accordingly, this class was named 'negative change'.

Of the tested pre-treatment characteristics, only the interpersonal problems (IIP-D; $F_{(4;\ 1224)} = 3.4$; $p < 0.05$) of the patients showed a significant relation to class membership. Tukey *post hoc* tests revealed that patients of the 'large

improvement–much initial progress' class (C1.1) were significantly less disturbed in terms of interpersonal problems than patients of the 'small improvement–little initial progress' group (C1.4). Thus, with higher levels of pre-treatment interpersonal problems, the probability to belong to class four (C1.4) rather than class one (C1.1) was significantly increased.

Membership in one of the five classes was also significantly related to treatment outcome and duration. As depicted in Table 1, analyses of covariance for both outcome measures showed a significant effect of class membership on symptomatic and interpersonal impairment at the end of therapy when pre-treatment differences were statistically controlled (BSI: $F_{(4;\ 1223)} = 5.4$, $p < 0.05$; IIP-D: $F_{(4;\ 1223)} = 5.3$, $p < 0.05$). Patients in classes five ('negative change') and four ('small improvement–little initial progress') showed significantly higher symptoms (BSI) at the end of therapy than the other three groups. With regard to interpersonal problems (IIP), patients in class one ('large improvement–much initial progress) showed significantly lower average levels of impairment than patients in the other classes. Additionally, patients in class three ('small improvement–much initial progress') showed significantly lower average IIP scores than patients in class four ('small improvement–little initial progress').

With regard to treatment duration, analysis of variance ($F_{(4;\ 1224)} = 2.67$, $p < 0.05$) and Tukey *post hoc* tests revealed that patients of the 'large improvement–much initial progress' group had significantly shorter treatments than patients of the 'small improvement–much initial progress' and 'negative change' groups (Table 1).

Conclusively, patients that perceived high progress from the very first session on and whose scores remain on that high level or even increased further over the subsequent five sessions showed the best outcome while at the same time receiving relatively short treatments.

Phase 2: Patterns of Change in Sessions 7–12

Model fit indices suggested a three class model as best solution for the description of the second treatment phase. The average change trajectories of the three groups are depicted in the phase 2 part of Figure 1. The first patient subgroup within this second treatment phase (C2.1; $n = 29$, 2.5%) showed clearly the lowest average assessment of progress after session seven ($SR_{7C2.1} = 2.9$). In the following sessions their perceived progress increased considerably ($slope_{C2.1} = 2.8$; $p < 0.05$). Thus, this class was labelled 'large improvement'.

Similarly to the first phase, the biggest patient group (C2.2; $n = 1100$, 96.2%) started this second phase with relatively high assessments of perceived progress ($SR_{7C2.2} = 4.8$) and showed small but significant improvement in their average score over the next five sessions ($slope_{C2.2} = 0.1$; $p < 0.05$). As there was no class showing comparatively low average change within this treatment phase, this group was labelled 'small improvement'.

The third patient subgroup (C2.3; $n = 14$, 1.2%) was characterized by the highest mean estimate for the seventh session ($SR_{7C2.3} = 5.3$). However, patients in this class showed fast average decrease in their scores from sessions 7–12 ($slope_{C2.3} = -2.1$; $p < 0.05$). Accordingly, this class was labelled 'negative change'.

None of the tested pre-treatment characteristics showed a significant relation to clients' membership in one of the three phase 2 groups. Thus, predictions about patients' patterns of change could not be derived for this treatment phase.

However, the pattern of change in the second phase was predictive for symptomatic impairment (BSI_{post}) at the end of therapy (Table 1) when differences in pre-treatment BSI_{pre} scores were controlled. Patients who showed 'negative change' in this second phase were significantly more averagely impaired at the end of treatment (BSI_{post}) than both other groups. This relation did not hold for patients' interpersonal problems (IIP_{post}) at the end of therapy or the average length of treatment.

Phase 3: Patterns of Change in Sessions 13–18

For the last investigated phase, a two class model was the best fitting class solution. The average change trajectories of the two groups are depicted in the phase 3 part of Figure 1. As in the previous periods, patients within one of these two classes assessed the first session of this phase (session 13) on a relatively low average level (C3.1; $SR_{12C3.1} = 3.4$) and showed significantly increasing progress over the following five sessions ($slope_{C3.1} = 0.4$; $p < 0.05$). Consistently, this relatively small class ($n = 37$; 3.7%) was labelled 'large improvement. The vast majority of patients ($n = 970$; 96.3%), who were still in treatment in this third phase, assessed their progress on a constantly high level (C3.2; $SR_{12C3.2} = 4.9$) without significant change over time ($slope_{C3.2} = 0.0$; $p > 0.05$). Accordingly, this class was named 'no change'.

Of the tested patients' intake characteristics, only the amount of interpersonal problems (IIP-D) was significantly related to class membership in this third phase. The 37 patients who report improved progress over time (C3.1) showed significantly lower IIP-D scores than the big 'no change' group (C3.2; $F_{(1;\ 1005)} = 4.7$, $p < 0.05$;). They did not differ with regard to any of the other pre-treatment variables. Thus, patients that experienced substantial improvement in their progress within this third phase began treatment significantly less impaired in terms of interpersonal problems.

Regarding the overall length of the treatment the two groups showed no significant differences.

Table 1. Symptom status (brief symptom inventory and inventory of interpersonal problems) of patients in the latent change groups at intake and termination, F values of analysis of variance on pre-treatment differences and analysis of covariance on post scores controlled for pre-treatment differences, pre-post effect sizes with 95% confidence interval

		Change pattern	M_{pre}	SD_{pre}	F_{Pre}	M_{post}	SD_{post}	F_{Ancova}	$d_{pre\text{-}post}$ [95% CI]	$M_{duration}$
Phase 1	BSI	Large improvement–much initial progress	1.19	0.68	0.84	0.45	0.58	5.37*	1.17 [0.58 1.76]	19.93
		Large improvement–little initial progress	1.25	0.70		0.66	0.65		0.87 [0.60 1.14]	26.89
		Small improvement–much initial progress	1.18	0.71		0.67	0.64		0.75 [0.66 0.84]	27.90
		Small improvement–little initial progress	1.30	0.70		1.00	0.86		0.38 [−0.05 0.81]	28.92
		Negative change	1.33	0.74		0.95	0.82		0.49 [0.07 0.90]	29.49
		Total	1.20	0.71		0.69	0.66		0.74 [0.66 0.82]	27.70
	IIP-D	Large improvement–much initial progress	1.34	0.62	3.41*	0.61	0.59	5.33*	1.21 [0.61 1.80]	19.93
		Large improvement–little initial progress	1.64	0.58		1.21	0.62		0.72 [0.45 0.99]	26.89
		Small improvement–much initial progress	1.61	0.55		1.16	0.66		0.74 [0.65 0.83]	27.90
		Small improvement–little initial progress	1.82	0.60		1.47	0.69		0.54 [0.10 0.97]	28.92
		Negative change	1.71	0.58		1.40	0.69		0.49 [0.07 0.90]	29.49
		Total	1.62	0.56		1.17	0.66		0.74 [0.66 0.82]	27.70
Phase 2	BSI	Large improvement	1.34	0.71	1.30	0.62	0.69	4.16*	1.03 [0.48 1.57]	29.89
		Small improvement	1.21	0.70		0.69	0.65		0.80 [0.71 0.89]	29.87
		Negative change	1.47	0.87		1.20	1.03		0.28 [−0.47 1.02]	30.09
		Total	1.22	0.71		0.69	0.66		0.77 [0.68 0.85]	29.89
	IIP-D	Small improvement	1.63	0.55	0.31	1.19	0.66	0.15	0.72 [0.63 0.81]	29.87
		Large improvement	1.56	0.55		1.10	0.70		0.66 [0.13 1.19]	29.89
		Negative change	1.58	0.58		1.21	0.56		0.65 [−0.12 1.41]	30.09
		Total	1.63	0.55		1.19	0.66		0.61 [0.53 0.69]	29.89
	BSI	Large improvement	1.17	0.64	0.25	0.60	0.57	0.38	0.99 [0.50 1.47]	33.38
Phase 3		No change	1.22	0.69		0.69	0.65		0.79 [0.70 0.88]	32.99
		Total	1.22	0.69		0.69	0.65		0.79 [0.70 0.88]	33.01
	IIP-D	Large improvement	1.46	0.57	4.65*	1.08	0.65	0.19	0.62 [0.15 1.08]	33.38
		No change	1.66	0.54		1.20	0.66		0.76 [0.67 0.85]	32.99
		Total	1.65	0.55		1.20	0.66		0.74 [0.65 0.83]	33.01

BSI = brief symptom inventory. IIP-D = inventory of interpersonal problems.

Note. Average pre-/post effect sizes were calculated using the formula: $d_{pre-post} = \frac{M_{pre}-M_{post}}{\sqrt{\frac{(n_{pre}-1)s^2_{pre}+(n_{post}-1)s^2_{post}}{n_{pre}+n_{post}}}}$ with M_{pre} representing the mean of all scores at the beginning of therapy, M_{post} the mean of the scores after the therapy, SD_{pre} and SD_{post} their respective standard deviations and n_{pre}/n_{post} the group sizes before and after the therapy. *$p < 0.05$.

Table 2. Cross classification of patients who were categorized to one of the five phase 1 change patterns and their group membership in the second treatment phase or the reason for not being included in the second phase (less than 11 sessions or more than one missing in phase 2)

	Change pattern in phase 2				
Change pattern in phase 1	Large improvement ($n = 29$)	Small improvement ($n = 1100$)	Negative change ($n = 14$)	<11 sessions ($n = 71$)	>1 missing ($n = 15$)
Large improvement–much initial progress ($n = 26$)	0 (0%)	20 (76.9%)	2 (7.7%)	3 (11.5%)	1 (3.8%)
Large improvement–little initial progress ($n = 114$)	4 (3.5%)	99 (86.8%)	1 (0.9%)	7 (6.1%)	3 (2.6%)
Small improvement–much initial progress ($n = 1001$)	14 (1.4%)	918 (91.7%)	8 (0.8%)	52 (5.2%)	9 (0.9%)
Small improvement–little initial progress ($n = 42$)	7 (16.7%)	27 (64.3%)	2 (4.8%)	4 (9.5%)	2 (4.8%)
Negative change ($n = 46$)	4 (8.7%)	36 (78.3%)	1 (2.2%)	5 (10.9%)	0 (0%)
Total ($n = 1229$)	29 (2.4%)	1100 (89.5%)	14 (1.1%)	71 (5.8%)	15 (1.2%)

Note. Patients with more than one missing value in phase 2 (>1 missing) or less than 11 sessions of therapy (<11 sessions) were not modelled in the second treatment phase.

Patterns of Group Transitions

Table 2 depicts the cross-classifications of class membership in phases 1 and 2 of the treatment. For each phase 1 group, the probability of its members to belong to one of the three phase 2 groups to terminate treatment before session 11 or to have more than one missing value in the second phase is given. As 60% of the cells had an expected frequency lower than five, no Chi-square tests could have been computed to inferentially test the differences between observed and expected cell frequencies. Nevertheless, it is obvious that the vast majority of patients (89.5%) from each of the five phase 1 groups transited into the 'small improvement' group in phase 2 (second column in Table 2). Patients that already belonged to the big 'small improvement–much initial progress' group in the first phase also showed the highest probability to belong to this similarly big 'small improvement' group in the second phase (91.7%). Patients of the 'small improvement–little initial progress' phase 1 group showed the lowest, yet still high, probability to belong to this phase 2 class (64.3%). The highest probability for treatment termination until session 11 had patients of the 'large improvement–much initial progress' (11.5%) and 'negative change' (10.9%) groups, followed by the 'small improvement–little initial progress' (9.5%) group. Thus, those patients with on average high and further increasing progress in the first six sessions, those with decreasing levels of perceived progress and those with continuously low levels of early progress had relatively high probabilities for a treatment termination until session 11.

Cross-classifications of class membership in phases 1 and 3 of the treatment descriptively confirmed the revealed pattern (Table 3). The overwhelming majority of the patients from the first phase transited into the big 'no change' class of phase 3 (78.9%). If only those patients were taken into account who were still in treatment in

Table 3. Cross classification of patients who were categorized to one of the five phase 1 change patterns and their group membership in the third treatment phase or the reason for not being included in the second phase (less than 17 sessions or more than one missing in phase 3)

	Change pattern in phase 3			
Change pattern in phase 1	Large improvement ($n = 37$)	No change ($n = 970$)	<17 sessions ($n = 202$)	>1 missing ($n = 20$)
Large improvement–much initial progress ($n = 26$)	1 (3.8%)	16 (61.5%)	9 (34.6%)	0 (0%)
Large improvement–little initial progress ($n = 114$)	4 (3.5%)	90 (78.9%)	18 (15.8%)	2 (1.8%)
Small improvement–much initial progress ($n = 1001$)	23 (2.3%)	806 (80.5%)	158 (15.8%)	14 (1.4%)
Small improvement–little initial progress ($n = 42$)	4 (9.5%)	29 (69%)	7 (16.7%)	2 (4.8%)
Negative change ($n = 46$)	5 (10.9%)	29 (63%)	10 (21.7%)	2 (4.3%)
Total ($n = 1229$)	37 (3%)	970 (78.9%)	202 (16.4%)	20 (1.6%)

Note. Patients with more than one missing value in phase 3 (>1 missing) or less than 17 sessions of therapy (<17 sessions) were not modelled in the second treatment phase.

Table 4. Cross classification of patients who were categorized to one of the three phase 2 change patterns and their group membership in the third treatment phase or the reason for not being included in the third phase model (less than 17 sessions or more than one missing in phase 3)

	Change pattern in phase 3			
Change pattern in phase 2	Large improvement ($n = 37$)	No change ($n = 962$)	<17 sessions ($n = 125$)	>1 missing ($n = 19$)
Large improvement ($n = 29$)	3 (10.3%)	23 (79.3%)	3 (10.3%)	0 (0%)
Small improvement ($n = 1100$)	33 (3.0%)	929 (84.5%)	120 (10.9%)	18 (1.6%)
Negative change ($n = 14$)	1 (7.1%)	10 (71.4%)	2 (14.3%)	1 (7.1%)
Total ($n = 1143$)	37 (3.2%)	962 (84.2%)	125 (10.9%)	19 (1.7%)

Note. Patients with more than two missing values in phase 3 (>1 missing) or less than 17 sessions of therapy (<17 sessions) were not modelled in the third treatment phase.

the third phase ($n = 1007$), even 97% belonged to this class ('no change'). In accordance to the transitions from phases 1 to 2, patients from the 'large improvement–much initial progress' and the 'negative change' phase 1groups had the highest probability for already having ended the treatment until session 17.

Table 4 depicts the cross-classifications of class membership in phases 2 and 3 of the treatment. Again, due to the fact that 50% of the cells had expected frequencies under five, chi-square tests were not applicable to test for unexpectedly high/low cell frequencies. Descriptively, all phase 2 groups showed the highest probability for the big 'no change' group of the third phase. Independent of the phase 2 change pattern, impressive 85% of all patients transited into this 'no change' group in phase 3. Quite interestingly, however, there was a slight tendency for patients experiencing 'large improvements' in their progress in the second phase (C2.1) to also experience positive change in the third phase (10.3%). Furthermore, besides their distinctiveness, all three phase 2 groups had about the same probabilities to end treatment until session 17.

Overall, investigating the several transition patterns comprehensively over the three phases, only one pattern showed substantial patient shares: altogether, 783 patients (63.7% of all patients or 77.8% of the patients being included in the phase 3 model) belonged in each of the three phases to the biggest class, i.e., a pattern of 'small improvements–much initial progress' (phase 1)/'small improvements' (phase 2)/'no change' (phase 2).

DISCUSSIONS

The goals of this study were the identification, exploration and characterization of patterns of patients subjectively perceived treatment progress in different phases of outpatient psychotherapy. In contrast to prior studies, for the first time, a progress measure (instead of a symptom measure) was used to describe patients' change over time, and a multiphase approach (instead of a single phase approach) to psychotherapy was adopted.

In a first step of analysis, for each of the three phases, the overall shape of change in patients' session reports has been defined. For the first two phases (sessions 1–6 and 7–12), a log-linear average trend over time best described the data. In the third phase (sessions 13–18), patients scores followed a linear average trend. This overall pattern is in accordance with the dose-effect model (Howard, Kopta, Krause, & Orlinsky, 1986). The dose-effect model postulates a negatively accelerating—approximately log-linear—relationship between the amount of therapy sessions and the number of improved patients over the complete treatment course. Following that, the probability for change is assumed to be high in early phases of the treatment and much lower in later stages. These predictions of the dose-effect model fit the revealed change patterns of the current study.

In a second step of analyses, separate GMMs were fitted to the session report scores from each of the three different treatment phases. Interestingly, a decreasing number of classes over time were identified. Although five patterns best described the individual change trajectories over the first six sessions, only three patterns were found in the second and two in the third phase. Presumably, this decreasing number of change patterns is most likely due to generally decreased change in patients' session reports in later stages of the treatment. Patients' ratings of their perceived progress seemed to be most diverse in an early phase of the treatment and levelled out on a relatively high level in later phases. To test this hypothesis, we additionally investigated each patient's individual slopes in the three phases. As would be suggested by the results of this study, the absolute average slope decreased significantly from phase 1 ($slope_{phase1} = 1.01$) to phase 2 ($slope_{phase2} = 0.2$) and even further from phase 2 to phase 3 ($slope_{phase3} = 0.02$). Additionally, the variation around these individual mean trends (root mean square error, RMSE) decreased notably from phase 1 ($RMSE_{phase\ 1} = 0.74$) to phase 2 ($RMSE_{phase\ 2} = 0.41$) and slightly further from phase 2 to phase 3 ($RMSE_{phase\ 3} = 0.38$). These supplementary analyses support the suggested hypothesis that most of the change in patients' perceived progress took place in the

early phase of the treatment. These phenomena in turn could have several reasons. Besides the fact that each phase comprised different patient subsamples due to drop out, specific characteristics of how patients use such kinds of progress measures could have caused the revealed pattern. In an early phase of the treatment, patients might have used the instrument to evaluate their progress more specifically with regard to the particular session after which they completed the assessment. In later phases, many patients who were still in treatment might generally perceive therapy sessions as helpful and facilitating good progress; otherwise, they might have quit treatment earlier. As a result, their estimates rather reflected their global perception of the helpfulness of the treatment. Thus, their progress ratings remained on a constantly high level without much change in these later phases.

Predominant in each phase is a big subgroup of patients who rated their progress on a relatively high level right at the start of the respective phase and showed small (C1.3, 'small improvement–much initial progress' and C2.2, 'small improvement') or no (C3.2, 'no change') further improvement in the following five sessions. In accordance with the finding of decreased change in progress ratings in later phases of the treatment, the proportion of patients belonging to this class, increased substantively from 81.5% in phase 1 to about 96% in phases 2 and 3. Besides these subgroups of patients with slow or not changing average trajectories, in each phase an initially rather sceptical group of patients was identified, which showed considerable improvement in their progress ratings in the course of the respective phase (C1.2, 'large improvement–little initial progress'; C2.1, 'large improvement'; and C3.1, 'large improvement'). Although in phase 1 this group comprised about 9% of all patients, this share decreased to about 3% in the later phases. The predominance of this pattern in a very early phase might reflect patients' uncertainty after the first sessions about whether therapy will help them. Descriptive but consistently, those who belonged to these positive changing groups showed rather good treatment outcome especially in terms of symptom changes (BSI).

Only present in the first two phases was a small subgroup of patients who perceived rather good progress after the first session of the respective phase but rated their progress in the following sessions less good ('negative change': C1.5; C2.3). These patients descriptively stayed in treatment rather long and showed to be at risk for bad treatment outcome. After being rather optimistic at the start of the treatment, patients of these classes might be disappointed with their further progress. These kinds of developments should be carefully monitored by therapists so that they can take respective means to hinder a potential treatment failure.

Two remaining groups showed to be unique for the first phase of the treatment. The first subgroup assessed their progress after the first session much below average while their ratings improved only slowly (C1.4); the second subgroup already rated their progress after the first session on a very high level but experienced further improvements over the course of the next sessions (C1.1). Although C1.1 reached in both outcome measures the descriptively lowest impairment of all groups at the end of the treatment and needed rather short therapies, the C1.4 class still had the highest impairment scores at the end of the treatment and showed the lowest average pre-effect to post-effect sizes. Conclusively, patients that experienced little progress after the first few sessions showed to be highly at risk for bad outcome, if they did not perceive substantial improvements in the next few sessions. A high level of pre-treatment interpersonal impairment was a dominant characteristic of this patient subgroup, which might be an indication of a possible relation between personality problems and patients' change trajectories.

Due to the fact that the class solutions in each phase are characterized by one rather big and other rather small classes, the given transition probabilities should be interpreted cautiously. Especially for the small classes even very small numbers can result in relatively high proportions. These unbalanced distributions also eliminated the possibility to derive inferential conclusions through chi-square tests. Nevertheless, some descriptive observations illustrate the potential of this stage-sequential method for future research. Due to the predominance of one big class in each of the three phases, belonging to each of these big classes was the most likely combination of class memberships. Thus, more than two thirds of the patients who had at least 17 sessions of therapy showed a phase transition pattern of 'small improvement–much initial progress'–'small improvement'–'no change' (C1.3–C2.2–C3.2). Again, this finding supported the suggestion of decreased amounts of change in patients' session reports from early to later phases of the treatment.

Only the group of patients who perceived consistently little progress in the first six sessions (C1.4, 'small improvement–little initial progress') showed relatively high probabilities to belong to another than the big (C2.2, 'small improvement') class in phase 2. Notably, patients of this class unexpectedly often showed a pattern of 'large improvement' (C2.1) in the second treatment phase. This patient subgroup might be initially rather critically minded towards the treatment but convinced of its helpfulness in its further course in later phases. A closer look at these patients could be profitable for future research, e.g., through the analysis of videotaped sessions to reveal what has helped these patients. Such efforts could help to identify the most effective interventions for these critical patients, which have also a high risk for treatment failure. Relatively little is known about in session processes and techniques, which are especially helpful for patients at risk for treatment failure. The combination of information about patients' early/late patterns of

treatment progress and ratings of videotaped sessions would bring together different lines of research and allow an empirical investigation of what is helpful for patients with a negative outcome prediction. Whipple et al. (2003), for example, applied a top-down approach to this question. They developed clinical support tools on the basis of theoretical considerations. The authors came up with a decision tree, which starts with inspecting possible problems in the therapeutic alliance. If no alliance problems are detectable in the Helping Alliance Questionnaire (Luborsky et al., 1996), the following potential problem areas are consecutively checked: patients' readiness to change, patients' social support, the diagnostic formulation and medication. Depending on the results of the consecutive steps of the decision tree, the clinician is referred to a specialized manual or to additional help from other service providers. These clinical support tools have shown to be an effective means in reducing the number of treatment failures and show the practical importance of this kind of research (Lambert, 2007; Lambert & Shimokawa, 2011; Shimokawa et al., 2010; Whipple et al., 2003).

Several aspects limit the scope of this study and give suggestions for future research.

The use of GMM has been discussed with regard to the following problems: the choice of the 'right' number of trajectory classes and the potential error to reify patient subgroups (Bauer, 2011; Nagin & Odgers, 2010). As shown in several simulation studies, widely used model selection criteria for the extraction of classes in GMM tend to result in an overestimation of the number of classes (Bauer, 2007). Taking this into account, more conservative model selection criteria were applied to destine the best fitting and most parsimonious model solution (Nylund, Asparouhov, & Muthén, 2007). Concerning the problem of reification, the revealed groups are not interpreted as 'literally distinct entities' but rather as parsimonious 'approximations of a more complex reality' (Bauer, 2007; Nagin & Odgers, 2010). Following that, the revealed patient subgroups should be understood as groups of patients with similar amounts of perceived progress over a defined period.

Quite a different methodology has been applied in the investigation of patterns of working alliance development (Kivlighan & Shaughnessy, 2000; Stiles et al., 2004). Kivlighan et al. (2000) used cluster analyses and Stiles et al. (2004) replicated this approach for a slightly different time interval (4 sessions versus 8/16 sessions). Kivlighan et al. (2000) found three distinct patterns of working alliance development of which the following two could have been replicated by Stiles et al. (2004): a subgroup of patients whose working alliance scores increased linearly over time and a subgroup whose alliance scores remained on a relatively high level from the first to the last inspected session. These patterns resemble two of the five classes from the first treatment phase examined in the current investigation: C1.2 ('large improvement–little initial progress') compares to the linearly improving alliance pattern and C1.3 ('small improvement–much initial progress') is akin to the pattern of constantly high alliance. Maybe patients' patterns of working alliance development are similar to their patterns of perceived progress. However, the proportion of patients showing one of these two patterns of perceived progress (90.8%) was somewhat higher than the proportion of patients showing one of these two patterns of working alliance development (72% Stiles et al., 2004). It would be interesting for future research to jointly investigate change patterns in patients' perceived progress and working alliance. By using the same methodology for the identification of patterns in both domains within the same dataset, it might be possible to detect possible relations between the course of patients' perceived progress and working alliance development over time. One suitable approach could be multidimensional GMM, which would allow the joint estimation and identification of patterns in both variables (Stulz & Lutz, 2007).

Future research should also critically examine the issue of phase definitions that have been used in the current investigation. The early phase was chosen to be as consistent as possible to the literature on early response. However, in the literature, there is still no consensual definition of how many sessions should be regarded as early response (e.g., Haas, Hill, Lambert, & Morrell, 2002; Lutz et al., 2009; Stulz et al., 2007). For later phases, the literature is even more sparse and lacks clear definitions. We have chosen the end of the second and third stages so that all three phases comprise the same number of sessions. It could be interesting for future research to identify different treatment phases on a theoretical or empirical basis. A theoretical framework for dividing treatments in several phases might be specific treatment protocols or guidelines. Some protocols or guidelines postulate several interventions and phases, which are connected to specific treatment sessions (e. g., Barlow, 2008).

An empirical rationale of phase definitions could be provided by abrupt shifts in patients' session reports, so called sudden gains and sudden losses. Sudden positive or negative shifts could be one potential marker of endpoints of a phase and the beginning of a new one (Eubanks-Carter et al., 2012; Lutz et al., 2012). On these grounds, for example, the average time until the first substantial gain or loss could possibly be used as an empirical definition of the first treatment phase in future studies.

Connected to the choice of phase intervals is the issue of drop outs. Especially patients showing 'negative change' (C1.5) and those showing 'large improvements and much initial progress' (C1.1) in the first treatment phase tend to drop out rather early and thus were not considered for phase 2 or/and phase 3 pattern identification. This tendency has surely influenced the shape of change patterns in these later treatment phases. With each temporally progressed

treatment phase, fewer patients might still be in treatment, who make very little progress. This development might result in a continuously growing proportion of patients who follow a pattern of small/no change on a relatively high average level of subjectively perceived progress. Clearly good or clearly bad early progressing patients seem to end their treatment, or in the latter case drop out of treatment, rather early.

Another area for future research could be the relation between change patterns and different characteristics of treatments and patients. Two interesting variables would be the kind of treatment (cognitive behavioural therapy [CBT], psychodynamic, interpersonal etc.) and the diagnoses of the patients. Generally, there are at least two different strategies how relations between change patterns and variables such as treatments or diagnoses could be studied: first, change patterns are identified in the whole dataset and afterwards their relation to subgroup indicators are analyzed via chi-square tests. Second, change patterns are identified within each of the different subsamples and the resulting patterns are compared afterwards. The first strategy was applied to compare different therapies in a reanalysis of the data of the National Institute of Mental Health Treatment of Depression Collaborative Research Program trial by Lutz et al. (2009). The authors found no significant association between the type of treatment (CBT, interpersonal psychotherapy, imipramine and pill-placebo) and the different early change patterns within a homogenous sample of depressed patients. However, the number of patients per treatment was too small to have an adequate power to detect treatment differences, and the investigation was restricted to an early treatment phase. Differences between different treatments might also reveal only in later treatment phases or in specific transition patterns.

Furthermore, the diagnostic category a patient is classified into could be an important issue in future research when investigating patterns in patients' change over time. In the present investigation, it was not possible to test for the relation between patients diagnoses and change patterns due to too many cells with an expected frequency under five. However, to get a first impression of possible relations, we conducted this analysis for the first phase of the treatment, by excluding the smallest class (C1.1; 'large improvement–much initial progress'), which circumvented the problem of too many cells under five. For this restricted sample, no relation between patients' diagnoses and early change patterns has been found. For such an endeavour, it might be more promising to separately examine diagnostically homogenous subgroups of patients rather than compare their frequencies in the different classes identified out of the whole sample. As comorbidity would be a matter of concern, especially samples of controlled trials might be appropriate for these kinds of investigations. Prior research has shown that, at least for an early treatment phase, patterns identified within a homogenous sample of depressed patients seem to be somewhat different compared with those identified in heterogeneous naturalistic samples (Stulz et al., 2007; Lutz et al., 2009). Especially the number of patients showing rapid early improvements was considerably higher in a solely depressed patient sample (about 61%; Lutz et al., 2009) than in a heterogeneous sample (about 12%; Lutz et al., 2007). This difference hints at the importance for future research to take into account diagnostic information when searching for patterns of change. As we could not control for the possible influence of patient diagnoses, the results of the present investigation might be in part due to the specific distribution of diagnoses in the patient sample. In the context of increased probabilities, to show particular change trajectories for special diagnostic subgroups and also samples with higher proportions of patients with personality disorders should be regarded in future research. A possible relation between change patterns and personality disorders might be suggested by the relation of pre-treatment levels of interpersonal impairment and class membership. However, the percentage of patients with personality disorder within this study sample was too small for a more in depth investigation.

We further want to note that all analyses are based on measures representing the view of the patients. Although we think that patients are one of the most important sources of information for psychotherapy process and outcome research as well as for the generation of feedback for therapists, their subjectively perceived progress might differ considerably from what therapists or outside observers might indicate. As in other studies, patient ratings within this study could also be biased by specific response styles (e.g., a possible tendency of some patients to use middle response categories), which might have influenced the observed patterns in the current investigation. Therefore, future research should integrate different perspectives on the course of change and compare the differences and similarities, thus approaching a more comprehensive picture of patient change.

A connected problem is the issue of practice effects due to the weekly assessment of progress with the same measure. As mentioned above, especially the reported levelling out of scores on a relatively high level in later phases of the treatment might be a result of an altered use of the measure in later phases compared with early phases (possibly due to a lack of measurement invariance over time). While in early phases the scores might be a more differentiated representation of the subjectively perceived progress, in later phases it might be a more global indicator of contentment with the course of the treatment. A similar phenomenon has been described by Fokkema, Smits, Kelderman and Cuijpers (2013) in a recent study. The authors found a significant response shift in patients beck depression inventory item scores from pre-treatment to post-treatment. Fokkema et al. (2013) interpreted their findings as a result of an

improved ability of patients to assess their depressive symptomatology due to an altered concept of depression. Longitudinal measurement invariance will definitely be an important topic for future patient-focused research as repeated assessments of progress or outcome measures are an essential component of this line of research.

Despite these limitations, results of this study expand prior findings regarding the early course of psychotherapy to a progress measure and gave first insights into change patterns in other than the early treatment phase. The applied modelling approach showed to be a useful method for a more in depth investigation of typical change patterns in different phases of psychotherapy and should be replicated in future research with different instruments and phase definitions. Different typical change patterns seem to be present in different treatment phases and should thus be respectively considered for example in decision support tools for therapists.

REFERENCES

Barlow, D.H. (2008). *Clinical handbook of psychological disorders: A step-by-step treatment manual* (4th ed.). New York: Guilford Press.

Bauer, D.J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*(4), 757–786. doi: 10.1080/00273170701710338

Bauer, D.J. (2011). Evaluating individual differences in psychological processes. *Current Directions in Psychological Science*, *20*(2), 115–118. doi: 10.1177/0963721411402670

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: John Wiley & Sons.

Cuijpers, P., van Lier, P.A.C., van Straten, A., & Donker, M. (2005). Examining differential effects of psychological treatment of depressive disorder: An application of trajectory analyses. *Journal of Affective Disorders*, *89*(1–3), 137–146. doi: 10.1016/j.jad.2005.09.001

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

Duncan, T.E., Duncan, S.C., & Strycker, L.A. (2006). *An introduction to latent variable growth curve modeling: concepts, issues, and applications* (2nd ed.). Mahwah, N.J: Lawrence Erlbaum Associates.

Eubanks-Carter, C., Gorman, B.S., & Muran, J.C. (2012). Quantitative naturalistic methods for detecting change points in psychotherapy research: An illustration with alliance ruptures. *Psychotherapy Research*, 1–17. doi: 10.1080/10503307.2012.693772

Flückiger, C., Grosse Holtforth, M., Znoj, H.J., Caspar, F., & Wampold, B.E. (2013). Is the relation between early post-session reports and treatment outcome an epiphenomenon of intake distress and early response? A multi-predictor analysis in outpatient psychotherapy. *Psychotherapy Research*, *23*(1), 1–13. doi: 10.1080/10503307.2012.693773

Flückiger, C., Regli, D., Zwahlen, D., Hostettler, S., & Caspar, F. (2010). Der Berner Patienten- und Therapeutenstundenbogen 2000. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *39*(2), 71–79. doi: 10.1026/1616-3443/a000015

Fokkema, M., Smits, N., Kelderman, H., & Cuijpers, P. (2013). Response Shifts in mental health interventions: An illustration of longitudinal measurement invariance. *Psychological Assessment* [Epub ahead of print]. doi: 10.1037/a0031669

Franke, G. (2000). *BSI. Brief Symptom Inventory: Deutsche version [BSI. brief symptom inventory: German version] manual*. Göttingen: Beltz.

Haas, E., Hill, R.D., Lambert, M.J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? *Journal of Clinical Psychology*, *58*(9), 1157–1172. doi: 10.1002/jclp.10044

Horowitz, L.M., Strauss, B., & Kordy, H. (2000). *Inventar intrpersonaler probleme (IIP-D): Deutsche version [Inventory of interpersonal problems (IIP-D): German version]*. Weinheim: Beltz Test GmbH.

Howard, K.I., Kopta, S.M., Krause, M.S., & Orlinsky, D.E. (1986). The dose-effect relationship in psychotherapy. *The American psychologist*, *41*(2), 159–164. doi: 10.1037/0003-066X.41.2.159

Howard, K.I., Moras, K., Brill, P.L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *The American Psychologist*, *51*(10), 1059–1064. doi: 10.1037//0003-066X.51.10.1059

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

Kim, S.-Y., & Kim, J.-S. (2012). Investigating stage-sequential growth mixture models with multiphase longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(2), 293–319. doi: 10.1080/10705511.2012.659632

Kivlighan, D.M., Jr., & Shaughnessy, P. (2000). Patterns of working alliance development: A typology of client's working alliance ratings. *Journal of Counseling Psychology*, *47*(3), 362–371. doi: 10.1037/0022-0167.47.3.362

Krause, M.S., Howard, K.I., & Lutz, W. (1998). Exploring individual change. *Journal of consulting and clinical psychology*, *66*(5), 838–845. doi: 10.1037/0022-006X.66.5.838

Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, *17*(1), 1–14. doi: 10.1080/10503300601032506

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*(2), 159–172. doi: 10.1037/0022-006X.69.2.159

Lambert, M.J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, *48*(1), 72–79. doi: 10.1037/a0022238

Luborsky, L., Barber, J.P., Siqueland, L., Johnson, S., Najavits, L.M., Frank, A., & Daley, D. (1996). The revised helping alliance questionnaire (HAq-II): Psychometric properties. *The Journal of psychotherapy practice and research*, *5*(3), 260.

Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.-A., Jorasz, C., Mocanu, S., Vocks, S., Schulte, D., & Tschitsaz-Stucki, A. (2012). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research*, 1–11. doi: 10.1080/10503307.2012.693837

Lutz, W., Saunders, S.M., Leon, S.C., Martinovich, Z., Kosfelder, J., Schulte, D., … & Tholen, S. (2006). Empirically and clinically useful decision making in psychotherapy: Differential predictions with treatment response models. *Psychological Assessment*, *18*(2), 133. doi: 10.1037/1040-3590.18.2.133

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*(1–3), 60–68. doi: 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*(4–5), 502–510. doi: 10.1080/10503300802688486

Morral, A.R., Iguchi, M.Y., Belding, M.A., & Lamb, R.J. (1997). Natural classes of treatment response. *Journal of Consulting and Clinical Psychology*, *65*(4), 673–685. doi: 10.1037/0022-006X.65.4.673

Muthén, B. (2006). The potential of growth mixture modeling. *Infant and Child Development*, *15*(6), 623–625. doi: 10.1002/icd.482

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

Nagin, D.S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, *4*, 139–157. doi: 10.1037/1082-989X.4.2.139

Nagin, D.S., & Odgers, C.L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, *6*(1), 109–138. doi: 10.1146/annurev.clinpsy.121208.131413

Nylund, K.L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and Growth Mixture Modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. doi: 10.1080/10705510701575396

Schulte, D., & Eifert, G.H. (2002). What to Do When Manuals Fail? The Dual Model of Psychotherapy. *Clinical Psychology: Science and Practice*, *9*(3), 312–328. doi: 10.1093/clipsy.9.3.312

Schwartz, G. (1978). Estimating dimensions of a model. *Annals of Statistics*, *6*, 461–464. doi: 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*(3), 298–311. doi: 10.1037/a0019247

Stiles, W.B., Glick, M.J., Osatuke, K., Hardy, G.E., Shapiro, D.A., Agnew-Davies, R., Reese, A., & Barkham, M. (2004). Patterns of alliance development and the rupture-repair hypothesis: Are productive relationships U-shaped or V-shaped? *Journal of Counseling Psychology*, *51*(1), 81–92. doi: 10.1037/0022-0167.51.1.81

Stiles, W.B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D.A., & Iveson, M. (2003). Early sudden gains in psychotherapy under routine clinic conditions: practice-based evidence. *Journal of Consulting and Clinical Psychology*, *71*(1), 14–21. doi: 10.1037/0022-006X.71.1.14

Stulz, N., Gallop, R., Lutz, W., Wrenn, G.L., & Crits-Christoph, P. (2010). Examining differential effects of psychosocial treatments for cocaine dependence: An application of latent trajectory analyses. *Drug and Alcohol Dependence*, *106*(2–3), 164–172. doi: 10.1016/j.drugalcdep.2009.08.009

Stulz, N., & Lutz, W. (2007). Multidimensional patterns of change in outpatient psychotherapy: The phase model revisited. *Journal of Clinical Psychology*, *63*(9), 817–833. doi: 10.1002/jclp.20397

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*(6), 864–874. doi: 10.1037/0022-006X.75.6.864

Tabachnick, B.G., & Fidell, L.S. (2009). *Using multivariate statistics*. Boston: Pearson Education.

Tang, T.Z., & DeRubeis, R.J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, *67*, 894–904. doi: 10.1037/0022-006X.67.6.894

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

Article 3

Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2014). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. ***Psychological Assessment, 27*****(2), 478-488.**

Psychological Assessment
2015, Vol. 27, No. 2, 478–488

1040-3590/15/$12.00 http://dx.doi.org/10.1037/pas0000060

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

Julian Rubel and Wolfgang Lutz
University of Trier

Stephen Mark Kopta
University of Evansville

Katharina Köck
University of Koblenz-Landau

Takuya Minami
University of Massachusetts Boston

Dirk Zimmermann
University of Trier

Stephen M. Saunders
Marquette University

Several different approaches have been applied to identify early positive change in response to psychotherapy so as to predict later treatment outcome and length as well as use this information for outcome monitoring and treatment planning. In this study, simple methods based on clinically significant change criteria and computationally demanding growth mixture modeling (GMM) are compared with regard to their overlap and uniqueness as well as their characteristics in terms of initial impairment, therapy outcome, and treatment length. The GMM approach identified a highly specific subgroup of early improving patients. These patients were characterized by higher average intake impairments and higher pre- to-posttreatment score differences. Although being more specific for the prediction of treatment success, GMM was much less sensitive than clinically significant and reliable change criteria. There were no differences between the groups with regard to treatment length. Because each of the approaches had specific advantages, results suggest a combination of both methods for practical use in routine outcome monitoring and treatment planning.

Keywords: routine outcome monitoring, decision rules, growth mixture modeling, reliable change, clinically significant change

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.

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

This article was published Online First December 15, 2014.

Julian Rubel and Wolfgang Lutz, Department of Clinical Psychology and Psychotherapy, University of Trier; Stephen Mark Kopta, Department of Psychology, University of Evansville; Katharina Köck, Outpatient Clinic for Psychotherapy, University of Koblenz-Landau; Takuya Minami, College of Education and Human Development, University of Massachusetts Boston; Dirk Zimmermann, Department of Clinical Psychology and Psychotherapy, University of Trier; Stephen M. Saunders, Department of Psychology, Marquette University.

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

Correspondence concerning this article should be addressed to Julian Rubel, Department of Clinical Psychology and Psychotherapy, University of Trier, Am Wissenschaftspark 25 + 27, Trier, Germany. E-mail: rubel@uni-trier.de

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 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.

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' 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[1] 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 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 $\alpha = .91$ (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 rs $= -.81$ 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).

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.

[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 M_0/s_0 and M_1/s_1 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.

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*.

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[2] (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 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.

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).

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.

[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; $\alpha = .91$) 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).

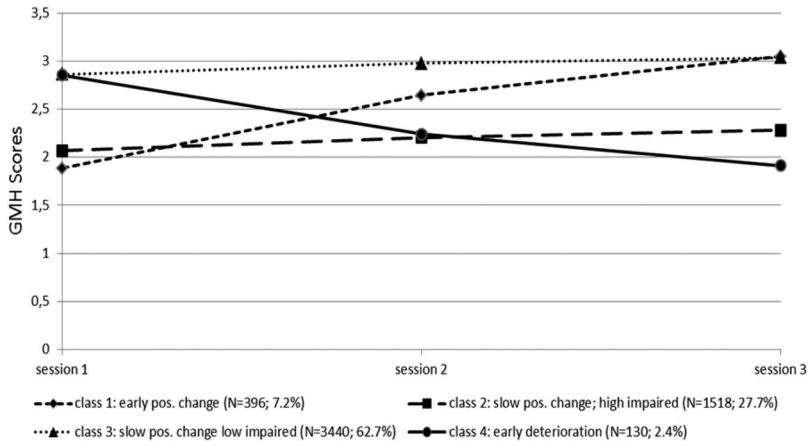

Figure 1. Estimated mean change trajectories over the first three sessions for a four-class growth mixture model solution.

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 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).

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.

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 (ds) and categorized change statuses after treatment (reliably improved, clinically 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 (ds = 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.

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

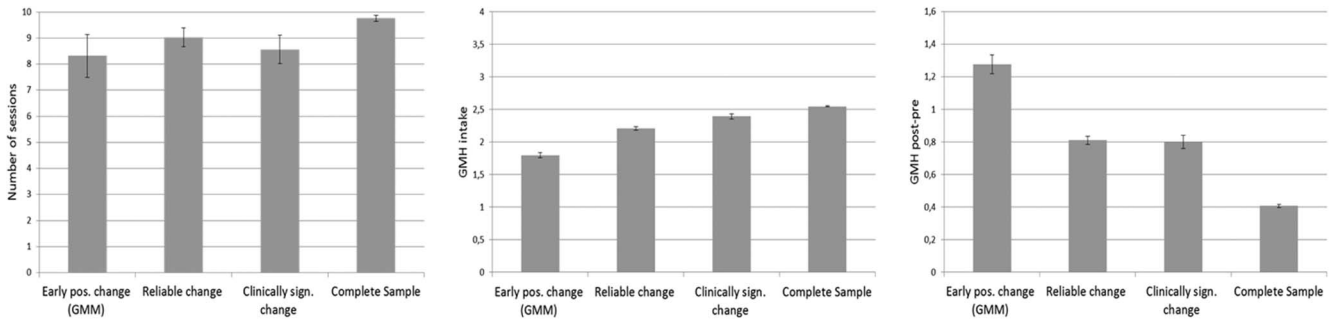

Figure 2. Mean numbers of sessions, mean intake general mental health (GMH) scores, mean differences between pre- and posttreatment GMH scores (high values indicating high positive changes from pretreatment to posttreatment), and 95% confidence intervals for each of the early positive response groups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

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 spec-

Table 2
*Frequencies and Final Treatment Outcomes (*d*s 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.

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.

ificity 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).

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.

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

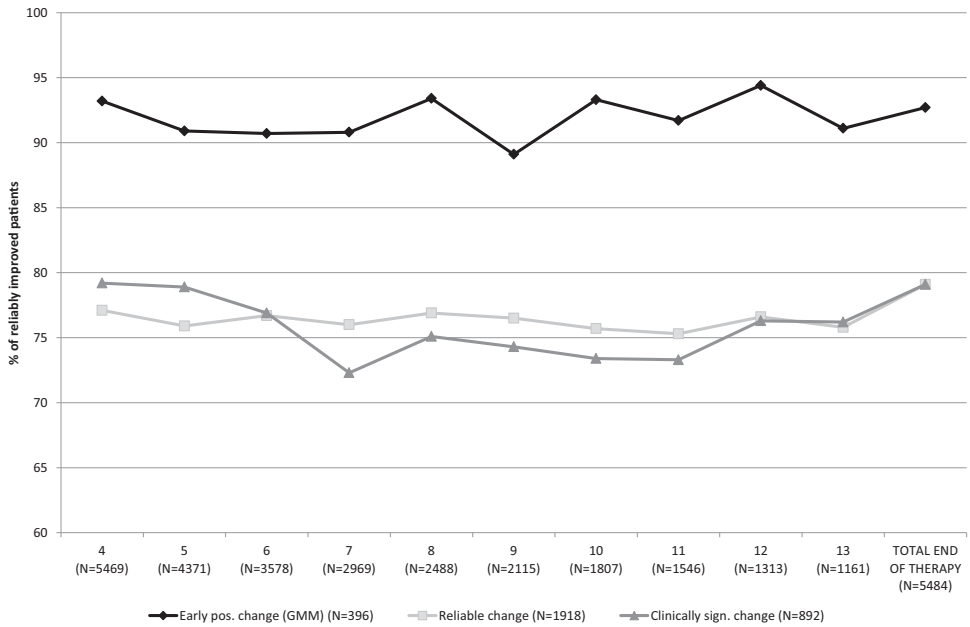

Figure 3. Percentages of reliably improved patients after Sessions 4–13 and after the end of the treatment for 5-patient subgroups defined on the basis of their change status after Session 3.

patient's score has to move from the range above an a priori defined cutoff score into the range below that cutoff score. Consequently, 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 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 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.

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 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 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.

References

Barlow, D. H. (2010). Negative effects from psychological treatments: A perspective. *American Psychologist, 65,* 13–20. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1080/10503300802688486

Martinovich, Z., Saunders, S., & Howard, K. (1996). Some comments on "assessing clinical significance." *Psychotherapy Research, 6,* 124–132. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1002/cpp.1868

Schwartz, G. (1978). Estimating dimensions of a model. *Annals of Statistics, 6,* 461–464. http://dx.doi.org/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. http://dx.doi.org/10.1037/a0019247

Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60,* 402–408. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1037/0022-006X.75.6.864

Received January 9, 2014
Revision received October 6, 2014
Accepted October 9, 2014 ■

Eidesstattliche Erklärung

Ich versichere, dass ich meine Dissertation ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Trier, den

Nachname: Rubel Vorname: Julian

Unterschrift: _____