Published in final edited form as:

J Consult Clin Psychol. 2014 April ; 82(2): 287–297. 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, SUNY

M. Katherine Shear,
Columbia University

Jack M. Gorman,
Franklin Behavioral Health Consultants

Scott W. Woods, and
Yale University

David H. Barlow
Boston University

Abstract

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 five 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 (CBT) 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 classification based on pre to post treatment comparisons (e.g., Aaronson et al., 2008). This approach defines change patterns based on the assumption that those patients who experience positive outcome at the end of the treatment have followed a positive response pattern. On the other hand, 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.

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., Nagin & Odgers, 2010; Muthén, 2006). 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 that 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 based on 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 found that quick, positive treatment response within the first three sessions related to positive outcome at termination and follow-up (e.g., Haas, Hill, Lambert & Morell, 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 utilized 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 6 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 6 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 6 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 that used a specific diagnostic group, major depression, 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 pre-treatment 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). Based on the rating of response status by a trained independent evaluator, patients were then triaged into two clinical trials. Responders were randomized to nine months of monthly booster sessions or no booster sessions, and then followed for one year (White et al., 2013). Non-responders were randomized to either three months of continued CBT or to three months of paroxetine, and then followed for an additional nine 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) was used to identify patterns of change over the course of treatment. Based on 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 4 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, 6 (1.8%) patients provided no information on their ethnicity. Treatment duration ranged from three to eleven 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/drop-out) were reported in Aaronson et al. (2008) and White et al. (2010), respectively (see Grilo et al., 1998 and 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. The present investigation focused on session-by-session PDSS-SR ratings collected at each visit during the acute open treatment phase of this trial.

Measures

Panic Disorder Severity Scale – Self-Report Version (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 current study, the sum of these items was used. The test-retest reliability for the PDSS-SR is $r_{tt} = .81$ and internal consistency is $\alpha = .92$ (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

¹The RCI is defined as the pre-treatment–post-treatment difference ΔRC that is large enough to be not attributable to measurement error and is calculated as: $\Delta_{RC} = z_{RC} \cdot \sqrt{2(SD \sqrt{1-r})^2} = 1.96 \cdot \sqrt{2(6.6 \cdot \sqrt{1-.92})^2} = 5.17$ 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).

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 $\alpha = 0.93$; White et al., 2010).

Work and Social Adjustment Scale (WSAS-SR; Mundt, Marks, Shear, & Greist, 2002)—The WSAS-SR is a 5-item self-report measure that assesses the degree of functional interference caused by symptoms in 5 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 $\alpha = 0.80$; 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 current investigation, the first five sessions were defined as the early treatment phase since all active ingredients of the treatment protocol had been delivered by the end of session five (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 variables 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 post-treatment.

Reliable change criteria as described above were used at the post-treatment assessment to classify patients into groups of *reliably improved* (pre to post improvement larger than the RCI of the PDSS-SR), *reliably deteriorated* (pre to post deterioration larger than the RCI) and *not reliably changed* (pre to post change that is not different from the RCI).

Associations between these change categories and early change patterns were investigated with χ^2 -tests. To identify predictors of latent class membership, separate ANOVAs and multinomial logistic regression analyses were conducted. Additionally, the pre-treatment variables described above were tested as potential predictors of treatment outcome and length of treatment.

Growth Mixture Modeling (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, 2006; 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, while 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 estimated³ (for similar approaches, also see Colder et al., 2002; Hunter, Muthen, 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; Schwartz, 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.

²Models 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.

³We also tested the more restrictive (LCGA) 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 8 and the latter in 2 latent change patterns. Both model solutions resulted in less clinically meaningful classes with lower predictive power than the final model.

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 5 sessions and no further change can be achieved after that.

An intercept-only linear model and a more complex quadratic model were compared to 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 analyses⁴. 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) (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 six to eleven. 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 ($M = 9.0$; 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 one to five. 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 five, 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).

⁴We re-ran 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.

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 six and eleven (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).

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 to 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 χ^2 -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 to 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 to 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 to 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.

Early Change Patterns, Treatment Outcome and Early Treatment Termination

Table 4 shows the relative frequencies of the PDSS-SR RCI *classes*, pre- to post effect sizes of change in PDSS-SR (d), number of treatment sessions attended (3–5 sessions, 6–10 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 ($\chi^2_{(df = 3)} = 77.79$)

and treatment length ($\chi^2_{(df=6)} = 16.80$) were analyzed with χ^2 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 post-treatment ($d = 2.11$), 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 post treatment ($d = -0.49$). 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 post effect 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 to 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 eleven sessions, whereas 10.1% left treatment after three to five sessions, and further 13.2% left treatment prior to session eleven. 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 three and five. These two classes also showed the highest rates of patients who ended treatment between the sixth and the tenth session (see Table 4). Approximately 60% of the patients in those two classes stayed in treatment until session eleven. 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* 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 post-treatment 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 comprised 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 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 post-treatment 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

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

frequency:
$$\text{standardized residual} = \frac{\text{observed}_{ij} - \text{model}_{ij}}{\sqrt{\text{model}_{ij}}}$$
; 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 χ^2 test and the relevance of the detected class.

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 pre-treatment predictors of treatment outcome. For each of these two predictors, a higher score at intake was associated with lower effect sizes from pre- to post-treatment when controlling for the other variables included in step two (see Table 5).

Adding the dummy coded variables for class membership in step three 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 three), 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 two, became non-significant in step three 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 one 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 5 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. Utilizing 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 to 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 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

⁶To evaluate the consequences of using patients with at least 3–5 sessions in the models, we re-ran all the analyses presented in Table 5 with only those individuals who completed 6 or more sessions and obtained consistent results.

explanation for this difference could be the difference in how “early” treatment was defined. Although in the current study the first 5 sessions were examined, the first 8 weeks were modeled in the NIMH TDCRP dataset. 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 post-treatment levels of depression and anxiety. The highly negative association between levels of pre-treatment 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 worth further investigation.

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

Only two other studies that investigated change patterns with a cluster analytic method reported a class of early deteriorating patients (Lutz et al., 2005; Morral et al., 1997), which we found in the current 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 drop-out. 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 highly 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, while 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 non-fixed 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 3 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 current 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). Based on the assessments of independent evaluators and the completer dataset, these studies reported slightly higher response rates (50.3% instead of 48.8%) than those based on the reliable change index used in the current investigation. The evaluator-rated response status proved to be quite stable among responders over an ensuing nine month period (White et al., 2013). Unfortunately, these independent evaluator assessments were only available for the completer sample and, thus, not applicable to the current 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 4 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 has been a topic of much discussion and debate (Bauer, 2007; Bauer, 2011; Nagin & Odgers, 2010; Nagin & Tremblay, 2005). Simulation studies have shown that modest specification errors could result in an over extraction 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.

Acknowledgments

This work was supported by National Institute of Mental Health grant R01 MH45963 (Dr. Gorman, principal investigator (PI)), MH45964 (Dr. Shear, PI), MH45965 (Dr. Barlow, PI), and MH45966 (Dr. Woods, PI). Dr. Lutz is supported by grants from the Humboldt Foundation (TransCoop Program, TCVERL-DEU/1133562).

References

- Aaronson CJ, Shear MK, Goetz RR, Allen LB, Barlow DH, White KS, Gorman JM. Predictors and time course of response among panic disorder patients treated with cognitive-behavioral therapy. *The Journal of Clinical Psychiatry*. 2008; 69:418–424. [PubMed: 18278989]
- Aderka IM, Nickerson A, Bøe HJ, Hofmann SG. Sudden Gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology*. 2012; 80:93–101.10.1037/a0026455 [PubMed: 22122290]
- Barlow, DH.; Craske, MG. *Mastery of your anxiety and panic*. Oxford: Oxford University Press; 2007.
- Barlow DH. Negative effects from psychological treatments: A perspective. *American Psychologist*. 2010; 65:13–20.10.1037/a0015643 [PubMed: 20063906]
- Bauer DJ. Observations on the use of Growth Mixture Models in psychological research. *Multivariate Behavioral Research*. 2007; 42:757–786.10.1080/00273170701710338
- Bauer DJ. Evaluating individual differences in psychological processes. *Current Directions in Psychological Science*. 2011; 20:115–118.10.1177/0963721411402670
- Boswell JF, Sauer-Zavala SE, Gallagher MW, Delgado NK, Barlow DH. Readiness to change as a moderator of outcome in transdiagnostic treatment. *Psychotherapy Research*. 2012; 22:570–578.10.1080/10503307.2012.688884 [PubMed: 22607634]
- Colder CR, Campbell RT, Ruel E, Richardson JL, Flay BR. A finite mixture model of growth trajectories of adolescent alcohol use: Predictors and consequences. *Journal of Consulting and Clinical Psychology*. 2002; 70:976–985.10.1037/0022-006X.70.4.976 [PubMed: 12182281]
- Cuijpers P, van Lier PAC, van Straten A, Donker M. Examining differential effects of psychological treatment of depressive disorder: An application of trajectory analyses. *Journal of Affective Disorders*. 2005; 89:137–146.10.1016/j.jad.2005.09.001 [PubMed: 16274750]
- Duncan, TE.; Duncan, SC.; Strycker, LA. *An introduction to latent variable growth curve modeling*. 2. New York: Psychology Press; 2010.
- Gibbons RD, Hedeker D, Elkin I, Waterneaux C, Kraemer H, Greenhouse JB, Shea TM, Imber S, Sotsky SM, Watkins JT. Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Archives of General Psychiatry*. 1993; 50:739–750. [PubMed: 8357299]
- Grilo CM, Money R, Barlow DH, Goddard AW, Gorman JM, Hofmann SG, Papp LA, Shear MK, Woods SW. Pretreatment patient factors predicting attrition from a multicenter randomized controlled treatment study for panic disorder. *Comprehensive Psychiatry*. 1998; 39:323–331.10.1016/S0010-440X(98)90043-8 [PubMed: 9829138]
- Gunlicks-Stoessel M, Mufson L. Early patterns of symptom change signal remission with interpersonal psychotherapy for depressed adolescents. *Depression and Anxiety*. 2011; 28:525–531.10.1002/da.20849 [PubMed: 21721071]
- Haas E, Hill RD, Lambert MJ, Morrell B. Do early responders to psychotherapy maintain treatment gains? *Journal of Clinical Psychology*. 2002; 58:1157–1172.10.1002/jclp.10044 [PubMed: 12209871]
- Hafner J, Marks IM. Exposure in vivo of agoraphobics: Contributions of diazepam, group exposure and anxiety evocation. *Psychological Medicine*. 1976; 6:71–88.10.1017/S0033291700007510 [PubMed: 6985]
- Hamilton M. The assessments of anxiety states by rating. *British Journal of Medical Psychology*. 1959; 32:50–55.10.1111/j.2044-8341.1959.tb00467.x [PubMed: 13638508]
- Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*. 1960; 23:56–62.10.1136/jnnp.23.1.56
- Hofmann SG, Barlow DH, Papp LA, Detweiler M, Ray S, Shear MK, Woods SW, Gorman JM. Pretreatment attrition in a comparative treatment outcome study on panic disorder. *American Journal of Psychiatry*. 1998; 155:43–47. [PubMed: 9433337]
- Hofmann SG, Schulz SM, Meuret AE, Moscovitch DA, Suvak M. Sudden gains during therapy of social phobia. *Journal of Consulting and Clinical Psychology*. 2006; 74:687–697.10.1037/0022-006X.74.4.687 [PubMed: 16881776]

- Hofmann SG, Smits JAJ. Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *The Journal of Clinical Psychiatry*. 2008; 69:621–632.10.4088/JCP.v69n0415 [PubMed: 18363421]
- Houck PR, Spiegel DA, Shear MK, Rucci P. Reliability of the self-report version of the Panic Disorder Severity Scale. *Depression and Anxiety*. 2002; 15:183–185.10.1002/da.10049 [PubMed: 12112724]
- Hunter AM, Muthén BO, Cook IA, Leuchter AF. Antidepressant response trajectories and quantitative electroencephalography (QEEG) biomarkers in major depressive disorder. *Journal of psychiatric research*. 2010; 44:90–98.10.1016/j.jpsychires.2009.06.006 [PubMed: 19631948]
- Kazdin AE. Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*. 2007; 3:1–27.10.1146/annurev.clinpsy.3.022806.091432
- Lambert MJ. Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*. 2007; 17:1–14.10.1080/10503300601032506
- Lewis CC, Simons AD, Kim HK. The role of early symptom trajectories and pretreatment variables in predicting treatment response to cognitive behavioral therapy. *Journal of Consulting and Clinical Psychology*. 2012; 80:525–534.10.1037/a0029131 [PubMed: 22730951]
- Lutz W. Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirically based clinical practice. *Psychotherapy Research*. 2002; 12:251–272.10.1080/713664389
- Lutz W, Ehrlich T, Rubel J, Hallwachs N, Röttger MA, Jorasch C, Tschitsatz-Stucki A. The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research*. 2013; 23:14–24.10.1080/10503307.2012.693837 [PubMed: 22708586]
- Lutz W, Leach C, Barkham M, Lucock M, Stiles WB, Evans C, Noble R, Iveson S. 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*. 2005; 73:904–913.10.1037/0022-006X.73.5.904 [PubMed: 16287390]
- Lutz W, Stulz N, Köck K. Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*. 2009; 118:60–68.10.1016/j.jad.2009.01.019 [PubMed: 19217669]
- Maier W, Buller R, Philipp M, Heuser I. The Hamilton Anxiety Scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. *Journal of Affective Disorders*. 1988; 14:61–68.10.1016/0165-0327(88)90072-9 [PubMed: 2963053]
- Manicavasagar V, Silove D, Wagner R, Hadzi Pavlovic D. Parental representations associated with adult separation anxiety and panic disorder agoraphobia. *Australian and New Zealand Journal of Psychiatry*. 1999; 33:422–428.10.1046/j.1440-1614.1999.00566.x [PubMed: 10442800]
- Morral AR, Iguchi MY, Belding MA, Lamb RJ. Natural classes of treatment response. *Journal of Consulting and Clinical Psychology*. 1997; 65:673–685.10.1037/0022-006X.65.4.673 [PubMed: 9256569]
- Mundt JC, Marks IM, Shear MK, Greist JM. The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*. 2002; 180:461–464.10.1192/bjp.180.5.461 [PubMed: 11983645]
- Muthén B. The potential of growth mixture modelling. *Infant and Child Development*. 2006; 15:623–625.10.1002/icd.482
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. 6. Los Angeles, CA: Muthén & Muthén; 2010.
- Nagin DS. Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*. 1999; 4:139–157.10.1037/1082-989X.4.2.139
- Nagin DS, Odgers CL. Group-Based Trajectory Modeling in clinical research. *Annual Review of Clinical Psychology*. 2010; 6:109–138.10.1146/annurev.clinpsy.121208.131413
- Nagin DS, Tremblay RE. Developmental trajectory groups: Fact or a useful statistical fiction? *Criminology*. 2005; 43:873–904.10.1111/j.1745-9125.2005.00026.x
- Nylund KL, Asparouhov T, Muthén BO. 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*. 2007; 14:535–569.10.1080/10705510701575396

- Otto MW, Tolin DF, Nations KR, Utschig AC, Rothbaum BO, Hofmann SG, Smits JA. Five sessions and counting: Considering ultra-brief treatment for panic disorder. *Depression and Anxiety*. 2012; 29:465–470.10.1002/da.21910 [PubMed: 22730311]
- Payne LA, White KS, Gallagher MW, Woods SW, Shear MK, Gorman JM, Farchione TJ, Barlow DH. 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.
- Schwartz G. Estimating the dimensions of a model. *Annals of Statistics*. 1978; 6:461–464.10.1214/aos/1176344136
- Shear MK, Brown TA, Barlow DH, Money R, Sholomskas DE, Woods SW, Gorman JM, Papp LA. Multicenter collaborative Panic Disorder Severity Scale. *American Journal of Psychiatry*. 1997; 154:1571–1575. [PubMed: 9356566]
- Shear MK, Rucci P, Williams J, Frank E, Grochocinski V, Vander Bilt J, Wang T. Reliability and validity of the panic disorder severity scale: Replication and extension. *Journal of Psychiatric Research*. 2001; 35:293–296.10.1016/S0022-3956(01)00028-0 [PubMed: 11591432]
- Speer DC. Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology*. 1992; 60:402–408.10.1037/0022-006X.60.3.402 [PubMed: 1619094]
- Stewart JW, Quitkin FM, McGrath PJ, et al. 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*. 1998; 55:334–343.10.1001/archpsyc.55.4.334 [PubMed: 9554429]
- Stulz N, Gallop R, Lutz W, Wrenn GL, Crits-Christoph P. Examining differential effects of psychosocial treatments for cocaine dependence: An application of latent trajectory analyses. *Drug and Alcohol Dependence*. 2010; 106:164–172.10.1016/j.drugalcdep.2009.08.009 [PubMed: 19782480]
- Stulz N, Lutz W, Leach C, Lucock M, Barkham M. Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*. 2007; 75:864–874.10.1037/0022-006X.75.6.864 [PubMed: 18085904]
- Tofghi, D.; Enders, CK. Identifying the correct number of classes in a growth mixture model. In: Hancock, GR., editor. *Mixture models in latent variable research*. Greenwich, CT: Information Age; 2007. p. 317–341.
- Uher R, Muthén B, Souery D, Mors O, Jaracz J, Placentino A, McGuffin P. Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine*. 2010; 40:1367–1377.10.1017/S0033291709991528 [PubMed: 19863842]
- van Calker D, Zobel I, Dykierck P, Deimel CM, Kech S, Lieb K, Schramm E. Time course of response to antidepressants: Predictive value of early improvement and effect of additional psychotherapy. *Journal of Affective Disorders*. 2009; 114:243–253.10.1016/j.jad.2008.07.023 [PubMed: 18849079]
- White KS, Allen LB, Barlow DH, Gorman JM, Shear MK, Woods SW. Attrition in a multicenter clinical trial for panic disorder. *The Journal of Nervous and Mental Disease*. 2010; 198:665–671.10.1097/NMD.0b013e3181ef3627 [PubMed: 20823729]
- White KS, Payne LA, Gorman JM, Shear MK, Woods SW, Saska JR, Barlow DH. Does maintenance CBT contribute to long-term response of panic disorder? A randomized controlled clinical trial. *Journal of Consulting and Clinical Psychology*. 2013; 81:47–57.10.1037/a0030666 [PubMed: 23127290]

Figure 1.
Mean latent growth curves for GMM solution with four latent classes within the first five sessions, and observed mean scores in the respective classes during sessions six to eleven.

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	0.0000	0.0073
3	8788.727	8744.319	8735.710	8737.060	0.595	0.0000	0.2195
4	8779.074	8725.151	8714.697	8716.684	0.686	0.0000	0.0103
5	8787.684	8724.245	8711.946	8714.700	0.721	0.2667	0.4320
6	8790.180	8717.226	8703.081	8706.737	0.733	0.0128	0.0584
7	8795.022	8712.551	8696.562	8701.258	0.749	0.0300	0.3556
8	8807.364	8715.378	8697.544	8703.422	0.740	0.6667	0.2209

Note. BIC = Bayesian Information Criterion; SABIC = Sample Size Adjusted BIC; AIC = Akaike Information Criterion; Corrected AIC; BLRT = Bootstrapped Likelihood Ratio Test; LRT = Lo-Mendel-Rubin Likelihood Ratio Test

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

Table 3

Prediction of group membership by patient intake characteristics via multinomial logistic regression analyses.

95% Confidence Interval for Odds Ratio					
	B (SE)	p	Lower	Odds Ratio	Upper
Class 1 vs. Class 2					
Intercept	2.51 (0.58)	<.001			
ASAC	−0.03 (0.22)	.320	.97	1.03	1.10
WSAS	0.24 (0.03)	.279	.52	.79	1.21
Class 1 vs. Class 3					
Intercept	−1.69 (0.73)	0.200			
ASAC	0.02 (0.04)	0.507	0.96	1.02	1.10
WSAS	0.81 (0.24)	0.001	1.41	2.25	3.60
Class 1 vs. Class 4					
Intercept	−0.75 (0.67)	0.262			
ASAC	0.00 (0.04)	0.982	0.93	1.00	1.07
WSAS	0.76 (0.23)	0.001	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	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	0.52 (0.11)	<.001	1.36	1.69	2.10
Class 3 vs. Class 4					
Intercept	0.94 (0.54)	.082			

95% Confidence Interval for Odds Ratio				
	B (SE)	p	Lower	Upper
ASAC	−0.02 (0.02)	.239	0.94	0.98
WSAS	−0.05 (0.13)	.679	0.74	0.95
				1.21

Note. $R^2 = .25$ (Cox & Snell), .29 (Nagelkerke). 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.

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.

	<i>n</i>	Final treatment outcome		Treatment completion status (number of sessions attended)				
		Reliable Improvement (%)	ES Change in PDSS-SR during treatment (<i>d</i>) [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	
<i>p</i>		<.001 ^b	<.001 ^c		<.001 ^b			.007 ^c

Note.

^aPDSS-SR scores from pre to post treatment.

^b χ^2 - tests were performed testing the association between class membership and categorized treatment outcome and length.

^cOne-way ANOVAs were performed testing the association between class membership and mean pre- to post change and mean treatment length.

* standardized residual >1.96.

Table 5

Hierarchical multiple regression analyses predicting final treatment outcome (standardized pre to post differences) and treatment length (log-transformed; base 10) by patient intake characteristics and patterns of early change

Predictor	Change During Treatment (Effect sizes)				Treatment Length			
	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.