

Factors of treatment success in psychotherapy: a within-therapist analysis of early session processes

Yan Zhu, Mark J. Hilsenroth

Derner School of Psychology, Adelphi University, New York, USA

ABSTRACT

While psychotherapy is generally effective, outcomes can vary significantly across patients, even when treated by the same therapist. This study employed a mixed-methods design to examine early indicators of treatment success by analyzing contrasting outcomes within individual therapists' caseloads. Treatment success was determined by patients achieving a clinically significant change on the Brief Symptom Inventory's Global Severity Index (BSI-GSI). The Reliable Change Index (RCI) was used to calculate these changes, adjusting to account for measurement error and regression to the mean. Unsuccessful cases are defined as those with no reliable change or deterioration. Data from six patients treated by three psychodynamic therapists were analyzed (each therapist having one successful and one unsuccessful case). Quantitative analyses of psychological assessments and early treatment measures, along with qualitative analyses of third session transcripts, were conducted. In successful cases, the patients reported higher levels of depression, interpersonal distress, and social avoidance/anxiety, but also greater personal insight prior to the treatment. Patients in successful cases also rated lower session smoothness compared to their therapists' ratings early in treatment. These divergent views are likely due to higher ratings of therapists' exploration of uncomfortable feelings in the successful treatment sessions. In addition, these external raters observed that in the successful cases, therapists employed higher overall levels of psychodynamic-interpersonal techniques during the third session than in their unsuccessful cases. Results highlight the importance of how initial severity and emotional expression in the early stage of psychotherapy may help lay the foundation for successful treatment outcomes.

Correspondence: Yan Zhu, Derner School of Psychology, Adelphi University, 158 Cambridge Ave, Garden City, 11530 NY, USA.
E-mail: yanzhu@mail.adelphi.edu

Citation: Zhu, Y., & Hilsenroth, M. J., (2025). Factors of treatment success in psychotherapy: a within-therapist analysis of early session processes. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 28(2), 848. doi: 10.4081/ripppo.2025.848

Conflict of interest: the authors have no conflict of interest to disclose.

Ethical approval and consent to participate: this study was reviewed and deemed exempt by the Adelphi University Institutional Review Board (Submission #021323, approved March 7, 2023).

Informed consent: informed consent was obtained from all participants included in the study.

Availability of data and materials: data and materials are available from the corresponding author upon reasonable request.

Received: 12 January 2025.
Accepted: 5 June 2025.

Publisher's note: all claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

©Copyright: the Author(s), 2025
Licensee PAGEPress, Italy
Research in Psychotherapy:
Psychopathology, Process and Outcome 2025; 28:848
doi:10.4081/ripppo.2025.848

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Key words: psychotherapy outcomes, reliable change, within-therapist effects, mixed-methods research.

Introduction

Psychotherapy is generally effective for a wide range of mental health issues; however, a significant number of patients fail to improve or even deteriorate during treatment (e.g., Hansen *et al.*, 2002; Howick *et al.*, 2022; Lambert, 2011; Tryon *et al.*, 2018). Studies are mixed on which patient, therapist, or treatment characteristics are most related to success or failure in psychotherapy. Research on initial symptom severity as a predictor of treatment outcomes has yielded conflicting results. Some studies indicate that higher initial severity predicts poorer outcomes. A meta-analysis study of Internet-based cognitive behavior therapy (ICBT) conducted by Rozental *et al.* (2019) found that patients with higher symptom severity at baseline have a higher rate of not responding to treatment. Furthermore, a multilevel modeling analysis study by Firth *et al.* (2015) discovered that patients' symptom severity at intake predicts the outcome. Specifically, patients with more severe symptoms of depression, anxiety, and impaired social functioning had poorer outcomes. However, other research indicates that higher initial severity may predict better outcomes. Bower *et al.* (2013) found a small but significant interaction, suggesting that more severely depressed patients may show slightly larger treatment effects. Similarly, Scholten *et al.* (2023) conducted a meta-analysis and found that higher social anxiety at baseline predicted better outcomes with cognitive behavioral therapy (CBT). These studies suggest that the relationship between initial severity and treatment success may be more nuanced, potentially reflecting im-

portant differences in how severity interacts with other patient characteristics. For example, Ackerman *et al.* (2000) found that patients with more disturbed interpersonal patterns and negative relationship expectations, but who maintained a capacity for emotional investment in relationships, were less likely to terminate and more likely to continue in psychodynamic treatment. Similar to this finding, one study using machine learning to predict treatment response in short-term dynamic psychotherapy found that non-responders (40% of patients) were characterized by higher initial distress and showed key predictors of poor outcomes, including difficulties in emotion regulation, interpersonal coldness, avoidant attachment, and low conscientiousness (Yonatan-Leus *et al.*, 2024). These findings suggest that understanding treatment success requires moving beyond simple severity measures to consider the complex interplay between symptom levels, patients' psychological capacities and interpersonal functioning.

In addition to patient variables, studies examining the effectiveness of therapy have consistently found that some therapists achieve better outcomes than others, even when treating similar types of patients within the same therapeutic approach. Luborsky *et al.* (1997) compared outcomes across 22 therapists and seven patient samples, concluding that differences in patient outcomes could not be explained by patients' severity or background characteristics and that the therapists included in the study differed significantly from each other in terms of effectiveness. Moreover, therapists who succeeded in one sample also succeeded in other samples. In a large-scale study of 1,841 patients treated by 91 therapists, Okiishi *et al.* (2003) found that some therapists were consistently more likely to have better treatment outcomes, although they were unable to identify specific therapist traits that accounted for these differences.

This variability in outcomes has led researchers to investigate therapist effects, which refer to the systematic differences in patient outcomes that can be attributed to the individual therapist, rather than to the specific treatment method or patient characteristics (e.g., Baldwin & Imel, 2013; Wampold & Imel, 2015). In a meta-analysis on treatment outcomes, Baldwin *et al.* (2011) suggested that the magnitude of therapist effects is typically estimated to account for around 5% of the variance in treatment outcomes. This finding aligns with Saxon and Barkham's (2012) large-scale study of 119 therapists treating 10,786 patients, which found that 6.6% of outcome variance could be attributed to therapists.

Studies have identified several important therapist-related factors that affect treatment success. Earlier research showed that general therapist characteristics, such as years of experience or theoretical orientation, do not strongly predict outcomes (Wampold & Brown, 2005). However, Baldwin *et al.* (2007) employed mixed-effects modeling to demonstrate that therapists who consistently formed stronger therapeutic alliances achieved better treatment outcomes. Similarly, Wampold and Imel (2013) report that the therapeutic alliance between therapist and patient has consistently and significantly contributed to the therapist effects. Consistent with this view, Zimmermann *et al.* (2021) found that a better therapeutic bond at session 3 predicted better outcomes.

Additional research has identified specific therapist interpersonal qualities that contribute to treatment success. Hilsenroth *et al.* (2012) demonstrate that empathy and alliance-building skills significantly contribute to patient improvement. Consistently, Lingardi *et al.* (2018) reviewed literature on psychodynamic psychotherapy and found that several

therapist traits are associated with treatment outcome. Specifically, they found that therapists with a more supportive and engaging approach, characterized by being warm, helpful, protective, and even gently persuasive, are effective in achieving more positive outcomes. Furthermore, research focused on patients' experiences found that a positive experience is often linked to therapist responsiveness and attunement to patients' needs (Levitt *et al.*, 2016).

Studies focused on unsuccessful treatment outcomes also yield similar findings, providing additional perspectives on the level of compatibility between therapist and patient, as well as therapist flexibility. A therapist's ability to mend ruptures in their relationship with the patient can prevent treatment failure (Eubanks-Carter *et al.*, 2018), while therapists' inflexibility can lead to adverse effects on psychotherapy (Curran *et al.*, 2019). A study focused on patients' experiences in unsuccessful therapies found that patients most commonly attributed treatment failure or early termination to three main factors: therapists' poor understanding and assessment of their problems, inflexible therapeutic approaches, and insufficient clinical knowledge (Alfonsson *et al.*, 2024). Vybíral *et al.* (2024) also identified several clusters of negative therapeutic experiences, including therapist misconduct, detrimental aspects of the therapeutic relationship, such as clients' feelings of insecurity, distrust, and confusion; and mismatched treatment approaches, for example, when clients needed more structure and concrete advice than the treatment provided. A recent study by Li *et al.* (2024), which specifically examined patient experiences within the UK's National Health Service, highlighted patients' preference for personalized treatment over a one-size-fits-all approach, the importance of the therapeutic alliance, and the demand for depth in therapy.

Research on within-therapist variables has advanced our understanding by examining the differences in outcomes across a single therapist's caseload, providing a more nuanced perspective on the factors influencing psychotherapy effectiveness. Earlier research (Zuroff *et al.*, 2010) identified substantial variation both within and between therapists, while within-therapist variance sometimes exceeded between-therapist differences (Baldwin, 2011). Boswell *et al.* (2013) further observed that individual therapists often demonstrated varying levels of competence across different patients. However, these studies, while establishing the existence of within-therapist effects, did not fully explicate the factors contributing to this variance.

Some studies have also found interactions between therapist and patient variables that influence treatment outcomes. Saxon and Barkham (2012) discovered that therapist effects are larger for more severe patients, indicating that the impact of the individual therapist on treatment outcomes may be more pronounced for patients with higher levels of symptom severity, suggesting a potential interaction between therapist variables and patient variables. Webart *et al.* (2019a) studied six successful cases and found that motivated therapists who reported themselves as being a good match for the particular patient contributed to positive outcomes at the outset of treatment.

Mixed-method studies have also emerged as valuable tools for understanding therapist differences in psychotherapy outcomes. In a qualitative study, Schattner *et al.* (2017) examined two contrasting cases within a single therapist's caseload, one being successful and the other being unsuccessful. Through clinical interviews and analysis of clients' childhood relational patterns with parents, they demonstrated that within-therapist effects could be explained through the interaction between the

therapist's and clients' developmental relational patterns. Werbart *et al.* (2019b) also employed a mixed-method design to conduct a deeper exploration of therapist effects. They categorized cases based on treatment outcomes, defining successful cases as those demonstrating clinically significant symptom reduction and unsuccessful cases as those showing no reliable improvement or deterioration. Their findings revealed that successful cases were characterized by three key elements: mutual understanding between therapist and patient regarding presenting problems, clearly defined treatment goals, and the therapist's ability to adapt their approach to meet patient-specific needs. Conversely, less successful cases were marked by discrepancies in perspectives between therapist and patient, with therapists often failing to effectively address challenges within the therapeutic process.

Building on Werbart *et al.*'s research, our study uses a mixed-method design to identify factors contributing to varying treatment outcomes within individual therapists' caseloads. While Werbart *et al.* focused on general treatment processes, our study specifically targets third session dynamics. The choice of third session analysis is empirically grounded, as Flückiger *et al.* (2018) demonstrated in their comprehensive meta-analysis that early alliance measures, particularly those assessed in sessions 3-5, predict treatment outcomes more strongly than later alliance assessments, highlighting the importance of examining this initial therapeutic timeframe.

Following Werbart *et al.*'s methodology, we categorize cases as successful and unsuccessful based on empirical patient outcomes at treatment termination using the Brief Symptom Inventory (BSI). However, we expand their approach through the integration of quantitative process measures with qualitative analysis of third session transcripts. This mixed-method examination of early treatment interactions represents a methodological advancement, allowing us to identify potential indicators of treatment success when therapeutic interventions can still be meaningfully adjusted.

By comparing contrasting cases within each therapist's caseload during this critical early phase, we aim to uncover the interaction between quantified treatment data and therapeutic interactions. We hypothesize that successful and unsuccessful cases will differ in both patient characteristics and early process variables. Specifically, we anticipate differences in patient symptom severity, interpersonal functioning, and psychological insight prior to treatment. Furthermore, treatment process measures will show meaningful discrepancies between successful and unsuccessful cases in the third session, and qualitative analysis will reveal distinct patterns in therapeutic interactions.

Materials and Methods

Data sources

The present study uses archival data from a naturalistic psychotherapy study conducted at a university-based outpatient clinic (Hilsenroth, 2007; Owen & Hilsenroth, 2014). Participants were individuals aged 18-65 who sought services at the center. The original data were collected as follows: after completing the standard clinical intake and psychological assessment, but prior to beginning formal psychotherapy, patients who agreed to take part in the study signed informed consent forms, which included permission to videotape therapy sessions for clinical and research purposes. No other exclusion criteria were applied.

Throughout treatment, patients and therapists met as they would in the normal course of therapy, with all sessions being videotaped. At different points during treatment, patients completed standard program evaluation measures and returned them to the clinic staff. The therapists had no access to their own patients' data.

External raters, who were clinically trained Ph.D. graduate students, viewed the videotaped sessions and provided ratings of the diagnosis and therapeutic process. The independent rating team was composed of clinical Ph.D. graduate students. External raters were blind to the therapists' ratings and other patient data, maintaining objectivity in their assessments. Prior to the rating sessions, the raters underwent supervised training on the scales. Their training included rating video recordings from sessions not included in this study. This process ensured they were well-versed in the application of the manuals of different scales. The external raters independently reviewed the videotapes of the third session from different therapists. Following their review, they discussed ratings until they reached a consensus on each item. To ensure consistency and prevent deviations in coding, reliability meetings were held throughout the analysis process. The interrater reliabilities were in the "good" to "excellent" range and discussed in another study (Stein *et al.*, 2009,2010).

Inclusion of cases

We identified contrasting cases within individual therapists' caseloads, replicating the selection methods of Werbart *et al.* (2019b) to examine patients showing the highest and lowest levels of improvement. Only patients who initially scored in the clinical range ($T\text{-score} \geq 60$) on the BSI at the beginning of treatment were considered for inclusion in the study. The BSI is a self-report inventory that measures symptom distress across nine domains. It contains 53 items using a 5-point Likert-type scale ranging from 0 (not at all) to 4 (extremely) (BSI; Derogatis, 1982). The BSI has demonstrated strong psychometric properties, including high internal consistency reliability, with Cronbach's alpha coefficients ranging from .71 to .85 for the nine dimensions (Derogatis & Melisaratos, 1983). Test-retest reliability coefficients range from .68 to .91 over a two-week interval, indicating good temporal stability (Derogatis & Melisaratos, 1983). The BSI has also shown good convergent and discriminant validity, correlating highly with corresponding scales on the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1940) and the Symptom Checklist-90-Revised (SCL-90-R) (Boulet & Boss, 1991; Derogatis & Melisaratos, 1983). Additionally, the BSI has demonstrated sensitivity to change in clinical settings, making it a useful tool for assessing treatment outcomes (Piersma *et al.*, 1994).

The BSI-Global Severity Index (BSI-GSI) (Derogatis, 1993), which is calculated by taking the mean score across all 53 items of the BSI, is used to assess overall psychological distress. To explore the experiences of patients who showed the most and least improvement, we employed a deviant case sampling method approach used by Werbart *et al.* (2019b). Treatment success was defined based on the degree of symptom change pre-post treatment, as measured by the Reliable Change Index (RCI) of the BSI-GSI scores. The RCI is calculated by dividing the difference between pre- and post-treatment scores by the standard error of difference, with a threshold of 1.96 representing statistically reliable change at the $p < .05$ level, accounting for measurement error and regression to the mean (Jacobson & Truax, 1991).

Five inclusion criteria were applied to select cases for analy-

sis. First, we examined only the caseloads of therapists who had two completed treatment cases, where each patient began treatment in the clinical range of scores. Second, successful cases were defined as patients who showed clinically significant improvement, moving from the clinical to the functional range on the BSI-GSI, with an RCI greater than 1.96 and a final BSI-GSI raw score below 0.93, signifying a shift from a clinical level of distress to a non-clinical level (Derogatis & Melisaratos, 1983). Third, unsuccessful cases were defined as patients who did not show reliable improvement (RCI less than 1.96), with little or no reliable change, as well as deterioration (RCI of 0 or negative value) during treatment. Fourth, to empirically contrast successful and unsuccessful treatments within a therapist's caseload, we selected cases whose RCI values differed by at least 2 points (e.g., RCI=2.2 vs. RCI=0.20). Finally, we required that therapists have at least one pair of cases meeting the criteria. Following application of these inclusion criteria, three therapists were identified, each having one pair of cases that met the criteria for the successful/unsuccessful case sampling method.

Participants

The three therapists (referred to as A, B, and C) selected for this study were second-year Ph.D. students in Clinical Psychology. A and C were both female, while B was male. All patients were female and comparable in age, treatment duration, and diagnosis (Table 1). Regarding diagnoses based on DSM-IV, all participants had a primary Axis I diagnosis related to depression, including major depressive disorder (MDD), depressive disorder not otherwise specified (DepDisNOS), or dysthymic disorder. On Axis II, all participants had personality disorder features or a diagnosed personality disorder, such as borderline personality

disorder (BPD), histrionic features, avoidant features, or personality disorder not otherwise specified (PD NOS), with Cluster B or Cluster C features. Three participants (A-1, B-1, and C-1) showed clinically significant improvement with GSI-RCI scores of 2.2, 3.8, and 3.8, respectively. The other three participants (A-2, B-2, and C-2) did not show reliable improvement, with GSI-RCI scores of 0.2, -0.1, and 0, respectively.

Quantitative data review

To investigate differences between successful and unsuccessful cases, we examined the diagnostic features, personality, and symptoms of the patients at the onset of treatment. In addition, we examined psychotherapy process variables (Table 2). A comparison was conducted between successful and unsuccessful cases within each therapist's caseload.

Measures

The Shedler-Westen Assessment Procedure (SWAP-200) (Shedler & Westen, 1998; Shedler & Westen, 2004) is a procedure in which clinicians sort 200 personality descriptors into eight categories, ranging from 0 (irrelevant or inapplicable to the patient) to 7 (highly descriptive of the patient), based on their clinical judgment and understanding of the patient's characteristics. The SWAP-200 has demonstrated strong psychometric properties in terms of test-retest reliability (Shedler & Westen, 2004) and good interrater reliability (Westen & Muderrisoglu, 2003) with coefficient ranging from .70 to .82. The SWAP Insight Scale, developed by Lehmann and Hilsenroth (2011), is a subscale of the SWAP-200 consisting of six items, and the SWAP Social Avoidance/Anxiety subscale (Katz & Hilsenroth, 2017) consisted of 8 items. Raters were not aware of what items

Table 1. Demographic, diagnostic, and patients' characteristics of successful and unsuccessful case pairs.

Patients' characteristics			Patients' designations					
			A-1	A-2	B-1	B-2	C-1	C-2
Gender			Female	Female	Female	Female	Female	Female
Age			40	49	36	35	18	18
Number of sessions			18	10	24	38	43	62
Therapist gender			Female	Female	Male	Male	Female	Female
Diagnosis	Axis I	MDD	DepDisNOS	Dysthymic Dis	DepDisNOS	DepDisNOS	MDD, GAD	
	Axis II	PD NOS (Cluster B Features)	Borderline & Histrionic Features	Avoidant Features	PD NOS (Cluster C Features)	Avoidant Features	BPD	
Outcome: BSI-GSI RCI Scores			2.2	0.2	3.8	-0.1	3.8	0

Patient designations indicate therapist (A, B, or C) and outcome category (1 = successful case with clinically significant improvement, 2 = unsuccessful case with no reliable improvement or deterioration); MDD, major depressive disorder; DepDisNOS, depressive disorder not otherwise specified; Dysthymic Dis, dysthymic disorder; PD NOS, personality disorder not otherwise specified; BPD, borderline personality disorder; BSI-GSI RCI, Brief Symptom Inventory-Global Severity Index Reliable Change Index.

Table 2. Pre-treatment psychological assessment and session process measures across pairs.

	Patient	Evaluator Therapist/evaluator	External rater
Diagnosis	BSI BSI-depression	DSM	DSM
Relational functioning	BSI-interpersonal distress	SWAP-200	
Process	SEQ CASF	SEQ	CPPS

BSI, Brief Symptom Inventory; DSM, Diagnostic and Statistical Manual of Mental Disorders; SWAP, Shedler-Westen Assessment Procedure; SEQ, Session Evaluation Questionnaire; CPPS, Comparative Psychotherapy Process Scale; CASF, Combined Alliance Short Form.

were used in these subscales when they completed their ratings. Therapists completed the SWAP-200 after the psychological assessment process and the first two psychotherapy sessions, prior to the third session.

The Session Evaluation Questionnaire (SEQ) (Stiles, 1980; Stiles & Snow, 1984; Stiles *et al.*, 2004) was used to assess the treatment process. The SEQ is a self-report instrument featuring 21 pairs of bipolar adjectives, each rated on a 7-point semantic differential scale. These items are grouped into four subscales: i) Smoothness (smooth/rough, comfortable/uncomfortable, easy/difficult, pleasant/unpleasant, relaxed/tense); ii) depth (powerful/weak, valuable/worthless, deep/shallow, full/empty, special/ordinary); iii) positivity (happy/sad, confident/afraid, pleased/angry, definite/uncertain, friendly/unfriendly); and iv) arousal (aroused/quiet, fast/slow, energetic/peaceful, moving/still, excited/calm (Stiles, 1980; Stiles *et al.*, 1994). The internal consistency reliability for the four subscales ranges from .78 to .91 (Stiles *et al.*, 1994; Stiles *et al.*, 2004), and the measure has shown good construct validity (Stiles *et al.*, 1994; Stiles *et al.*, 2004). In this study, SEQ was completed by both patients and therapists at the end of the third session.

The Combined Alliance Short Form - patient version (CASF-P) (Hatcher and Barends, 1996) was derived from a factor analysis of responses from 231 outpatients at a university clinic who completed popular alliance measures including the Penn Helping Alliance Questionnaire (Luborsky *et al.*, 1983), the Working Alliance Inventory (Horvath & Greenberg, 1989), and the California Psychotherapy Alliance Scales (Gaston, 1991). The CASF-P consists of 20 items rated on a 7-point Likert-type scale. The measure has demonstrated good construct validity and internal consistency with a coefficient alpha of .93 for the Total Scale (Hatcher & Barends, 1996). The patient rated this after the third session.

The Comparative Psychotherapy Process Scale (CPPS) (Hilsenroth *et al.*, 2005) is a 20-item measure that assesses therapist activity and psychotherapy techniques during therapy sessions. Developed from an extensive review of empirical studies comparing psychotherapy processes (Blagys & Hilsenroth, 2000, 2002), the CPPS consists of two subscales: one measuring psychodynamic-interpersonal (PI) features (CPPS-PI) and one measuring cognitive-behavioral (CB) features (CPPS-CB). Items are scored using a 7-point Likert scale that ranges from 0 (not at all characteristic) to 6 (extremely characteristic). Studies have validated the CPPS's reliability and clinical utility (Hilsenroth *et al.*, 2001; Hilsenroth *et al.*, 2005). External raters who had achieved excellent levels of interrater reliability completed CPPS ratings by watching videotapes of the third sessions independently and blind to all other information. Regular meetings were held to address rater drift.

Qualitative analysis

To gain an in-depth view of the patients' diagnosis and treatment process at the early stage of treatment, the third session *verbatim* transcripts were analyzed using qualitative methods. The research team comprised two members. The first author is an Asian female Ph.D. student with two years of clinical experience in the United States and an integrative theoretical orientation. The first author received training in qualitative analysis as part of her doctoral coursework; no additional training was undertaken for this study. The second author is a White male professor serving as the first author's research mentor. The second author had supervised the original data collection process. To maintain objec-

tivity, the first author had no personal connections to any of the study participants or assessment material prior to the review of data. Throughout the research process, both authors engaged in ongoing reflexivity, discussing their positional perspectives with each other and during group meetings to manage potential biases.

The first author transcribed the third therapy sessions of both successful and unsuccessful cases conducted by therapists B and C, watching the original videotapes (videotapes of therapist A's sessions were not available). Initially, Otter.ai was used to generate the transcripts of the sessions. Transcripts were then carefully reviewed and corrected by the research team to ensure accuracy. Given the focus on early treatment, only third-session data were analyzed. Due to the limited sample, thematic saturation was not reached. The analysis followed an inductive, experiential thematic analysis (Braun & Clarke, 2006). To be specific, the analysis process adhered to Braun and Clarke's six-phase framework: familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and writing up the report. The first author independently read through the transcripts line by line, identifying and categorizing relevant content into thematic domains aligned with the study's research questions. Each transcript was then reviewed separately. After viewing, verbatim quotations were selected to illustrate the identified themes and provide a nuanced understanding of the therapeutic process data. Within-therapist and cross-case comparisons were conducted to examine similarities and differences between successful and unsuccessful cases.

Combining quantitative data and qualitative data

The findings from qualitative analysis were examined and integrated with the quantitative data. Diagnoses, relational functioning, and treatment processes were analyzed to identify factors that could explain outcomes in both successful and unsuccessful cases.

Results

Quantitative data review

A review of assessment and process data revealed discrepancies between the successful cases and the unsuccessful cases (Table 3). Prior to entering treatment, patients in successful cases reported higher levels of depression. On the BSI depression subscale, the average difference in T-scores was 8.33, indicating a large effect size (T-score=8 is the standardized unit of measurement, corresponding to Cohen's $d=.8$). Additionally, prior to treatment, successful cases demonstrated greater interpersonal distress on the BSI Interpersonal Sensitivity subscale. The average difference in BSI Interpersonal Sensitivity T-scores was 7, also indicating a moderate to large effect size. Successful cases also showed higher levels of social anxiety on the SWAP-200 Social Anxiety/Avoidance subscale. The average raw score difference on this scale is 0.89. However, successful cases also demonstrated greater personal insight on SWAP-200 Insight. The average difference in raw scores is 0.87.

Regarding psychotherapy process measures, there were incongruences between therapist and patient ratings of the third session on the SEQ smoothness scale. Patients in successful cases rated lower on the smoothness subscale (difference=-2.27). Moreover, in successful cases, therapists' rating on the smoothness subscale was higher than patients' rating (difference=1.47).

In contrast, in unsuccessful cases, therapists rated lower than patients on the smoothness subscale (difference=2.13). The CPPS item #1, “The therapist encourages the exploration of feelings regarded by the patient as uncomfortable”, revealed a discrepancy between therapists exploring uncomfortable feelings in successful cases *versus* unsuccessful cases during the third session. Specifically, external raters coded therapists as exploring more uncomfortable feelings in successful cases than in unsuccessful cases in the third session (difference=1.67). In addition, these external raters observed that in the successful cases, therapists used more psychodynamic-interpersonal techniques than in unsuccessful cases during session 3.

Qualitative analysis

The quantitative findings revealed several key patterns distinguishing successful from unsuccessful cases during the third session. Specifically, successful cases were characterized by patients rating sessions as less smooth, therapists exploring more uncomfortable feelings, and patients demonstrating higher initial insight and symptom severity.

To better understand how these quantitative differences manifest in therapeutic interactions, we conducted a semantic analysis of third session transcripts. Through semantic analysis, we found that exploring uncomfortable feelings was a consistent theme across all cases. However, in successful cases, this exploration was characterized by increased insight development, while in unsuccessful cases, it was marked by avoidance and externalization of difficulties. Furthermore, the lower smoothness ratings observed in successful cases may reflect patients’ willingness to explore uncomfortable feelings and tolerate difficult emotions.

The following clinical vignettes illustrate differences in emotional exploration and insight development or externalization across cases. Each example contrasts the reactions of two patients to similar questions from the same therapist. For each patient, there is an overview of the case background, followed by an excerpt from the third session transcript. This comparative approach highlights the differences in emotional exploration and insight development between successful and unsuccessful cases (*i.e.*, greater symptom severity, SWAP Insight Scale, CPPS-item #1).

Therapist B

Therapist B is a Caucasian male and worked with two patients: B-1 (RCI=3.8) and B-2 (RCI=-0.1). Patient B-1 was a 36-year-old Caucasian female who completed 24 sessions with Therapist B. She presented with depression and avoidant features. Her therapeutic goals included reducing depressive and anxious symptoms, improving emotional expression, enhancing relationship formation, and increasing self-esteem. She described her parents as critical and dismissive. Patient B-2 was a 35-year-old Caucasian female who completed 38 sessions with Therapist B. She presented with depression and personality disorder features with anxious and avoidant traits. Her therapeutic goals included reducing depression and anxiety, improving emotional connections with others, enhancing emotional expression, and increasing parenting confidence. She described her childhood environment as having an unsupportive, critical mother and a warmer father.

Vignettes from Therapist B’s successful and unsuccessful patients (B-1 & B-2)

Response to exploration of uncomfortable feelings

The following excerpts from the third session demonstrate the patient’s ability to explore and articulate feelings of frustration. The first excerpt is from patient B-1. Prior to this excerpt, the patient had been discussing her dissatisfaction at work.

T: Is that a part of that kind of frustration?
B-1: It is, yeah. It is something... yeah, you know, my own choosing, when I was quite young, but I started there, a few years ago... (continues talking about how she got her current job in relation to feelings of frustration).

Patient B-1 willingly engaged when the therapist attempted to explore uncomfortable feelings of frustration. She not only acknowledged these emotions but also provided detailed information about their origins. Her statement that it was “my own choosing” reveals significant insight, reflecting self-awareness of her role in her current situation rather than external circumstances.

Later in the same session, a powerful emotional moment

Table 3. Within-therapist comparisons of assessment and process variables between successful and unsuccessful cases.

Therapists			Therapist A			Therapist B			Therapist C		
Case Category			A-1	A-2	Diff	B-1	B-2	Diff	C-1	C-2	Diff
Diagnosis	BSI	IST-E	70	62	8	71	60	11	70	68	2
		DEPT-E	72	55	17	71	65	6	70	68	2
	SWAP-200	Insight	2.83	2.17	0.6	6.17	5.5	0.67	3.83	2.5	1.33
		Social anxiety/avoidance	1.63	1.38	0.25	4.25	3.63	0.62	4	2.25	1.75
Process	SEQ-3-P	Smooth	3.6	6.6	-3	4.8	6.2	-1.4	2.4	4.8	-2.4
	SEQ-3-T	Smooth	5.2	4	1.2	5	3.2	1.8	5	4	1
	CASF	3rd	4.5	6.35	-1.85	6.8	6.8	0	6.25	6.7	-0.45
	CPPS	3rd-I	4	3	1	4	2	2	6	4	2
		3rd PI Scale	32	28	4	36	33	3	39	30	9

This table displays raw score differences between successful cases (A-1, B-1, C-1) and unsuccessful cases (A-2, B-2, C-2) within each therapist’s caseload. The “Diff” column represents the score difference across all three therapist pairs. For BSI measures, T-scores are listed, and differences between T-scores are calculated in “Diff” (T-score=8 is a standardized unit of measurement, corresponding to Cohen’s d=.8). For the other measures, due to the small sample size (n=3 pairs), effect size was not calculated. BSI, Brief Symptom Inventory; IST-E, Interpersonal Sensitivity Scale of BSI T scores at evaluation; DEPT-E, Depression Scale of BSI T scores at evaluation; SWAP-200, Shedler-Westen Assessment Procedure; SEQ-3-P, Session Evaluation Questionnaire 3rd session patient ratings; SEQ-3-T = Session Evaluation Questionnaire 3rd session therapist ratings; Smooth, Smoothness subscale; CASF-3rd, Combined Alliance Short Form-Patient Version ratings from 3rd session; CPPS, Comparative Psychotherapy Process Scale; CPPS 3rd session, Comparative Psychotherapy Process Scale, independent ratings of 3rd session by external raters; CPPS Item #1, “The therapist encourages the exploration of feelings regarded by the patient as uncomfortable (*e.g.*, anger, envy, excitement, sadness, or happiness)”; CPPS PI Scale, Psychodynamic-Interpersonal Technique Scale.

emerged when discussing B-1's relationship with her father. The following excerpt demonstrates affect tolerance when the therapist explored vulnerable emotions.

T: You just want him to ask: how you are doing... maybe sometimes to ask how you were doing, maybe to ask if you need anything, anything I can do for you?

B-1: (silent for period, starting to tear up)

T: You feel sad at the moment.

B-1: (big sigh, crying) Yeah. It's just really hard.

T: It is really hard.

The therapist's articulation of B-1's unexpressed emotional needs triggered a profound reaction. She began crying, demonstrating a willingness to experience the uncomfortable feelings, allowing the difficult emotion to emerge and be expressed. Therapist B then supported and validated her expression of this difficult emotion by verbally acknowledging, "It is really hard."

The third excerpt, occurring after the previous emotional expression, illustrates continued engagement with uncomfortable feelings and developing insight.

B-1: It just was like, either he would like to talk to me more, or, not that I want to get rid of him, but he would just move on after a little bit.

T: Move on where?

B-1: Maybe move down to Florida, or something. That I don't want, he was like...he was at home, but not really there.

T: It's (his behavior) confusing.

B-1: Yeah.

T: What do you think is the reason he does that?

B-1: I don't know. It's easier said than done. I can't imagine that, why...

T: Do you think he has an idea how it makes you feel?

B-1: Uh, I mean, even if I talk to him, he is like, huh.. you know, he is like, he says, "Call me and leave a message, ah, text is OK". That's it. I don't know... he would say, "How were things back at home?", but like that's it.

T: Small talk, on the surface, always.

B-1: Yeah.

This excerpt demonstrates B-1's sustained engagement with difficult material following her emotional expression. She articulated her feelings about her father, describing him as "not really there". Despite initially responding, "I don't know" when asked about her father's perspective, she continued exploring the relationship dynamics. This sustained exploration led to a deeper understanding of their relational pattern, "on the surface, always". This progression illustrates the development of insight into her father's communication style and their relationship dynamic.

Contrasting unsuccessful case: avoidance of uncomfortable feelings

The following excerpt from B-2's third session demonstrates response patterns when therapists explore uncomfortable emotions. This excerpt followed a discussion of B-2's emotional response to perceived judgment.

T: How about now? Or when have you felt more inhibited here when we talk about things?

B-2: Um...

T: When things are not so comfortable to talk about.

B-2: I don't know. Um... I'm trying to think. Pretty much nothing really. I mean, I told you things that, probably, I wouldn't want to tell people, but I did. Um, like the fact that, you know, I, um, you know, I had a, I wouldn't say wild streak, but I did things that were just a little bit, shouldn't have done, I guess, um, you know, like, It's like when I was telling you that...um, yeah, and I guess it's hard because you know, other people have judged me in that area.

In contrast to B-1's engagement, B-2 initially denied uncomfortable feelings by saying, "pretty much nothing really". She demonstrated difficulty sustaining exploration of internal experiences. When she did acknowledge discomfort, she attributed her difficulties to external factors by saying, "I guess it's hard because you know, other people have judged me in that area". This pattern reflects limited psychological insight, as B-2 focused on external attribution rather than developing self-understanding. While B-1 demonstrated the ability for insight and emotional tolerance, B-2 exhibited avoidance and externalization, illustrating lower emotional accessibility that may influence therapeutic outcomes.

Therapist C

Therapist C is a Caucasian female who has two patients, C-1 (BSI-GSI RCI=3.8) and patient C-2 (BSI-GSI RCI=0). Patient C-1 was an 18-year-old Caucasian female who completed 43 sessions with Therapist C. She presented with depression and avoidant features, along with limited family and relationship support. Her therapeutic goals focused on addressing depressive symptoms and interpersonal difficulties. Patient C-2 was an 18-year-old Caucasian female who completed 62 sessions with Therapist C. She presented with depression, anxiety, and borderline personality features. She sought therapy for academic difficulties, anxiety management, and relationship issues.

Vignettes from Therapist C's successful and unsuccessful patients (C-1 & C-2)

Articulation of uncomfortable feelings in successful cases

The following excerpts from the third session demonstrate the patient's C-1 ability to articulate her uncomfortable feelings.

C-1: I mean. I'm working all the time. He gets to go out with his friends, but like every time when I was not working, he doesn't really care.

T: And what are your feelings?

C-1: It's like, like I'm trapped. I'm just boring. It was like I can't bring anything new to this relationship (with her partner).

In the first excerpt, patient C-1 began by describing events in her relationship with her boyfriend. When asked, she articulated her feelings, using "trapped" and "boring" to describe her emotions. Moreover, she provided insight into the underlying reasons "I can't bring anything new to this relationship"; she also appeared to recognize her own sadness, through which she also developed a new insight: "It was like I can't bring anything new to this relationship".

Topic switching and avoidance in unsuccessful cases

The following excerpt from C-2's third session illustrates avoidance of sustained exploration of uncomfortable feelings.

T: How do you feel right now? About...things about you, there is a bar (referring to C-2 use of word; i.e., a limit).

C-2: Yeah, yeah.

T: What do you think makes the bar?

C-2: Um, I guess, part of it was pessimistic, maybe. Well,...no.

T: Very hard to...

C: Yeah.

T: I have a sense... it's hard for you to tolerate the positive feelings.

C-2: Yeah. Um... I'm waiting for the one thing. Um, oh, yeah, there is a thing that I want to mention. And after the conversation. Um, as I mentioned minutes ago, after I came home the Saturday night... (started to talk about her friend).

When the therapist explored C-2's difficulty tolerating positive emotions, the patient initially acknowledged the limitation but demonstrated difficulty sustaining exploration. Notably, she deflected the conversation from her internal experience to external topics (discussing with her friend), illustrating the pattern of avoiding sustained engagement with uncomfortable emotional material observed in unsuccessful cases.

Discussion

This study makes a significant contribution to psychotherapy research by identifying specific early session indicators that predict treatment success within individual therapists' caseloads. Unlike previous research that has examined general therapeutic processes throughout treatment (Werbart, 2019b), our study is the first to systematically analyze third session dynamics as predictive markers of outcome, revealing early therapeutic patterns that influence treatment success. Furthermore, while Werbart *et al.* (2019b) demonstrated the value of analyzing contrasting outcomes within therapists' caseloads, our study uniquely combines quantitative measures with qualitative transcript analysis focused on early treatment, providing both statistical evidence and clinical understanding of how therapeutic success begins to emerge.

The current study reveals that successful cases began treatment with higher levels of depression, interpersonal distress, and social anxiety. Importantly, these patients also demonstrated greater personal insight, which aligns with research showing better outcomes for patients who are experiencing higher distress but as well as psychological resources such as insight and emotional investment capacity. Moreover, we identified a distinctive pattern in early sessions where successful cases were characterized by divergent therapist-patient perspectives on session smoothness, with patients rating sessions as less smooth than their therapists. This divergence appears to be linked to greater therapist exploration of uncomfortable feelings in successful cases, as confirmed by independent ratings using the CPPS. These results suggest that productive therapeutic tension, combined with patients' capacity for insight and both parties' willingness to explore difficult emotions, may be crucial early indicators of treatment success.

Our findings challenge the mixed evidence regarding the relationship between initial severity and treatment outcomes. While some studies have found that higher symptom predicts poorer outcomes (Firth *et al.*, 2015; Rozenal *et al.*, 2019), our results align with research suggesting more complex patterns. Like Bower *et al.* (2013), who found more severely depressed patients showed larger treatment effects, and Scholten *et al.* (2023), who found higher baseline social anxiety predicted bet-

ter CBT outcomes, we found that successful cases began treatment with higher levels of depression, interpersonal distress, and social avoidance/anxiety. Crucially, while supporting the findings of Ackerman *et al.* (2000), who found that patients with challenging relationship patterns, who retained a capacity for emotional investment/engagement, were more likely to continue in therapy, our study demonstrates that longer treatment duration alone does not ensure better outcomes. In fact, for therapists B and C, the unsuccessful cases remained in treatment longer yet showed no reliable improvement. This suggests that the combination of higher initial severity, emotional engagement, and insight may be more predictive of outcome than treatment length, highlighting the importance of considering patients' capacity for insight and affective engagement alongside clinical severity.

Our findings extend research on within-therapist variations by revealing complex relationships between early session processes and treatment outcomes. Research examining session smoothness has shown significant correlations with therapeutic alliance (Lingiardi *et al.*, 2011; Marmarosh & Kivlighan, 2012), but its direct relationship to treatment outcomes is uncertain (Stiles *et al.*, 1990). While Pesale *et al.* (2012) found that higher patient early session smoothness predicted better treatment outcomes, our results suggest that perhaps this issue may be more complex than a simple association with higher smoothness scores, but instead that greater therapist-patient discrepancy in smoothness perceptions during early sessions needs to be better understood by therapists and may even be associated with better outcomes. This divergence in the current study appears to be linked to a deeper emotional exploration, as successful cases maintained affective engagement during challenging discussions. In unsuccessful cases, similar discomfort led to avoidance patterns, despite therapists' attempts to re-engage with affective content. These findings suggest that discrepancies in smoothness perceptions are not inherently positive or negative, but rather serve as potential indicators of underlying therapeutic processes. When accompanied by sustained effort in emotional processing and the therapist's skill in managing difficult emotions, such discrepancies may be more beneficial than surface-level comfort or smoothness alone.

The current findings diverge from Werbart *et al.* (2019b) in ways that illuminate the temporal development of therapeutic processes. Key differences emerge in our focus on early session dynamics *versus* their examination of the full treatment course. While Werbart *et al.* (2019b) found that therapists in less successful cases did not take initiative to overcome the obstacles in the process, our qualitative analysis showed that the therapists continued to explore the uncomfortable feelings with care, although the patients remained emotionally closed in the early treatment. This divergence may reflect the temporal development of therapeutic processes, suggesting that rupture-repair patterns and therapist adaptability issues may emerge later in treatment rather than during early sessions.

Our study extends Werbart's work by identifying specific patient characteristics that may make early therapeutic tension productive rather than detrimental. While Werbart *et al.* (2019b) identified the importance of mutual understanding and clear treatment goals, our study uniquely highlights how early emotional exploration may set treatment trajectories before these broader therapeutic agreements are fully established. These findings demonstrate how mixed-method approaches can reveal not only which factors differentiate successful from unsuccessful cases, but also when these differences begin to emerge in the therapeutic process. This highlights the importance of examining

not just isolated session qualities, but how different process elements interact with patient variables to create conditions for therapeutic change.

Some limitations of the current study should be noted. The study was based on archival data, which prevented obtaining post-treatment interviews with therapists and patients that could have provided valuable retrospective insights like those captured in Werbart *et al.* (2019b), potentially limiting our understanding of how participants themselves perceived the therapeutic process. For therapist A's cases, the lack of available videotaped sessions limited our ability to conduct the same depth of qualitative analysis as possible with the other cases, which may create an imbalance in the richness of data across therapists. Additionally, the small sample of three therapists, while allowing for detailed within-therapist comparisons, limits the generalizability of findings.

Another limitation is that the study relied on a single outcome measure (symptom reduction) rather than examining multiple domains of therapeutic change, which may have obscured important treatment benefits not captured by symptom measures alone. Future research would benefit from larger samples of therapist-patient dyads and the inclusion of multiple outcome measures to capture different aspects of therapeutic change. Studies examining these processes across different therapeutic modalities and patient populations would also help establish the generalizability of these findings beyond psychodynamic therapy with young adults. Furthermore, research incorporating continuous session recordings throughout treatment could help illuminate how the early dynamics identified in this study evolve over time and shape therapeutic trajectories.

The findings from this study have important implications for clinical practice and training. Our results suggest that a patient's capacity to engage with emotional discomfort during initial sessions serves simultaneously as a diagnostic indicator of treatment readiness and a prognostic marker for therapeutic outcomes. The differential patterns observed between successful and unsuccessful cases indicate that resistance to emotional exploration in early therapeutic encounters may signal potential treatment challenges requiring specialized clinical approaches or enhanced therapeutic attention to affective avoidance. These findings underscore that early-session process indicators may serve as valuable predictive markers for treatment outcomes, suggesting that early-session dynamics may be crucial in shaping treatment trajectories and ultimate outcomes.

References

- Ackerman, S. J., Hilsenroth, M. J., Clemence, A. J., Weatherill, R., & Fowler, J. C. (2000). The effects of social cognition and object representation on psychotherapy continuation. *Bulletin of the Menninger Clinic*, 64(3), 386-408.
- Alfonsson, S., Fagerlös, S., Beckman, M., & Lundgren, T. (2024). Psychotherapist factors that patients perceive are associated with treatment failure. *Psychotherapy*, 61(3), 241-249. doi: 10.1037/pst0000527
- Baldwin, S. A., Murray, D. M., Shadish, W. R., Pals, S. L., Holland, J. M., Abramowitz, J. S., & Watson, J. (2011). Intraclass correlation associated with therapists: Estimates and applications in planning psychotherapy research. *Cognitive Behaviour Therapy*, 40(1), 15-33.
- 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). Wiley.
- Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology*, 75(6), 842-852. doi: 10.1037/0022-006X.75.6.842
- Blagys, M. D., & Hilsenroth, M. J. (2000). Distinctive features of short-term psychodynamic interpersonal psychotherapy: A review of the comparative psychotherapy process literature. *Clinical psychology: Science and practice*, 7(2), 167-188. doi: 10.1093/clipsy.7.2.167
- Blagys, M. D., & Hilsenroth, M. J. (2002). Distinctive activities of cognitive-behavioral therapy: A review of the comparative psychotherapy process literature. *Clinical Psychology Review*, 22(5), 671-706. doi: 10.1016/S0272-7358(01)00117-9
- Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S., Bullis, J., Gorman, J. M., Shear, M. K., & Barlow, D. H. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology*, 81, 443-454. doi: 10.1037/a0031437
- Boulet, J., & Boss, M. W. (1991). Reliability and validity of the Brief Symptom Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(3), 433-437. doi: 10.1037/1040-3590.3.3.433
- Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., Knowles, S., Cuijpers, P., Andersson, G., Christensen, H., Meyer, B., Huibers, M., Smit, F., Van Straten, A., Warmerdam, L., Barkham, M., Bilich, L., Lovell, K., & Liu, E. T.-H. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: Meta-analysis of individual patient data. *BMJ*, 346(2), f540-f540. doi: 10.1136/bmj.f540
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. doi: 10.1191/1478088706qp0630a
- Curran, J., Parry, G. D., Hardy, G. E., Darling, J., Mason, A.-M., & Chambers, E. (2019). How Does Therapy Harm? A Model of Adverse Process Using Task Analysis in the Meta-Synthesis of Service Users' Experience. *Frontiers in Psychology*, 10, 347. doi: 10.3389/fpsyg.2019.00347
- Castonguay, L. G., & Hill, C. E. (Eds.). (2017). *How and why are some therapists better than others?: Understanding therapist effects*. American Psychological Association.
- Derogatis, L. R. (1982). Brief Symptom Inventory [Database record]. Retrieved from PsycTESTS. doi: 10.1037/t00789-000
- Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine*, 13(3), 595-605. doi: 10.1017/S0033291700048017
- Derogatis, L. R. (1993). BSI: Administration, scoring and procedures manual for the Brief Symptom Inventory II (3rd ed.). Minneapolis, MN: National Computer Systems.
- Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance rupture repair: A meta-analysis. *Psychotherapy*, 55(4), 508-519. doi: 10.1037/pst0000185
- Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, 55(4), 316-340. doi: 10.1037/pst0000172
- Firth, N., Barkham, M., Kellett, S., & Saxon, D. (2015). Therapist effects and moderators of effectiveness and efficiency in psy-

- chological wellbeing practitioners: A multilevel modelling analysis. *Behaviour Research and Therapy*, 69, 54–62. doi: 10.1016/j.brat.2015.04.001
- Firth, N., Saxon, D., Stiles, W. B., & Barkham, M. (2019). Therapist and clinic effects in psychotherapy: A three-level model of outcome variability. *Journal of Consulting and Clinical Psychology*, 87(4), 345–356. doi: 10.1037/ccp0000388
- Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, 73(5), 914–923.
- Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose–response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice*, 9(3), 329–343. doi: 10.1093/clipsy.9.3.329
- Hathaway, S. R., & McKinley, J. C. (1940). A Multiphasic Personality Schedule (Minnesota) : I. Construction of the Schedule. *The Journal of Psychology*, 10(2), 249–254. doi: 10.1080/00223980.1940.9917000
- Hill, C. E., Spiegel, S. B., Hoffman, M. A., Kivlighan Jr, D. M., & Gelso, C. J. (2017). Therapist expertise in psychotherapy revisited. *The Counseling Psychologist*, 45(1), 7–53.
- Hilsenroth, M. (2007). A Programmatic Study of Short-Term Psychodynamic Psychotherapy: Assessment, Process, Outcome and Training. *Psychotherapy Research*, 17, 31–45.
- Hilsenroth, M. J., Cromer, T. D., & Ackerman, S. J. (2012). How to make practical use of therapeutic alliance research in your clinical work. In R. A. Levy, J. S. Ablon, & H. Kächele (Eds.), *Psychodynamic psychotherapy research: Evidence-based practice and practice-based evidence* (pp. 361–380). Humana Press/Springer Nature. doi: 10.1007/978-1-60761-792-1_22
- Hilsenroth, M. J., Blagys, M. D., Ackerman, S. J., Bonge, D. R., & Blais, M. A. (2005). Measuring Psychodynamic-Interpersonal and Cognitive-Behavioral Techniques: Development of the Comparative Psychotherapy Process Scale. *Psychotherapy: Theory, Research, Practice, Training*, 42(3), 340–356. doi: 10.1037/0033-3204.42.3.340
- Hilsenroth, M. J., Ackerman, S. J., & Blagys, M. D. (2001). Evaluating the Phase Model of Change During Short-Term Psychodynamic Psychotherapy. *Psychotherapy Research*, 11(1), 29–47. doi: 10.1080/713663851
- Hilsenroth, M. J., Kivlighan, D. M., & Slavin-Mulford, J. (2015). Structured Supervision of Graduate Clinicians in Psychodynamic Psychotherapy: Alliance and Technique. *Journal of Counseling Psychology*, 62(2), 173–183. doi: 10.1037/cou0000058
- Howick, J., Koletsi, D., Ioannidis, J. P. A., Madigan, C., Pandis, N., Loef, M., Walach, H., Sauer, S., Kleijnen, J., Seehra, J., Johnson, T., & Schmidt, S. (2022). Most healthcare interventions tested in Cochrane Reviews are not effective according to high quality evidence: A systematic review and meta-analysis. *Journal of Clinical Epidemiology*, 148, 160–169. doi: 10.1016/j.jclinepi.2022.04.017
- Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67(3), 300–307. doi: 10.1037/0022-006X.67.3.300
- Katz, M., & Hilsenroth, M. (2017). Is Therapist Evaluation of Social Anxiety/Avoidance Traits Associated with Patient-Reported Attachment Style? *Psychiatry Research*, 257, 226–232.
- Lambert, M. J. (2011). What Have We Learned About Treatment Failure in Empirically Supported Treatments? Some Suggestions for Practice. *Cognitive and Behavioral Practice*, 18(3), 413–420. doi: 10.1016/j.cbpra.2011.02.002
- Lehmann, M. E., & Hilsenroth, M. J. (2011). Evaluating Psychological Insight in a Clinical Sample Using the Shedler-Westen Assessment Procedure. *Journal of Nervous & Mental Disease*, 199(5), 354–359. doi: 10.1097/NMD.0b013e3182175138
- Levitt, H. M., Pomerville, A., & Surace, F. I. (2016). A qualitative meta-analysis examining clients’ experiences of psychotherapy: A new agenda. *Psychological Bulletin*, 142(8), 801–830. doi: 10.1037/bul0000057
- Li, E., Kealy, D., Aafjes-van Doorn, K., McCollum, J., Curtis, J. T., Luo, X., & Silberschatz, G. (2024). “It Felt Like I Was Being Tailored to the Treatment Rather Than the Treatment Being Tailored to Me”: Patient Experiences of Helpful and Unhelpful Psychotherapy. *Psychotherapy Research*, 1–15. doi: 10.1080/10503307.2024.2360448
- Lingiardi, V., Colli, A., Gentile, D., & Tanzilli, A. (2011). Exploration of session process: Relationship to depth and alliance. *Psychotherapy*, 48(4), 391–400. doi: 10.1037/a0025248
- Lingiardi, V., Muzi, L., Tanzilli, A., & Carone, N. (2018). Do therapists’ subjective variables impact on psychodynamic psychotherapy outcomes? A systematic literature review. *Clinical Psychology & Psychotherapy*, 25(1), 85–101. doi: 10.1002/cpp.2131
- Luborsky, L., McLellan, A. T., Diguer, L., Woody, G., & Seligman, D. A. (1997). The psychotherapist matters: Comparison of outcomes across twenty-two therapists and seven patient samples. *Clinical Psychology: Science and Practice*, 4(1), 53–65.
- Marmarosh, C. L., & Kivlighan, D. M. (2012). Relationships among client and counselor agreement about the working alliance, session evaluations, and change in client symptoms using response surface analysis. *Journal of Counseling Psychology*, 59(3), 352–367. doi: 10.1037/a0028907
- Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy*, 10(6), 361–373. doi: 10.1002/cpp.383
- Owen, J., & Hilsenroth, M. J. (2014). Treatment Adherence: The Importance of Therapist Flexibility in Relation to Therapy Outcomes. *Journal of Counseling Psychology*, 61(2), 280–288. doi: 10.1037/a0035753
- Pesale, F. P., Hilsenroth, M. J., & Owen, J. J. (2012). Patient early session experience and treatment outcome. *Psychotherapy Research*, 22(4), 417–425. doi: 10.1080/10503307.2012.662607
- Piersma, H. L., Boes, J. L., & Reaume, W. M. (1994). Unidimensionality of the Brief Symptom Inventory (BSI) in Adult and Adolescent Inpatients. *Journal of Personality Assessment*, 63(2), 338–344. doi: 10.1207/s15327752jpa6302_12
- Rozental, A., Andersson, G., & Carlbring, P. (2019). In the Absence of Effects: An Individual Patient Data Meta-Analysis of Non-response and Its Predictors in Internet-Based Cognitive Behavior Therapy. *Frontiers in Psychology*, 10, 589. doi: 10.3389/fpsyg.2019.00589
- Saxon, D., & Barkham, M. (2012). Patterns of Therapist Variability: Therapist Effects and the Contribution of Patient Severity and Risk. *Journal of Consulting and Clinical Psychology*, 80(4), 535–546. doi: 10.1037/a0028898
- Schattner, E., Tishby, O., & Wiseman, H. (2017). Relational Pat-

- terns and the Development of the Alliance: A Systematic Comparison of two Cases. *Clinical Psychology & Psychotherapy*, 24(2), 555–568. doi: 10.1002/cpp.2019
- Scholten, W., Seldenrijk, A., Hoogendoorn, A., Bosman, R., Muntingh, A., Karyotaki, E., Andersson, G., Berger, T., Carlbring, P., Furmark, T., Bouchard, S., Goldin, P., Kampmann, I., Morina, N., Kocovski, N., Leibing, E., Leichsenring, F., Stolz, T., Van Balkom, A., & Batelaan, N. (2023). Baseline Severity as a Moderator of the Waiting List–Controlled Association of Cognitive Behavioral Therapy With Symptom Change in Social Anxiety Disorder: A Systematic Review and Individual Patient Data Meta-analysis. *JAMA Psychiatry*, 80(8), 822. doi: 10.1001/jamapsychiatry.2023.1291
- Shedler, J., & Westen, D. (1998). Refining the Measurement of Axis II: A Q-Sort Procedure for Assessing Personality Pathology. *Assessment*, 5(4), 333–353. doi: 10.1177/107319119800500403
- Shedler, J., & Westen, D. (2004). Refining Personality Disorder Diagnosis: Integrating Science and Practice. *American Journal of Psychiatry*, 161(8), 1350–1365. doi: 10.1176/appi.ajp.161.8.1350
- Stein, M. B., Hilsenroth, M., Pinsker-Aspen, J. H., & Primavera, L. (2009). Validity of DSM-IV Axis V Global Assessment of Relational Functioning Scale: A Multimethod Assessment. *The Journal of Nervous and Mental Disease*, 197(1), 50. doi: 10.1097/NMD.0b013e3181923ca1
- Stein, M. B., Pesale, F. P., Slavin, J. M., & Hilsenroth, M. J. (2010). A training outline for conducting psychotherapy process ratings: An example using therapist technique. *Counseling and Psychotherapy Research*, 10(1), 50–59. doi: 10.1080/14733140903229457
- Stiles, W. B. (1980). Measurement of the impact of psychotherapy sessions. *Journal of Consulting and Clinical Psychology*, 48(2), 176–185. doi: 10.1037/0022-006X.48.2.176
- Stiles, W. B., & Snow, J. S. (1984). Counseling session impact as viewed by novice counselors and their clients. *Journal of Counseling Psychology*, 31(1), 3–12. doi: 10.1037/0022-0167.31.1.3
- Stiles, W. B., Glick, M. J., Osatuke, K., Hardy, G. E., Shapiro, D. A., Agnew-Davies, R., Rees, 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., Reynolds, S., Hardy, G. E., Rees, A., Barkham, M., & Shapiro, D. A. (1994). Evaluation and description of psychotherapy sessions by clients using the Session Evaluation Questionnaire and the Session Impacts Scale. *Journal of Counseling Psychology*, 41(2), 175–185. doi: 10.1037/0022-0167.41.2.175
- Stiles, W. B., Shapiro, D. A., & Firth-Cozens, J. A. (1990). Correlations of session evaluations with treatment outcome. *British Journal of Clinical Psychology*, 29(1), 13–21. doi: 10.1111/j.2044-8260.1990.tb00844.x
- Tryon, G. S., Birch, S. E., & Verkuilen, J. (2018). Meta-analyses of the relation of goal consensus and collaboration to psychotherapy outcome. *Psychotherapy*, 55(4), 372–383. doi: 10.1037/pst0000170
- Tryon, G. S. (1990). Session depth and smoothness in relation to the concept of engagement in counseling. *Journal of Counseling Psychology*, 37(3), 248–253. doi: 10.1037/0022-0167.37.3.248
- Vybíral, Z., Ogles, B. M., Řiháček, T., Urbancová, B., & Gocieková, V. (2024). Negative experiences in psychotherapy from clients' perspective: A qualitative meta-analysis. *Psychotherapy Research*, 34(3), 279–292. doi: 10.1080/10503307.2023.2226813
- Wampold, B. E., & Imel, Z. E. (2015). The great psychotherapy debate: The evidence for what makes psychotherapy work (2nd ed.). Routledge.
- Wampold, B. E., & Brown, G. S. (Jeb). (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, 73(5), 914–923. doi: 10.1037/0022-006X.73.5.914
- Werbart, A., Missios, P., Waldenström, F., & Lilliengren, P. (2019a). “It was hard work every session”: Therapists' view of successful psychoanalytic treatments. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 29(3), 354–371. doi: 10.1080/10503307.2017.1349353
- Werbart, A., von Below, C., Engqvist, K., & Lind, S. (2019b). “It was like having half of the patient in therapy”: Therapists of nonimproved patients looking back on their work. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 29(7), 894–907. doi: 10.1080/10503307.2018.1453621
- Westen, D., & Shedler, J. (1999). Revising and Assessing Axis II, Part II: Toward an Empirically Based and Clinically Useful Classification of Personality Disorders. *American Journal of Psychiatry*, 156(2), 273–285. doi: 10.1176/ajp.156.2.273
- Westen, D., & Muderrisoglu, S. (2003). Assessing personality disorders using a systematic clinical interview: Evaluation of an alternative to structured interviews. *Journal of Personality Disorders*, 17(4), 351–369. doi: 10.1521/pedi.17.4.351.23967
- Wienicke, F. J., Beutel, M. E., Zwerenz, R., Brähler, E., Fonagy, P., Luyten, P., Constantinou, M., Barber, J. P., McCarthy, K. S., Solomonov, N., Cooper, P. J., De Pascalis, L., Johansson, R., Andersson, G., Lemma, A., Town, J. M., Abbass, A. A., Ajilchi, B., Connolly Gibbons, M. B., Driessen, E. (2023). Efficacy and moderators of short-term psychodynamic psychotherapy for depression: A systematic review and meta-analysis of individual participant data. *Clinical Psychology Review*, 101, 102269. doi: 10.1016/j.cpr.2023.102269
- Yonatan-Leus, R., Gwertzman, G., & Tishby, O. (2024). Using machine learning methods to identify trajectories of change and predict responders and non-responders to short-term dynamic therapy. *Psychotherapy Research*, 1–17. doi: 10.1080/10503307.2024.2420725
- Zimmermann, D., Wampold, B. E., Rubel, J. A., Schwartz, B., Poster, K., Schilling, V. N. L. S., Deisenhofer, A.-K., Hehlmann, M. I., Gómez Penedo, J. M., & Lutz, W. (2021). The influence of extra-therapeutic social support on the association between therapeutic bond and treatment outcome. *Psychotherapy Research*, 31(6), 726–736. doi: 10.1080/10503307.2020.1847344
- Zuroff, D. C., Kelly, A. C., Leybman, M. J., Blatt, S. J., & Wampold, B. E. (2010). Between-therapist and within-therapist differences in the quality of the therapeutic relationship: Effects on maladjustment and self-critical perfectionism. *Journal of Clinical Psychology*, 66(7), 681–697. doi: 10.1002/jclp.20683