

Running title: Effectiveness of Modified Alliance-Focused Training with Doubling

Modified Alliance-Focused Training with Doubling as an integrative approach to improve therapists' competencies in dealing with alliance ruptures and prevent negative outcome in psychotherapy for depression. Study protocol of a randomized controlled multicenter trial

Antje Gumz 1, Denise Kästner 2, Laurence Reuter 3, Carmen Martinez Moura 4, Klarissa Ehlers 5, Anne Daubmann 6, Catherine F. Eubanks 7, J. Christopher Muran 8, Timothy Anderson 9, Ramona Stöckl 10, Georg Schwanitz 11, Lena Stegemann 12, Lana Rohr 13, Ulrike Willutzki 14, Frank Jacobi 15, Antonia Zapf 16

1 Antje Gumz, a.gumz@phb.de; Psychologische Hochschule Berlin, Germany, Department of Psychosomatics and Psychotherapy, Berlin, 10179; 00491635026408

2 Denise Kästner, d.kaestner@phb.de; Psychologische Hochschule Berlin, Germany, Department of Psychosomatics and Psychotherapy, Berlin

3 Laurence Reuter, l.reuter@phb.de; Psychologische Hochschule Berlin, Germany, Department of Psychosomatics and Psychotherapy, Berlin

4 Carmen Martinez Moura, cm.moura@tp.phb.de; Psychologische Hochschule Berlin, Germany, Department of Psychosomatics and Psychotherapy, Berlin

5 Klarissa Ehlers, k.ehlers@vt.phb.de; Psychologische Hochschule Berlin, Germany, Department of Psychosomatics and Psychotherapy, Berlin

6 Anne Daubmann, adaubman@uke.de; Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Germany

7 Catherine F. Eubanks, ceubanks@adelphi.edu; Derner School of Psychology, Adelphi

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
University, Garden City, New York, USA

8 J. Christopher Muran, jcmuran@adelphi.edu; Derner School of Psychology, Adelphi University, Garden City, New York, USA; Psychotherapy Research Program, Icahn School of Medicine at Mount Sinai

9 Timothy Anderson, andersot@ohio.edu; Department of Psychology, Ohio University, Athens, Ohio

10 Ramona Stöckl, ramona.stoeckl@charite.de, Clinical Trial Office, Charité Universitätsmedizin Berlin, Germany

11 Georg Schwanz, georg.schwanz@charite.de, Clinical Trial Office, Charité Universitätsmedizin Berlin, Germany

12 Lena Stegemann, lena.stegemann@charite.de, Clinical Trial Office, Charité Universitätsmedizin Berlin, Germany

13 Lana Rohr, lane.rohr@charite.de, Clinical Trial Office, Charité Universitätsmedizin Berlin, Germany

14 Ulrike Willutzki, ulrike.willutzki@uni-wh.de, University Witten/Herdecke, Department of Psychology and Psychotherapy, Witten, Germany

15 Frank Jacobi, f.jacobi@phb.de, Psychologische Hochschule Berlin, Germany, Department of Clinical psychology and Psychotherapy, Berlin, 10179; 004930209166220

16 Antonia Zapf, a.zapf@uke.de; Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Germany

Correspondence concerning this article should be addressed to Prof. Dr. Antje Gumz, Department of Psychosomatics and Psychotherapy, Psychologische Hochschule Berlin, Am Köllnischen Park 2, 10179 Berlin. Email: a.gumz@phb.de

Abstract

Introduction: Alliance ruptures constitute a high risk of premature treatment termination and poor psychotherapy outcome. The Alliance-Focused Training (AFT) is a promising transtheoretical approach to enhance therapists' skills in dealing with alliance ruptures.

Methods and analysis: To evaluate the effectiveness of Modified AFT with doubling (MAFT-D), a randomized, patient and evaluator-blinded, multicenter trial was designed comparing MAFT-D (delivered to trainee therapists and supervisors) and psychotherapy training/ treatment as usual (TAU) for therapists (n=120) and their patients with depressive disorders (n=240). A total of 16 cooperating centers, each offering either cognitive-behavioral or psychodynamic psychotherapy training, will contribute to recruitment. Stratification by center (both for therapists and patients) and hence therapeutic approach (cognitive-behavioral vs. psychodynamic psychotherapies), and by comorbid personality disorder (yes vs. no, for patients) will be carried out. The two hierarchically ordered primary hypotheses are: In MAFT-D compared to TAU, a stronger reduction of depressive symptoms and a lower rate of patient dropout is expected from baseline to 20 weeks after baseline. Follow-up assessments are planned at 35 weeks, 20 months, and 36 months post-baseline to evaluate the persistence of effects. Secondary patient- and therapist-related outcomes as well as predictors, moderators, and mediators of change will be investigated. Mixed models with repeated measures will be used for the primary analyses.

Ethics and dissemination: Ethical approvals were obtained by the institutional Ethics review board of the main study center as well as by review boards in each federal state where one or more cooperating centers are located (secondary votes). Following the Consolidated Standards of Reporting Trials statement for non-pharmacological trials, results will be reported in peer-reviewed scientific journals and disseminated to patient organizations and media.

Trial registration number DRKS00014842 (German Clinical Trial Register).

Strengths and limitations of this study:

- This large randomized controlled multicenter trial aims to deepen our understanding of alliance-focused trainings and holds great potential to substantially advance evidence-based psychotherapy training across therapeutic approaches.
- The study is designed to assess the effects of MAFT-D (delivered to trainee therapists and supervisors) on both therapists and, most importantly, on their patients.
- Involving multiple psychotherapy training institutes not only ensures generalizability, but may also facilitate transfer into clinical training and practice.
- By focusing on depression while stratifying for comorbid personality disorders, the study also stands to clarify potential differential effects of MAFT-D depending on patients' psychopathological characteristics.
- The supervision frequency (one session per four therapy sessions), aligned with local real-world practices, and the primary measurement point (20 weeks), chosen for comparability with other trials, together pose a risk of insufficient dose, with effects potentially becoming apparent only at later follow-ups.

Keywords: Alliance-Focused Training; depression; psychotherapy; randomized controlled trial; therapeutic alliance; alliance ruptures

Background

Depressive disorders are among the most common mental disorders (1) and the leading cause for disability worldwide (2). The course of the disorder is often recurring or chronic and the consequences for the individual and the society can be severe (1). The guideline-recommended treatments for depression are pharmacotherapy, psychotherapy or a combination of both (depending on severity and chronicity). Cognitive-behavioral and psychodynamic psychotherapies (CBT and PDT) are evidence-based and comparably effective (1, 3-5). However, unsatisfactory outcomes in depression treatments are frequent (up to 41% of patients do not reliably improve, up to 1/3 drop out (6-8)). Thus, factors closely linked to outcome and dropout need to be addressed. Here, the evidence clearly points to the importance of common factors (relevant across therapeutic approaches), especially the therapeutic alliance (9-12).

Alliance ruptures (i.e., periods of tension or breakdown of the collaborative relationship between patient and therapist) inevitably occur in the course of any psychotherapy (13). Across various therapeutic approaches, there is an evergrowing wealth of evidence establishing a link between alliance ruptures and treatment outcome, whereas the therapists' ability to repair ruptures has been demonstrated to improve outcome and to prevent dropout (14-27). However, therapists often fail to notice ruptures or lack the skills to deal with them constructively (13, 28-32).

Consequently, it is of crucial importance to improve the competence of therapists in dealing with alliance ruptures. This is in line with current recommendations according to which psychotherapy trainings should give more emphasis to common therapy principles like improving the alliance (33-35). Despite the common occurrence of alliance ruptures and the serious implications resulting from failures to resolve them, the curricula of psychotherapy training do not systematically address the topic as of yet.

Various therapist trainings focus on therapeutic relational skills and/or the process of alliance rupture and repair. Among them, the most prominent is Alliance-Focused Training

(AFT, (36, 37)), which follows a transtheoretical approach and is grounded in an evidence-based model developed by Safran, Muran, and Eubanks for successfully resolving alliance ruptures (13). Its main objective is to strengthen three skills: 1) to sensitively notice alliance ruptures, 2) to tolerate difficult affects, and 3) to meta-communicate about them in a helpful way. Video-recorded sessions and role plays are fundamental elements of AFT (37-39). Building on our pilot study (funded by the Heigl Foundation), we made several modifications to the training, the most significant of which was the incorporation of the psychodrama technique of doubling after role plays to enhance the focus on affective communication and to foster a secure, non-judgemental atmosphere in both training and supervision (40, 41) (details see section “Experimental intervention”).

Previous research emphasizes the potential of therapist training programs targeting the therapeutic alliance in general, and of AFT specifically (20, 42, 43). A meta-analysis of empirically supported relationship factors (44) describes the systematic resolution of alliance ruptures as one (out of three) of the most promising relationship elements that warrant further research. Two recent studies demonstrated an effect of AFT on therapists’ skills (42, 43). Evidence on patient outcomes appears promising regarding the reduction of dropout rates, but remains unclear in terms of symptom improvement (20, 22). Overall, the evidence on the effectiveness of trainings focusing on the alliance seems still rather limited. The studies included in the existing reviews and meta-analyses (20-22) were highly heterogeneous with respect to the investigated patient populations, treatment lengths, scope and content of the training intervention. Moreover, the sample size of most studies was small. Promising results were found with respect to patients with depressive disorders (evidence base: study 1: n=31 intervention group (IG) 1; n=34 IG 2; n=38 control group (CG); 8 sessions problem solving therapy (45); study 2: n=11 IG, n=11 CG, 16 sessions cognitive therapy (46)).

Despite the central role of relationship problems among patients with depression (47-49) and the high proportion of negative or insufficient therapeutic outcome, there is no large-scale

trial investigating the benefits of trainings with a focus on the therapeutic alliance in depression. More studies are needed to investigate the effectiveness of AFT and to link the increase in therapist competence to outcome (50).

Against this background, our study intends to investigate the benefits of our modified form of AFT (MAFT-D) vs. psychotherapy training as usual (TAU) for patients with depressive disorders. The decision to investigate patients with depressive disorders is based on the facts that a) to date there is no large-scale study investigating the effects of alliance-focused trainings in depression, b) depressive disorders are highly prevalent and debilitating, c) interpersonal problems are key factors in models on the development and maintenance of depression (48, 49), d) negative outcome in depression treatments is particularly frequent (6-8). The project investigates CBT and PDT therapists, because of the transtheoretical nature of AFT, the great significance of common factors, and the urgent need for evidence-supported methods in psychotherapy training. Training and therapy under routine conditions represents the reference for efficacy and safety for the experimental intervention, because CBT and PDT are established evidence-based treatments for depression, and psychotherapy training at the state-approved institutes can also be considered as the current gold standard.

Additionally, the project intends to investigate mediators of MAFT-D effectiveness as well as further evidence-based mediators of change, as well as predictors and moderators of therapy success.

Methods

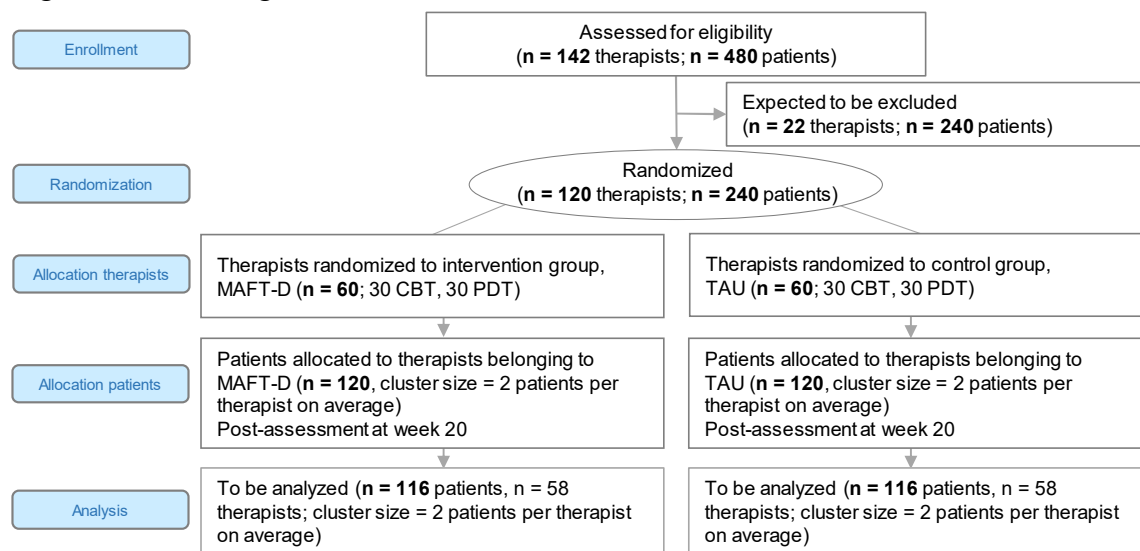
Study design

Patients, their trainee therapists and therapists' supervisors will take part in the randomized controlled multicenter trial. The trial intends to investigate the benefits of MAFT-D vs. TAU for patients with depressive disorders, stratified by the cooperating center (psychotherapy training institute), subsequently by therapeutic approach (CBT vs. PDT) and comorbid personality disorder (SCID-5-PD, yes vs. no), as well as their trainee therapists,

stratified by cooperating center (psychotherapy training institute) and therapeutic approach (CBT vs. PDT). Therapists and supervisors belonging to the intervention group will receive MAFT-D. Therapists and supervisors belonging to the the control group will receive/conduct psychotherapy training as usual (TAU). Figure 1 illustrates the study design. The trial is registered at the German Clinical Trial Register (DRKS00014842).

- Please insert Figure 1 here -

Figure 1. Trial design.



CBT: Cognitive-behavioral therapy; PDT: Psychodynamic therapy; MAFT-D: Modified Alliance-Focused Training with doubling; TAU: Psychotherapy Training as usual.

Hypotheses and research questions

The two hierarchically ordered primary hypotheses are: In MAFT-D compared to TAU, a stronger reduction of patient-rated depressive symptoms (first primary endpoint) and a lower rate of patient dropout (second primary endpoint) is expected from baseline to 20 weeks after baseline.

Secondary patient-related hypotheses are: In MAFT-D compared to TAU, stronger reduction of patient-rated depressive symptoms after 35 weeks, and 20 and 36 months, lower rate of dropout at week 35, and month 20 and 36, stronger reduction of observer-rated depressive symptoms, and patient-rated anxiety, somatic complaints, personality structure deficits, interpersonal problems, and quality of life after 20 and 35 weeks, and 20 and 36 months.

Secondary therapist-related hypotheses are: In MAFT-D compared to TAU, stronger increase in observer-rated interpersonal skills after 35 weeks, and therapist-rated therapeutic skills, trait-like relational manner in therapy, emotional suppression, alexithymic traits (difficulties identifying and describing feelings) and satisfaction with supervision after 20 and 35 weeks, and 20 and 36 months.

The hypotheses regarding mediators of MAFT-D effectiveness are: MAFT-D generates more favorable outcomes via an improved therapeutic alliance and increased therapist skills to deal with alliance ruptures. Corresponding mediator variables are: 1) therapeutic alliance (51), 2) ratio of unresolved/ resolved alliance ruptures (22), 3) convergence of patients' and therapists' alliance ratings over time (14, 52, 53), 4) interventions referring to the therapeutic relationship in the here and now (13), 5) nondirective supportive therapist's techniques (54, 55), 6) therapists' competence in session (empathy) (56), 7) adherence to MAFT-D (13, 22).

Subsidiary research questions concern predictors, moderators, and mediators of a) therapeutic macro and micro outcomes (e.g., dropout, alliance), b) the effectiveness of MAFT-D (e.g., presence of a personality disorder or therapeutic orientation as moderators of MAFT-D effectiveness), c) observer-rated interpersonal therapeutic skills and skills development, and d) the successful resolution of alliance ruptures.

Methods Against Bias

Randomization and stratification: Therapists are randomly assigned by a 1:1 ratio to receive MAFT-D vs. TAU stratified by study center and approach (CBT/PDT). Patients are randomly assigned to MAFT-D vs. TAU stratified by study center, subsequently by approach (CBT/PDT),

and comorbid personality disorder (SCID-5-PD, yes vs. no). Block randomization with variable block length is used. Patients are assigned to therapists appropriate to the random group. Randomizations will be performed centrally by Prof. Zapf's group, University Medical Center Hamburg Eppendorf. Allocation is done via the electronic case report forms (implemented in secuTrial®).

Blinding: Therapists will be aware of allocation, patients will be blinded. Interviewers, raters and persons conducting the statistical analyses will be blinded. There will be strictly separate role assignments in the study team (raters/interviewers vs. unblinded study team members). Raters/ interviewers will work physically separated from unblinded staff.

Allegiance bias: Principal investigator (PI), research group members and national or international collaborators are PDT, CBT or systemic therapists, recruiting centers are from both therapeutic approaches, statisticians have no affiliation to a therapeutic approach.

Contamination bias: All MAFT-D-supervisors and -therapists will be asked to treat the contents of the intervention highly confidential and to give a standardized response to respective questions. MAFT-D-therapists will receive a two-staged informed consent (second stage: description of MAFT-D, after randomization). Part of the intervention (workshop) will be offered to TAU-therapists at the end of follow-up. Supervisors in the intervention group will document the total number of cases discussed during each supervision session and the number in which MAFT-D was used. The supervisors are also asked retrospectively about the perceived usefulness of individual MAFT-D elements. Additionally, the number of hours therapists spent engaging with crucial elements of the intervention - such as observing role plays, participating in role plays, watching video-recorded therapy sessions of other therapists, and reviewing their own video-recorded sessions - is included in the "basic characteristics, experience, and practice" section of the assessment battery. Therapists complete this section at all fixed measurement points (BA, Post, FUP 1, 2, 3). Differences between therapists in the intervention group and the control group regarding these items will be analyzed. To estimate the

extent of contamination bias, an adherence measure score (Beth Israel Fidelity Scale (57)) will be applied to the video-recorded therapy sessions and analyzed as dependent variable using a linear mixed model.

Other biases: Data monitoring will be conducted by the Clinical Study Office Charité Berlin (CTO). In order to minimize potential bias due to video-recordings, all therapy sessions will be recorded. The number of theoretical hours and supervision sessions completed will be compared for MAFT-D- and TAU-therapists. The use of additional treatments of the patients is monitored and analyzed (Client Sociodemographic and Service Receipt Inventory (58)). Therapists' training preferences are analyzed as efficiency modifiers. To assure generalizability, we use a multi-center design. To assess selection bias, we will document ineligible persons among patients (basic demographic information, reason for non-participation). Cases of withdrawal who allow to use their data will be compared with participants who do not drop out regarding baseline differences. We will document reasons for study discontinuation. Incentives are used to reduce missing values in other outcomes (e.g., 150€ dropout post-assessment).

Inclusion and exclusion criteria for patients

Inclusion criteria: a) Agreement to start CBT or PDT outpatient therapy, weekly 50 min sessions; b) Age ≥ 18 years; c) Present DSM-5 diagnosis of depressive and related disorders (major depressive disorder - single or recurrent episode; persistent depressive disorder, other specified depressive disorder). Operationalization: Structured Clinical Interview for DSM-5 Disorders, Clinician Version, SCID-5-CV (59, 60), conducted by trained raters. The SCID-5-CV shows excellent to satisfactory interrater-reliability, via phone and face-to-face (61, 62); d) Reaching or exceeding the cut-off value indicating moderate depression severity. Operationalization: Beck Depression Inventory II, BDI-II (63, 64), score ≥ 20 (1, 65); e) Signed informed consent form

Rationale for the criteria: Depression is the most common mental disorder in Germany, associated with considerable costs (66). Psychotherapy is recommended for patients with

moderate or severe depression (1). Interpersonal problems are key factors in models on the development and maintenance of depression (47-49, 67).

Exclusion criteria: a) Current or past diagnosis of a psychotic or bipolar disorder; b) Substance dependency disorder (current or during the last 12 months); SCID-5-CV at timepoint of screening; c) Acute suicidal plans at timepoint of screening or suicide attempts within the last 6 months; modified (last six months) Columbia Suicide Severity Rating Scale (CSSR-S) (68), which is a reliable and valid instrument for identification of suicide risk (68, 69); d) Psychopharmacotherapy other than antidepressants; change of antidepressant regimen during the previous month or during study participation; modified Client Sociodemographic and Service Receipt Inventory (CSSRI) (58); which is one of the most widely used instruments for a standardized assessment of health care utilization (including medications; medication classification will be based on ATC-codes) (58); e) Planned concurrent psychological treatments (Self-help groups can be continued.); CSSRI; f) Insufficient command of German language, below level B2.

Rationale for the criteria: Severe comorbidities are excluded for reasons of patients' safety. Concomitant treatments are limited to minimize performance bias. The number of exclusion criteria is kept low to ensure representativeness for routine care.

Inclusion criteria for therapists and supervisors

Therapists: a) graduated psychologist or licensed physician; b) in advanced CBT or PDT training (with a license to practice outpatient therapy under supervision, which is usually obtained after 2 years of training); c) capacity to start treatment with two new patients, d) capacity to visit one of the workshops offered for study therapists.

Supervisors: a) approved by state and cooperating institute, b) no scheduling conflicts (e.g., workshop dates, regular dates for group supervision), c) agreement to decline requests for supervision from study therapists of the other condition during the study period.

There are no exclusion criteria for therapists and supervisors.

Cooperating study centers

The main study center is at Psychologische Hochschule Berlin where study coordination and data collection take place. The cooperating centers across Germany are state-approved, well established psychotherapy training institutes offering courses in individual outpatient psychotherapy (CBT or PDT). They support the recruitment of participants and are responsible for organization and carrying out of the study-independent regular therapy training and patient treatments. The following centers agreed to cooperate (alphabetically ordered, person responsible in brackets):

1. Akademie für Psychotherapie Erfurt (Prof. M. Geyer)
2. Berliner Akademie für Psychotherapie (Prof. F. Jacobi)
3. Centrum für Integrative Psychologie Bamberg (Dr. J. Siegl)
4. DGVT-Ausbildungszentrum Berlin (Dr. M. Rotter)
5. Dresdner Institut für Psychodynamische Psychotherapie (Dr. C. Schilling, Dr. S. Seifert)
6. ppt-Institut für Psychologische Psychotherapie und Beratung Berlin (Dr. L. Hauten, Dr. R. Spielberg, E. Stahl, S. Ulm)
7. IFT – Psychotherapeutische Ambulanz, München (S. Gmeinwieser)
8. Institut für Psychotherapie und Angewandte Psychoanalyse, Jena (Dr. U. Wutzler)
9. Leipziger Ausbildungsinstitut für Psychologische Psychotherapie und Universität Leipzig (Prof. C. Exner, Dr. S. Koranyi)
10. Regionalinstitut Sachsen der DGVT (K. Sturmeit)
11. Sächsisches Institut für Psychoanalyse und Psychotherapie (Prof. K. v. Klitzing, Dr. O. Krauß, A. Kazzler)
12. Universität Greifswald (Prof. E.-L. Brakemeier, M. Tewes)
13. Universität Witten/Herdecke und DGVT-Ausbildungszentrum Dortmund (Prof. U. Willutzki)

14. Weiterbildungsstudiengang Psychodynamische Psychotherapie, Klinik und Poliklinik für Psychosomatische Medizin und Psychotherapie, Universitätsmedizin der JGU Mainz (Prof. M. Beutel)
15. Saarländisches Institut für Tiefenpsychologisch fundierte Psychotherapie, Saarbrücken (Dr. E. Hahn, W. Bauer-Neustädter)
16. Zentrum für Ausbildung in Psychologischer Psychotherapie, Friedrich-Alexander-Universität Erlangen-Nürnberg (Prof. M. Berking, Dr. K. Zierhut)

Other centers will be involved in the event of recruitment difficulties.

Recruitment

Recruitment will take place at the cooperating study centers (see above). Additional centers will be involved in the event of recruitment difficulties. Recruitment started in July 2024 and is ongoing.

The standard group size for group supervision is four participants which applies to the study and also to routine therapy training. Consequently, units of eight therapists must usually be recruited per institute. Some institutes allow smaller group sizes. However, a group consisting of only two participants on a long-term basis (more than 4 months) is considered a protocol violation. Combining study therapists from different training institutes at one location in a supervision group is permitted. The groups can also be filled with non-study participants. Changing the format from group to individual supervision is generally not permitted within the context of the study. Recruitment methods for therapists include project promotion through e-mail, advertisement in lectures/ seminars. Patients are recruited from the regular outpatient pool. Recruitment period is 21 months for patients and 16 months for therapists. Patients, therapists and supervisors will receive incentives for study related expenses (assessments, video recordings, travel).

Patient treatments

Basic conditions of treatments will be equal between MAFT-D and TAU. Patients will receive weekly 50-minute sessions (regular frequency and length). Post-assessment will take place at week 20, first follow-up at week 35, 2nd and 3rd follow-up after 20 and 36 months. Treatments can be short- or long-term depending on patient's and therapist's agreement and the sessions granted by payers. All sessions will be video recorded. Therapists will treat two study patients on average. The costs of treatments are fully covered by statutory health insurances in Germany.

Psychotherapy training

TAU- and MAFT-D-therapists will continue to participate in their regular training courses. Group supervisions of treatments will be carried out by experienced supervisors (usual frequency, one supervision session per four therapy sessions, in person, additional individual supervision is permitted in accordance with the regulations of the specific institute in the event of crises or for questions regarding the report for the health insurance company) either with MAFT-D specific focus or according to the usual procedure. The supervision groups are fixed and in order to participate the therapist must have already started at least one study therapy. Group supervisions will occur monthly or bi-weekly with patients being discussed for varying amounts of time (10 to 50 minutes per case). Supervisors will receive regular supervision fees. The hours of the workshop for the MAFT-D-group will replace other non-mandatory courses of the regular training.

Experimental intervention

Therapists randomized to MAFT-D will receive a two-day workshop (approx. 15 teaching units, 11.5 hours) followed by group supervision carried out by supervisors who will also be trained in MAFT-D. The initial workshops will take place in person (for therapists and supervisors separately). Central elements are a theoretical introduction (150-180 min.) and role plays based on individual case examples (approx. 45 min. per participants). The number of lecturers depends on the overall group size (one lecturer and max. eight therapists per role play

subgroup). In the case of unexpected absences (e.g., illness) and for therapists who are unable to attend at the originally planned workshop dates, catch-up workshops will be offered. Catch-up workshops consist of smaller groups and are shorter (approx. 10 teaching units, 7.5 hours), but are more intense due to more frequent participation in role plays.

AFT is a transtheoretical training concept initially developed by Muran, Safran, and Eubanks (36) focusing on alliance ruptures and aiming at enhancing therapists' interpersonal skills to a) sensitively recognize alliance ruptures, b) tolerate negative affects, and c) meta-communicate using video-recorded sessions, role plays, and mindfulness exercises (13, 36, 40, 41). We have translated the AFT into German and tested and modified it in pilot studies (37, 40, 41). Our modifications concern (31, 70, 71): a) strong standardisation of the supervision process, b) verbalization of the affective experience exclusively via "doubling" (psychodrama technique) immediately after the role plays, c) structured focussing on therapist's own contribution (vulnerabilities, biographical relationship experiences; based on pre work (72, 73)), d) active role of supervisors in practicing meta-communication ("prompting"). The modification were made, because in our experience, there is a tendency in supervision groups to rationalize, intellectualize and cognitively exchange concepts and strategies, which can also conceal difficulties in opening up affectively (resistance to the perception of subjective or objective countertransference (72, 74)) or "hiding" of alexithymic tendencies (i.e. difficulties in perceiving and expressing emotions; (75)). The modifications aim to enhance the focus on the perception and verbalization of emotions, minimize intellectualizing after role plays, and foster a more non-judgmental atmosphere within the supervision group. Supervision follows the following standardized procedure (duration approx. 45 minutes per patient): A) Mindfulness exercise (to increase awareness and openness to one's own emotional experience), 3-4 minutes; B) Introduction of the patient and characteristics of the therapeutic relationship with this patient (B1: verbally, B2: video sequence from last therapy session, B3: "screenplay" of an alliance rupture), 8-10 minutes; C) First role-play and "doubling" (C1: Supervisee in therapist role,

supervision group member in patient role, C2: “Doubling”, psychodrama technique, to deepen the affective insight and enable access to blind spots), 7-9 minutes; D) Second role-play and “doubling” (D1: Supervisee in patient role, further supervision group member in therapist role, D2: “Doubling”, D3: Supervisor asks group members who were in the patient role how they felt), 10-13 minutes; E) Third role-play and own contributions (E1: Supervisee in therapist role, tries out new behavior, practises meta-communication, uses suggestions from the supervisor – prompting, E2: Supervisor asks what the rupture may have to do with supervisee's vulnerabilities, whether own biographical issues may have been ‘triggered’, which feelings may be difficult to bear due to personal experiences), 6-8 minutes.

Supervisors belonging to MAFT-D will receive a two-day workshop before start of study therapies (same contents as for therapists but separately), followed by a monthly two-hour online training to refine MAFT-D-specific supervision skills and discuss questions.

To ensure consistency across centers, MAFT-D-therapists of different centers will be trained together. The same applies to supervisors. There will be a fixed group of lecturers (experienced psychotherapists and experienced in conducting MAFT-D-specific workshops and supervision).

Sample size calculation, number of study participants

Depressive symptoms: There are no prior studies that closely align with our study in terms of study population, intervention, and measurement instruments. In our view, the chosen reference (Bambling et al., 2006; (45)) represents the best available compromise, as it examines a population of depressed patients, employs an alliance-focused supervision approach, and uses the BDI-II as a measurement instrument. Accordingly, we assume a higher reduction in BDI-II after 20 weeks in MAFT-D compared to TAU of 4.0 BDI-II units with a standard deviation of 7.3 BDI-II units. To detect such an effect with a type I error of .05 (two-sided hypothesis), a power of .90, an intra-cluster correlation coefficient of .05, and a cluster size of 2 patients per

therapist on average, we need 76 therapists and 152 patients in total (38 therapists and 76 patients per group, PASS 15, tests for two means in a cluster-randomized design).

Treatment dropout: We expect that MAFT-D will reduce the treatment dropout rate after 20 weeks from 30% to 12.5%, corresponding to an odds ratio of 3.0. The expected dropout rate of 30% for TAU is a conservative estimate based on a recent meta-analysis (35% dropout) (7) and numbers in our own clinic (n= 3.357 depressed patients; 33% dropout). Previous studies on different alliance-focused training approaches reported a difference in treatment dropout rates of 17.5% or greater (Bambling et al.: IG1=6.1%, IG2=3%, CG=35%; Constantino et al.: IG=0%, CG=27.3%; Muran et al.: IG=20%, CG1=37%, CG2=46%) (45, 46, 76). With identical design effects, 58 therapists and 116 patients per group (116 therapists, 232 patients in total) are needed to detect this difference (PASS 15, test for two proportions in a cluster-randomized design). We expect that 3% of the randomized sample are not part of the full analysis set (oriented on numbers from the trial of (77)) due to the lack of any data post-randomization (78). Thus, based on the larger sample size required for dropout, a sample of 120 therapists and 240 patients in total has to be enrolled.

Based on previous psychotherapy trials (77, 79), we expect a loss to follow-up of 15% (of therapists with their patients) at post-assessment in BDI-II measurement. With the planned number of sample size and the assumed loss to follow-up, we can prove the assumed effect. We assume a non-compliance rate for the (per protocol) BDI-II analysis (e.g., failure to complete 12 sessions within 20 weeks, change in antidepressant medication, use of additional treatments) of 35%.

We expect that 20% of patients and therapists decline study participations and 15% of therapist and 50% of patients do not meet inclusion criteria. Consequently, 178 therapists and 600 patients need to be initially invited in order to assess eligibility in 142 therapists and 480 patients, in order for 120 therapists and 240 patients to participate in the study and finally analyze 116 therapists and 232 patients.

Endpoints, variables, instruments

Primary endpoints: Change in patient-rated depressive symptoms will be measured using the total score of Beck Depression Inventory (BDI-II, change=difference from baseline to 20 weeks after baseline) (65, 80) which is one of the most commonly used self-rating scales to assess depressive symptoms and is particularly recommended for monitoring the severity of depressive symptoms over time in clinical populations (1). Items are rated from 0 ‘not present’ to 3 ‘severe’. The total score is the sum of all 21 items (0-63). Good psychometric properties (reliability, validity, sensitivity to change) were demonstrated for the German version (65).

Premature therapy dropout is a dichotomous variable determined by patient interview. It is defined as the patient's decision to end therapy contrary to the initial agreement with the therapist, excluding premature terminations unrelated to therapy quality, such as those due to serious illness of the patient or therapist, or moving to a different city. In case patients cannot be reached, interviews will be conducted with the therapist, and the patient is classified as dropout.

Patient-related secondary endpoints:

- Change in patient-rated depressive symptoms (BDI-II total score) and therapy dropout at the remaining measuring times.

Relating to difference from baseline to 20 and 35 weeks, 20 and 36 months after baseline:

- Observer-rated depressive symptoms (Hamilton Rating Scale for Depression, GRID-HAM-D (81, 82), total score of the 21-item version). Excellent interrater-reliability (ICC, intra-class correlation = .95), Cronbach's α : .78.
- Patient-rated anxiety (Generalized Anxiety Disorder Scale-7, GAD-7 (83), total score, rated on a 4-point Likert scale, 0 ‘not at all’ to 3 ‘nearly every day’). Cronbach's α : .92, good test-retest reliability (ICC = .83) (83).

- Patient-rated somatic symptoms (Patient Health Questionnaire-15, PHQ-15 (84), total score, ratings of problems in the last four weeks, 0 = ‘not bothered at all’ to 2 ‘very bothered’). Cronbach's α : .80, validity was proven (85).
- Patient-rated personality structure deficits (Operationalized Psychodynamic Diagnosis-Structure Questionnaire, short version, OPD-SQS (86), total score, 12-items assessing personality functioning, rated on a 4-point Likert scale, 0 ‘does not apply at all’ to 4 ‘does apply completely’). Cronbach's α : .89 (86).
- Patient-rated interpersonal problems (Inventory of Interpersonal Problems-Revised, IIP-32 (87), total score, 32 items scored on a 5-point Likert scale). Adequate internal consistency was demonstrated. Cronbach's α of the subscales: .68 to .90 (88).
- Patient-rated quality of life (World Health Organization Quality of Life, short version, WHOQOL-BREF (89), 5-point Likert scale, 1 ‘not at all’ to 5 ‘completely/to an extreme amount’). We will use two dimensions: physical and psychological quality of life. Chronbach's α of the subscales: .57 to .88.

Therapist-related secondary endpoints:

- Difference in observer-rated interpersonal skills from baseline to week 35 after baseline of the first patient (Facilitative Interpersonal Skills Performance Test, FIS (90-94), total score, eight items measuring general interpersonal skills based on verbal responses to six video clips displaying challenging therapy situations - verbal fluency; hope; persuasiveness; emotional expression; warmth, acceptance and understanding; empathy; alliance bond capacity and rupture-repair responsiveness), 1 ‘low level’ to 5 ‘high level’ of skills). The German FIS version shows good interrater-reliability (total score ICC = .81). Cronbach's α : .95 to .96 (91, 93).
- Therapist-rated retrospective satisfaction with supervision at week 20, 35, months 20, 36 after baseline of the first patient (modified version of the Client Satisfaction Questionnaire-

8 (95), total score, measuring satisfaction with supervision instead of treatment, CSQ-8-mod (40). Cronbach's α : .87 to .93, good concurrent validity for the total score (96).

Relating to difference from baseline to 20 and 35 weeks, 20 and 36 months after baseline (all therapist rated):

- Trait-like relational manner and therapeutic competence score (relational manner - trait section (35 items) and skills change index section (10 items) of the Adapted Trainee Current Practice Report which is originally used in the SPRISTAD trial and currently undergoing psychometric evaluation (97)).
- Emotional suppression (Emotional Regulation Questionnaire subscale score, ERQ (98), 4 items, 1 'totally disagree' to 7 'totally agree'). Cronbach's α suppression subscale: .74.
- Alexithymia (Toronto Alexithymia Scale, TAS-26 (99), total score and the difficulties a) identifying and b) describing feelings subscales, 5-point likert scale, 1 'does not apply at all or not true at all' to 5 'completely true'). Cronbach's α : .80 to .86 (99).

Variables to assess potential mediators for M-AFT-D effectiveness:

- Average level of therapeutic alliance across sessions (German short revised version of the Working Alliance Inventory, WAI-SR (100-102), assessed after each session, 12 items). Cronbach's α for the total score: $>.90$. Convergent validity is established (101).
- Convergence of patient's and therapist's alliance ratings over time (patient and therapist versions of the WAI-SR, total scores of session ratings; see above).
- Ratio of unresolved/ resolved alliance ruptures (Post-Session-Questionnaire – modified version, PSQ-mod (103, 104). The first item asks whether a "problem or tension" occurred in the relationship between patient and therapist during the session. If answered with "yes", the intensity of the rupture and the degree to which it could be solved is specified. Despite the simple nature of the items, there are clear indications of their validity (103, 104).
- Therapists' usage of interventions referring to the therapeutic relationship in the here and now as well as usage of nondirective supportive techniques (mean of the categories

“repeating, paraphrasing, summarizing“, „expression of emotional sympathy“, „validation“); Psychodynamic Intervention List, PIL, observer-ratings based on transcribed therapy sessions on the level of verbal statements (105), 0 ‘category characteristics do not apply at all’ to 5 ‘category characteristics completely apply’. Satisfactory reliability and convergent validity; interrater-reliabilities ICC “expression of emotional sympathy”: .69, ICC “validation”: .68, ICC “repeating, paraphrasing, summarizing”: .80, ICC “referring to the therapeutic relationship: .78, ICC topic “therapist”: .87 (55, 105).

- Therapists’ empathy (Empathy Scale (106)), total score, 10-items, observer-ratings of video-taped sessions, 0 ‘no agreement at all’ to 3 ‘very strong agreement’. Cronbach’s α : .86 (107).
- Therapists’ basic communication skills (Clinical Communication Skills Scale – short version, CCSS-S; (108), total score, 14 items, observer-ratings of video-taped sessions, 4-point scale ‘not at all appropriately’ to ‘entirely appropriately’. Reliability and validity was demonstrated (108).
- Therapists’ adherence to AFT, CBT, PDT (Beth Israel Fidelity Scale, BIFS, (57), observer-ratings of video-taped sessions, 12-items each (three subscales). Good psychometric properties, criterion validity and adequate inter rater agreement have been demonstrated repeatedly (109).

Variables for subsidiary research questions:

- Mediators of change (patient-rated, Mediators of change in Psychotherapy Inventory, MoCPI (110)), after each session, 22 items and five subscales: quality of the therapeutic relationship (7 items), patients’ perception and expression of feelings and thoughts (5 items), patients’ becoming conscious of biographical experiences (4 items), gaining a new view on problems (4 items), patients’ self-disclosure (2 items). A five-factor structure with good psychometric properties could be found (110).
- Parenting behavior of therapist's parents (German translation of the Measure of Parental Style Questionnaire, Fragebogen dysfunktionaler elterlicher Beziehungsstile, FDEB (111, 112),

parenting style of father and mother, 15 items each, 1 'does not apply at all' to 4 'does completely apply'). Cronbach's α : .48 to .93 (112).

- Therapists' personality traits (Big Five Inventory, short version, BFI-10 (113), measuring extraversion, agreeableness, conscientiousness, emotional stability, openness, 10 items). Cronbach's α : .49 to .62 (114).
- Therapists' interpersonal problems (IIP-32, see patient outcomes).
- Therapists' sensitivity to aversive interpersonal behaviour (Interpersonal Sensitivities Circumplex Questionnaire, ISC-G (115), 40 items assessing the extent to which individuals are bothered by others' interpersonal behaviors, 1 'not at all, never' to 8 'extremely, always'). Cronbach's α : .74 to .89.
- Therapists' abilities to mentalize (Certainty About Mental States Questionnaire, CAMSQ (116), 20 items, 0 'never' to 1 'always', measuring the perceived capacity to understand mental states of the self and others. McDonald's Ω : .88 to .91, good validity and test-retest reliability (116).
- Therapist's dispositional fear of negative evaluation (Fear of Negative Evaluation Scale, FNE (117), German version - Skala Angst vor negativer Bewertung, SANB (118), 5 items, 1 'almost never applies' to 4 'almost always true'). Cronbach's α : .84 to .94.
- Therapists: Relational manner – state (97), 35 items, see therapist-related endpoints.
- Therapists' perception of own competence and skills (Facilitative interpersonal skills, FIS, self rating; see therapist-related endpoints).
- Patients' therapy expectations and therapy evaluation, PATHEV (119), 11 items, 1 'not true at all' to 5 'completely true', three subscales: improvement, fear of change, suitability. Chronbach's α : .73 to .89.
- Patients' early depressive symptom changes (pre to 5th session, BDI-II (65, 80), see above).

- Distinctive internal reactions (Impact message Inventory, IMI-C (120-122), observer-ratings, 64 items, 1 'not true at all' to 4 'absolutely true', dimensions: affiliation and dominance. Chronbach's α : .84 to .89, interrater-reliability, ICC: .30 to .58.
- Planned for following funding period: Rupture Resolution Rating System, 3RS (123), observer-ratings, presence of a rupture and its clarity and intensity, 5-point Likert-scale. Interrater-reliability withdrawal rupture ICC: .85, confrontation ruptures ICC: .98 (124).
- Selected sessions will be analysed in terms of nonverbal behavior and linguistic variables.

Measurement time points

The data collection is summarized in **Table 1**.

Tab. 1: Data and measurement time points.

Infor- mant	Instrument	No. of items	Measurement points
Patient	BDI-II	21	Screening, BA, after the 5 th therapy session, Post, FUP 1, 2, 3
	SCID-5-SPQ	106	Screening
	GAD-7	7	BA, Post, FUP 1, 2, 3
	PHQ-15	15	BA, Post, FUP 1, 2, 3
	OPD-SQS	12	BA, Post, FUP 1, 2, 3
	IIP-32	64	BA, Post, FUP 1, 2, 3
	WHOQOL-BREF	16	BA, Post, FUP 1, 2, 3
	PATHEV	11	BA
	Basic characteristics	app. 40	BA
	WAI-SR	12	After every session
	PSQ-Mod	3	After every session
	MOCPI	22	After every session

Infor- mant	Instrument	No. of items	Measurement points
Observer interviews and ratings	Symptom severity VAS	1	After every session
	additional treatments	3	After every session
	INEP	21	Post, FUP 1, 2, 3
	SCID-5-CV	Adaptive	Screening
	SCID-5-PD	Adaptive	Screening or BA (depending on interview duration)
	adapted CSSRI	Adaptive	Screening, Post, FUP 1, 2, 3
	CSSR-S	23	Screening
	GRID-HAM-D-21	21	BA, Post, FUP 1, 2, 3
	Dropout-Interview	3	Flexibly
	BIFS-AFT	12	one session each of session 1-4/5-10, 11-16, 17-20 (if available)
	PIL	37	Sessions selected based on process data
	Empathy Scale	10	Sessions selected based on process data
	CCSS-S	14	Sessions selected based on process data
	IMI-R	64	Sessions selected based on process data
	3RS	22	Sessions selected based on process data
	FIS	6*9	BA, FUP1

Infor- mant	Instrument	No. of items	Measurement points
Therapist	Basic characteristics, experience, practice	40	BA, Post, FUP 1, 2, 3
	Skills Change Index	10	BA, Post, FUP 1, 2, 3
	Relational Manner	35	BA, Post, FUP 1, 2, 3
	ERQ	10	BA, Post, FUP 1, 2, 3
	TAS-26	26	BA, Post, FUP 1, 2, 3
	FDEB	30	BA
	BFI-10	10	BA
	ISC-G	40	BA
	IIP-32	32	BA
	CAMSQ	20	BA
	SANB	5	BA
	FIS self-rating	9	BA, FUP 1
	CSQ-8-mod	8	Post, FUP 1, 2, 3
	WAI-SR	12	After every session
	PSQ-Mod	3	After every session
	Report on important events	Adaptive	flexibly
	Relational manner – state	35	After the 5 th session
Supervisor	Basic characteristics, experience, practice	app. 10	BA
	Attendance sheet	4-5	After every supervision

Note: Screening = Baseline screening; BA = Baseline Assessment; FUP = Follow-up assessment; Post = Post-assessment. Instruments: 3RS: Rupture Resolution Rating System (124); BDI-II: Beck Depression Inventory–II (65, 80); BIFS: Beth Israel Fidelity Scale (57); BFI-10: Big Five Inventory – 10 (113); CAMSQ: Certainty about Mental States Questionnaire (116); CCSS-S: Clinical Communication Skills Scale – short (108); CSSRI: Client Sociodemographic and Service Receipt Inventory (58); CSSR-S: Colombia Suicide Severity Rating Scale (68); CSQ-8: Client Satisfaction Questionnaire-8, modified to supervision satisfaction (95); ERQ: Emotion Regulation Questionnaire (98); Empathy Scale (106); FDEB: Fragebogen dysfunktionaler elterlicher Beziehungsstile (111); FIS: Facilitative Interpersonal Skills (90); GAD-7: Generalized Anxiety Disorder Scale-7 (83); GRID-HAM-D: GRID Hamilton Rating Scale for Depression (81, 82); IIP-32: Inventory of Interpersonal Problems-32 (87); IMI-R: Impact Message Inventory (125); INEP: Inventory for the assessment of Negative Effects of Psychotherapy (126); ISC-G: Interpersonal Sensitivities Circumplex – German (127); MOCPI: Mediators of Change in Psychotherapy Inventory (110); OPD-SQS: OPD-Structure Questionnaire, short version (86); PATHEV: Patientenfragebogen zur Therapieerwartung und Therapieevaluation (119); PHQ-15: Patient Health Questionnaire-15 (84); PIL: Psychodynamic Interventions List (105); PSQ-Mod: Post-Session-Questionnaire, modified version (103, 104); Relational Manner – state and trait, and the Skills Change Index are both part of the Trainee Current Practice Report originally used in the SPRISTAD trial and currently under psychometric evaluation (128); SANB: Skala Angst vor negativer Bewertung (129); SCID-CV/PD: Structured Clinical Interview for DSM-5 Disorders, Clinician Version/ Personality Disorders (59, 60); VAS: Visual analogue scale (130); WAI-SR: Working Alliance Inventory, short revised (101); WHOQOL–BREF: World Health Organization Quality of Life, short version (89).

The number of items in the report on important events depends on which event (e.g., study dropout, therapy dropout, hospitalization) occurred. Important events should be reported as soon as possible and the reporting is independent from scheduled sessions.

Statistical analyses

Results are reported according to the CONSORT statement (131). Analyses are carried out with standard software: SPSS, Version 29 or newer (Armonk, NY: IBM Corp.), SAS, Version 9.4 or newer (Cary, NC: SAS Institute Inc.), or R (Version 4.3.3). A statistical analysis plan that contains the details of all planned analyses will be finalized before database lock.

Descriptive data are presented overall and separated by randomized group. As descriptive statistics, mean, standard deviation, median, first and third quartile or minimum and

maximum are calculated for continuous variables. For categorical variables, absolute and relative frequencies are reported.

First primary outcome (hypothesis I): We use a linear mixed model with repeated measures with changes from baseline as dependent variable, intervention group, therapeutic approach (CBT vs. PDT), time and comorbid SCID-5 personality disorder (yes vs. no) as fixed effects, baseline BDI-II score as covariate, time by intervention group interaction, and cooperation center, supervision group (if possible), therapist (= cluster) and patient as nested random effects. The interaction term is eliminated from the model if the corresponding interaction p-value is >0.15 . The contrast of the intervention group after 20 weeks is the result of the primary analysis. This result is considered in a confirmatory manner according to the approach described below.

Second primary outcome (hypothesis II): We perform a mixed logistic regression with the binary variable patient dropout after 20 weeks (yes vs. no) as dependent variable, intervention group, therapeutic approach (CBT vs. PDT), and comorbid SCID-5 personality disorder (yes vs. no) as fixed effects, and cooperation center, supervision group (if possible), and therapist as nested random effects.

The two primary hypotheses are hierarchically ordered and tested one after the other at the two-sided 5% level. Hypothesis II is only tested if hypothesis I could be rejected. If hypothesis I cannot be rejected, hypothesis II is no longer examined confirmatory but exploratory. This procedure ensures a family wise type I error of 5%.

These analyses will be conducted in the full analysis set which is as complete as possible and as close as possible to the intention-to-treat ideal of including all randomized participants [120]. Missing values are treated by direct maximum-likelihood imputation to allow an intention to treat analysis, which results in unbiased estimations under the missing-at-random assumption. Missing values for treatment dropout can only be due to both patient and therapist leaving the study.

Sensitivity analysis for primary endpoints: To investigate the robustness of the results of the primary analyses we will conduct sensitivity analyses with multiple imputation by chained equation as imputation strategy. Additionally, primary analyses are repeated in the per protocol (PP) sample. Definition PP sample: patients who receive therapy by trained or non-trained and accordingly supervised therapists as intended by their group allocation without major protocol deviations (i. e., at least 12 sessions within 20 weeks for BDI-II outcome analysis; no changes with respect to psychopharmacotherapy, additional treatments as allowed for BDI-II outcome and treatment dropout) and who complete the assessments within the defined time frames.

Secondary outcomes: Secondary outcomes are examined exploratory with analogous methods appropriate for the scale level (i.e. mixed linear, logistic, ordinal, or poisson/negative binomial regression). The contrasts of the intervention vs. control group at week 5, 10, 15, 35 of BDI-II changes from baseline of the model of the first primary outcome are used as results of secondary outcome analysis.

Mediators: In path models, we test the hypotheses that patient's outcome is mediated by the average level of therapeutic alliance across sessions (WAI patient-rated); ratio of unresolved/resolved alliance ruptures (PSQ-Mod patient-rated); convergence of patient's and therapist's alliance ratings over time (WAI patient-/therapist-rated); therapists' usage of interventions referring to the therapeutic relationship in the here and now; therapists' usage of nondirective supportive techniques (validation, expressing emotional sympathy, repeating paraphrasing, summarizing; PIL); therapists' competence in session (empathy); therapists' adherence to AFT techniques (BIFS-Subscale AFT).

Additional analyses: a) To test whether MAFT-D is equally effective in CBT and PDT, we examine the interaction between intervention group and therapeutic approach. Prior to this analysis, we examine baseline differences (depression severity, chronicity, comorbidity) between therapeutic approaches (CBT/PDT). Variables for which we detect clinically relevant group differences will be added as covariates. Cooperation center will be included as random

effect. b) We examine the adherence measure BIFS (adherence with AFT-techniques, observer-rated) as an outcome, with the same approach as in primary analyses (extent of contamination bias). c) Safety endpoints are determined using frequency tables and if possible (mixed) logistic or linear regressions to compare the event frequencies/groups. d) Subsidiary research question concern predictors, moderators, and/ or mediators (patient-, therapist-, patient-therapist match- and process-related variables) of a) therapeutic macro and micro outcomes (e.g., dropout, alliance), b) the effectiveness of MAFT-D (e.g., presence of a personality or therapeutic orientation as moderators of MAFT-D effectiveness), c) observer-rated interpersonal therapeutic skills and skills development and d) the successful resolution of alliance ruptures.

Interim analyses are not planned. For secondary outcomes and analyses, missing values are not imputed and no adjustment for multiple testing is conducted.

Ethics and dissemination

The study is being conducted in compliance with the protocol, Good Clinical Practice, the declaration of Helsinki and the CONSORT statement (131). All study patients, therapists and supervisors need to provide written informed consent to the main study team prior to inclusion. Study participation is completely voluntary and can be cancelled at any time without negative consequences. Patients who terminate study participation can continue their treatment. The study was approved by the Ethics Committee of the Psychologische Hochschule Berlin (EK2024/11). Additionally, we obtained secondary votes in each federal state where one or more study centers are located. Any major modifications to the study protocol will be covered by amendments and communicated to the relevant parties (e.g., Data Safety and Monitoring Board, DSMB, see below). All personal data collected as part of the study, following the consent of the study participant, are subject to confidentiality and the provisions of the General Data Protection Regulation. All questionnaire, interview, and rating data are collected pseudonymously within the secuTrial eCRF and REDCap surveys. This means that the participant's name and all direct identifiers are replaced with a pseudonym making re-identification outside of the study no

longer possible. Access to the code key list that enables the association between study data and personal information of the study participant is limited to study staff explicitly authorized by the principal investigator. The code key list is stored securely and separately from all other study data (e.g., questionnaire data). The individual code key will be permanently erased following the completion of data collection and after the obligatory archiving time frame.

There is subject insurance for study patients. They receive the insurance confirmation and insurance conditions together with the study information. Assessments might be experienced as time-consuming. We do not expect other risks or adverse events due to study participation in patients or therapists. Positive effects resulting from treatments in both groups can be expected. There is no placebo or waiting list group. Both evidence-based treatments will be conducted according to state-of-the-art CBT/PDT and will be supervised by certified experienced psychotherapists. Frequency and dose of supervision corresponds to usual standards.

Results will be published in peer-reviewed scientific journals following the ICMJE guidelines and disseminated to the general public, patient organisations and media.

Public Involvement: Implications and recommendations for psychotherapy studies and further psychotherapy training are to be derived from the project results. These proposals will be discussed with relevant stakeholders (e.g., the clinical directors of the training institutes).

Quality Assurance and Safety

During the trial, quality control and quality assurance will be ensured through central monitoring based on the risk assessment of the study according to ICH Good Clinical Practice (GCP). The Clinical Trial Office (CTO) will perform coordination, implementation, and conduction of monitoring at the main study center following its applicable Standard Operating Procedures (SOPs). Monitoring will be performed according to SOPs and study specific requirements in accordance with the patient recruitment. The monitoring strategy will be defined in a monitoring plan with special attention to critical data and processes. The verification will focus on the key study data, e.g. signed informed consent, adherence to inclusion and exclusion

criteria and participant safety or adverse events. All the other data are checked on the basis of a representative sample, e.g. documentation on primary objectives. Unclear and incomplete data will trigger increasing in-depth monitoring of the respective data.

Data management creates a data validation plan, containing additional logic checks for plausibility and consistency not available within the electronic case report forms (eCRF). Query analysis will be performed in SAS on an up-to-date export from the CRF. Found inconsistencies, implausible or missing data will be flagged by a data management query within the CRF. A possibly false information can be corrected through directly changing values within the eCRF or must be flagged as accepted failures or false positives. Changes within the documented information will be tracked via the audit trail, which is a core function of the eCRF.

After the last participant had its last visit, all queries and all centers were closed, the database will be locked. Electronic documentation includes all exported files (SPSS, SAS, CSV, EXCEL), SAS scripts, data protocol and the closed database.

An independent DSMB will be regularly provided with all safety aspects of the trial and will review the safety data. Members of the DSMB are: Prof. Dr. Christiane Steinert, International Psychoanalytic University, Berlin, licensed psychotherapist; Prof. Dr. Elmar Brähler, Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Mainz, experienced researcher and DSMB member in former psychotherapy trials, and Dr. Theresa Keller, Charité - Universitätsmedizin Berlin, independent statistician. At annual intervals, a meeting of the board will be scheduled to review whether the recruitment plan is on target, to ensure compliance with ethical principles, adherence to protocol, to check data quality and accuracy, and to advise whether to continue, modify, or stop the trial. The DSMB will serve as link to the funding organisation and provide them with information and advice. In the event of a respective advice from the DSMB, the steering committee will finally decide whether to stop the trial (members: Prof. A. Gumz, PI; Prof. A. Zapf, statistician, Prof. S. Singer - Institute for Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the

Johannes Gutenberg University Mainz, independent member, and Prof. L. White - Clinical Child and Adolescent Psychology and Psychotherapy, Bremen University, independent member).

Important events (e.g., therapy dropouts, hospitalization) can be reported by the therapist at any time. The study team will be alerted about incoming reports via email. Supervisors will be available at all study centers to decide about crisis interventions (e.g., need for referrals). We will assess side effects/ risks with questionnaires and interviews.

Discussion

This study protocol presents a comprehensive and innovative project designed to examine the impact of MAFT-D compared to psychotherapy training and treatment-as-usual (TAU) for trainee therapists and their patients with depressive disorders.

Strengths of the study

The proposed project focuses on one of the most prevalent disorders, is dedicated to a topic highly relevant to everyday practice, and uses a randomized, controlled multi-center design including about 16 training institutes across Germany. Thus, the results have great potential to substantially contribute to an evidence-based psychotherapy training and improved outcomes for patients with depression. By involving numerous institutes, the study not only ensures generalizability within a naturalistic setting, but it could also directly impact clinical training, especially if positive results lead to the integration of MAFT-D into psychotherapeutic training curricula. The intervention itself has a transtheoretical nature and could be incorporated into various existing training and continued education programs at comparatively low cost (132).

Notably, AFT has so far been more widely studied in the context of personality disorders (42, 133), with limited evidence available for patients with depression. This study focuses on patients with depression, but includes those with comorbid personality disorders, stratifying for this variable. Thus, the study also stands to clarify potential differential effects of MAFT-D depending on patients' psychopathological characteristics, contributing to a more nuanced understanding of alliance-focused approaches across various disorders.

Moreover, the study provides valuable insights into the active ingredients of psychotherapy for depression. The chosen design allows for the consideration and differentiation of patient and therapist effects, as well as various common factors measured from different perspectives (patients, therapist, observer). The knowledge generated may contribute to a better understanding of change mechanisms in psychotherapy overall, which constitutes a critical step toward more effective treatments.

Methodological considerations and potential challenges

We aim for a rigorous methodology and different study design components are implemented to effectively minimize bias: Randomization, blinding of raters, prevention of contamination, control of therapy allegiance and of overlapping treatments as well as for confounding factors are described and warranted by corresponding measures. Additionally, the sample size is sufficiently large to enable generalizable findings across broader populations with depressive disorders.

The study's hierarchical testing framework first assesses symptom reduction, followed by dropout rates. While this sequence reasonably suggests that dropout rate effects might be less relevant without significant symptom reduction, we would like to point out that one could also argue that both hypotheses hold independent value and could merit separate evaluation, given their relevance to treatment quality.

The central aim of this study is to examine how therapist training affects therapy outcomes for patients. This means that the primary outcomes are measured not in those who receive the intervention but in the patients they treat. Although this approach might seem self-evident in the context of therapist trainings, few studies actually adopt it. Most evaluations of therapist trainings focus instead on therapist outcomes, such as improvements in therapeutic skills (e.g. (43)). While one could argue that improved therapeutic skills should lead to better therapy outcomes, this remains an assumption rather than a proven fact. Therefore, this study

was designed to assess both changes in therapeutic skills (as perceived by observers and therapists themselves) and the more essential patient outcomes.

Limitations

A key methodological aspect to consider is the supervision frequency, set at one session for every four therapy sessions. While this represents the common standard and legal requirement for trainee therapists in our country, this frequency might deviate from other clinical trials or the clinical practice in others regions. The supervision frequency may influence outcomes but also reflects local real-world practice conditions, adding ecological validity to the study.

For the purpose of comparability with other studies, the primary outcome was set to be measured at 20 weeks. However, this introduces a risk that the intervention dose may be insufficient by this time point, and effects might only become apparent at later follow-ups.

Another limiting factor is that, due to the large number of cooperating psychotherapy training institutes and the relatively small number of participating therapists per institute (typically eight), it remains unclear whether we will be able to disentangle the effects of the center from those of the supervisor or supervision group. If feasible, we plan to include supervision group as a random effect.

Additionally, it might have been valuable to record not only therapy sessions but also supervision sessions. This could have provided deeper insights into adherence to MAFT-D and rupture-repair processes occurring during supervision. Unfortunately, this was not feasible within the scope of the current study.

Finally, selective attrition poses a risk: if MAFT-D indeed reduces dropout rates, this might lead to differences in attrition between the intervention and control groups. This is particularly relevant to the analysis in the intention-to-treat population using the mixed linear model, which leads under the “Missing-at-Random” (MAR) assumption—that missing data occur at random rather than systematically (134). Various mechanisms are implemented to

prevent study dropout in patients who drop out of therapy (e.g., incentives). However, this aspect will require careful consideration and control.

Conclusion

In summary, this protocol presents an innovative and carefully designed study to test the effectiveness of MAFT-D for trainee therapists and their patients with depression. The study holds the potential to significantly enhance the quality of psychotherapy training and to contribute to a deeper understanding of mechanisms of change in psychotherapy. With positive findings, the study could have immediate implications for psychotherapeutic education and clinical practice. Despite some limitations and methodological challenges, the study's naturalistic approach underscores its relevance and ecological validity, positioning it to make a meaningful contribution to improving treatment outcomes and reducing dropout rates in depression therapy.

Declarations

Conflict of Interest Statement

The authors declare that they have no competing interests.

Funding

This work is funded by the German Research Foundation (DFG); project number: 504346851; GU 1564/6-1; Principal Investigator: Antje Gumz). The pilot study for the project was funded by the Heigl-Foundation (grant number: 02122015; Principal investigator: Antje Gumz). The funding bodies were not involved in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication.

Acknowledgments

We would like to thank all current and former members of our research group who contributed to our research described in this article. In particular, we would like to thank Jelka Berger, Merle Longley, Dr. Christopher Marx, Dr. Thomas Munder, Dr. Kai Rugenstein, Carina Schlipfenbacher, Kyra Toussaint, Lena Walther, and Miriam Warmuth. We also thank our partners at the cooperating training institutes with whom we have worked together or who have agreed to cooperate in the future.

Statement of Ethics

The study was reviewed and approved by the Ethics Committee of the Psychologische Hochschule Berlin (AZ: EK2024/11). Secondary votes were obtained for each federal state where a study center is located. Written informed consent was obtained from all study participants (patients, therapists, supervisors).

Data availability

On request, anonymized raw data will be made available to scientific colleagues after finishing the trial. During the publication process, relevant trial data will be uploaded to the Dryad Digital Repository.

Author Contributions

AG formulated the research question and conception of the study. AG, DK, LR, AZ and AD designed the study. CE and CM developed the AFT, provided their original materials and supported the research team with information and their experience. TA contributed to the implementation of measurement of therapeutic competence (Facilitative interpersonal skills) as an outcome und predictor variable by providing his original materials and supporting the research team with information and his experience. RS, GS, LS and LRo supported the study team with their experience in data collection and management and in the implementation of the clinical study. AG tested the described MAFT-D procedures in clinical context. UW provided a questionnaire. AG, DK, LR, CMM, KE, UW, and FJ contributed to the recruitment of cooperating training institutes. AG drafted the manuscript. DK, AZ, and AD provided input to

the first draft. All authors revised it critically for important intellectual content. All authors gave their final approval of the version to be published. All authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Schneider F, Härter M, Schorr S. S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression: Springer-Verlag; 2017.
2. World Health Organization. Depression and other common mental disorders: Global health estimates. Geneva; 2017.
3. National Institute for Health and Care Excellence. Depression in adults: recognition and management (clinical guideline). 2009.
4. Driessen E, Van HL, Don FJ, Peen J, Kool S, Westra D, et al. The efficacy of cognitive-behavioral therapy and psychodynamic therapy in the outpatient treatment of major depression: a randomized clinical trial. *Am J Psychiatry*. 2013;170(9):1041-50.
5. Driessen E, Van HL, Peen J, Don FJ, Twisk JWR, Cuijpers P, et al. Cognitive-behavioral versus psychodynamic therapy for major depression: Secondary outcomes of a randomized clinical trial. *J Consult Clin Psychol*. 2017;85(7):653-63.
6. Cuijpers P, Karyotaki E, Ciharova M, Miguel C, Noma H, Furukawa TA. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica*. 2021;144(3):288-99.
7. Fernandez E, Salem D, Swift JK, Ramtahal N. Meta-analysis of dropout from cognitive behavioral therapy: Magnitude, timing, and moderators. *J Consult Clin Psychol*. 2015;83(6):1108-22.
8. Swift JK, Greenberg RP, Tompkins KA, Parkin SR. Treatment refusal and premature termination in psychotherapy, pharmacotherapy, and their combination: A meta-analysis of head-to-head comparisons. *Psychotherapy (Chic)*. 2017;54(1):47-57.
9. Fluckiger C, Del Re AC, Wampold BE, Symonds D, Horvath AO. How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. *J Couns Psychol*. 2012;59(1):10-7.
10. Sharf J, Primavera LH, Diener MJ. Dropout and therapeutic alliance: a meta-analysis of adult individual psychotherapy. *Psychotherapy (Chic)*. 2010;47(4):637-45.
11. Wampold BE, Imel ZE. *The Great Psychotherapy Debate*. New York: Routledge; 2015.
12. Zilcha-Mano S, Muran JC. Thinking Trans theoretically About Alliance and Rupture: Implications for Practice and Training. *Clin Psychol Eur*. 2024;6(Spec Issue):e12439.
13. Safran JD, Muran JC. *Negotiating the therapeutic alliance: A relational treatment guide*. New York, NY, US: Guilford Press; 2000.
14. Gumz A, Bauer K, Brähler E. Corresponding instability of patient and therapist process ratings in psychodynamic psychotherapies. *Psychother Res*. 2012;22(1):26-39.
15. Gumz A, Brähler E, Geyer M, Erices R. Crisis-repair sequences - considerations on the classification and assessment of breaches in the therapeutic relationship. *BMC Med Res Methodol*. 2012;12:10.
16. Gumz A, Kästner D, Geyer M, Brähler E. [Abrupt transient deteriorations in the experienced interaction and their relationship to the experience of other important persons--operationalisation of transference aspects in therapeutic processes]. *Z Psychosom Med Psychother*. 2010;56(4):373-84.
17. Stiles WB, Glick MJ, Osatuke Kea, Hardy GE, Shapiro DA, Agnew-Davies R, et al. Patterns of alliance development and the rupture-repair hypothesis: are productive relationships U-shaped or V-shaped? *J Couns Psychol*. 2004;51:81-91.
18. Strauss JL, Hayes AM, Johnson SL, Newman CF, Brown GK, Barber JP, et al. Early alliance, alliance ruptures, and symptom change in a nonrandomized trial of cognitive therapy for avoidant and obsessive-compulsive personality disorders. *Journal of Consulting and Clinical Psychology*. 2006;74(2):337-45.
19. Larsson MH, Falkenström F, Andersson G, Holmqvist R. Alliance ruptures and repairs in psychotherapy in primary care. *Psychother Res*. 2018;28(1):123-36.
20. Safran JD, Muran JC, Eubanks-Carter C. Repairing alliance ruptures. *Psychotherapy (Chic)*. 2011;48(1):80-7.
21. Muran JC. Confessions of a New York rupture researcher: An insider's guide and critique. *Psychotherapy Research*. 2019;29(1):1-14.
22. Eubanks CF, Muran JC, Safran JD. Alliance rupture repair: A meta-analysis. *Psychotherapy (Chic)*. 2018;55(4):508-19.
23. O'Keeffe S, Martin P, Midgley N. When adolescents stop psychological therapy: Rupture-repair in the therapeutic alliance and association with therapy ending. *Psychotherapy (Chicago, Ill)*. 2020.

24. Humer E, Schramm E, Klein JP, Härter M, Hautzinger M, Pieh C, et al. Effects of alliance ruptures and repairs on outcomes. *Psychother Res.* 2021;31(8):977-87.
25. Babl A, Berger T, Gómez Penedo JM, Grosse Holtforth M, Caspar F, Eubanks CF. Disaggregating between- and within-patient effects of ruptures and resolutions on the therapeutic alliance and symptom severity. *Psychotherapy (Chic).* 2022.
26. Murphy ST, Garcia RA, Cheavens JS, Strunk DR. The therapeutic alliance and dropout in cognitive behavioral therapy of depression. *Psychother Res.* 2022:1-8.
27. Gumz A, Kästner D, Geyer M, Wutzler U, Villmann T, Brähler E. Instability and discontinuous change in the experience of therapeutic interaction: an extended single-case study of psychodynamic therapy processes. *Psychother Res.* 2010;20(4):398-412.
28. Gumz A. Kritische Momente im Therapieprozess – Chance oder Sackgasse? [Critical moments in the course of therapy - Chance or an impasse?]. *Psychotherapeut.* 2012;27:256-62.
29. Owen J, Wampold BE, Kopta M, Rousmaniere T, Miller SD. As good as it gets? Therapy outcomes of trainees over time. *J Couns Psychol.* 2016;63(1):12-9.
30. Kline KV, Hill CE, Morris T, O'Connor S, Sappington R, Vernay C, et al. Ruptures in psychotherapy: Experiences of therapist trainees. *Psychother Res.* 2019;29(8):1086-98.
31. Gumz A. Umgang mit Spannungen und Krisen in der Therapiebeziehung [Dealing with ruptures in the therapeutic relationship. Concepts and techniques]. *Psychotherapeut.* 2020;65(2):119-32.
32. Cirasola A, Midgley N, Muran JC, Eubanks CF, Hunter EB, Fonagy P. Repairing alliance ruptures in psychodynamic psychotherapy with young people: The development of a rational–empirical model to support youth therapists. *Psychotherapy.* 2024;61(1):68-81.
33. Beck JG, Castonguay, L. G., Chronis-Tuscano, A., Klonsky, E. D., McGinn, L. K., Youngstrom, E. A. Principles for training in evidence-based psychology: Recommendations for the graduate curricula in clinical psychology. *Clinical Psychology: Science and Practice.* 2014;21:410–24.
34. Laska KM, Gurman AS, Wampold BE. Expanding the lens of evidence-based practice in psychotherapy: a common factors perspective. *Psychotherapy (Chic).* 2014;51(4):467-81.
35. Castonguay LG, Eubanks CF, Goldfried MR, Muran JC, Lutz W. Research on psychotherapy integration: building on the past, looking to the future. *Psychother Res.* 2015;25(3):365-82.
36. Eubanks-Carter C, Muran JC, Safran JD. Alliance-focused training. *Psychotherapy.* 2015;52(2):169-73.
37. Gumz A, Rugenstein KM, T. Allianz-Fokussiertes Training (AFT). *Schulenübergreifender Weg zum Umgang mit Krisen in der therapeutischen Beziehung [Alliance-focused training. Integrative approach to repairing ruptures in the therapeutic alliance]. Psychotherapeut.* 2018;63(1):55-61.
38. Gumz A. *Kompetent mit Spannungen und Krisen in der therapeutischen Beziehung umgehen: Techniken und didaktische Konzepte: Vandenhoeck & Ruprecht; 2020.*
39. Eubanks CF, Muran JC, Samstag LW. Alliance-focused training: Teaching therapists to navigate alliance ruptures. *Becoming better psychotherapists: Advancing training and supervision.* Washington, DC, US: American Psychological Association; 2023. p. 107-26.
40. Gumz A, Reuter L, Flückiger C, Marx C, Rugenstein K, Schlipfenbacher C, et al. Umgang mit Spannungen und Krisen in der therapeutischen Beziehung: Erste Erfahrungen mit einem handlungsorientierten Ausbildungs- und Supervisionskonzept [Maintaining a Strong Therapeutic Alliance: Initial Experiences with the Alliance-Focused Training (AFT) in Germany]. *Psychother Psychosom Med Psychol.* 2020;70(03/04):122-9.
41. Reuter L, Walther L, Gumz A. Wie erlebten Ausbildungskandidaten das allianzfokussierte Training im Rahmen der Psychotherapieausbildung? – Eine qualitative Studie. *Psychother Psychosom Med Psychol.* 2021;71(05):185-91.
42. Muran JC, Safran JD, Eubanks CF, Gorman BS. The effect of alliance-focused training on a cognitive-behavioral therapy for personality disorders. *J Consult Clin Psychol.* 2018;86(4):384-97.
43. Perlman MR, Anderson T, Foley VK, Mimnaugh S, Safran JD. The impact of alliance-focused and facilitative interpersonal relationship training on therapist skills: An RCT of brief training. *Psychother Res.* 2020;30(7):871-84.
44. Norcross JC, Wampold BE. Evidence-based therapy relationships: research conclusions and clinical practices. *Psychotherapy (Chic).* 2011;48(1):98-102.
45. Bambling M, King, R., Raue, P., Schweitzer, R., & Lambert, W. Clinical supervision: Its influence on client-rated working alliance and client symptom reduction in the brief treatment of major depression. *Psychotherapy Research.* 2006;16(03):317-31.

46. Constantino MJ, Marnell ME, Haile AJ, Kanther-Sista SN, Wolman K, Zappert L, et al. Integrative cognitive therapy for depression: A randomized pilot comparison. *Psychotherapy (Chic)*. 2008;45(2):122-34.
47. Gómez Penedo JM, Schwartz B, Deisenhofer AK, Rubel J, Babl AM, Lutz W. Interpersonal clarification effects in Cognitive-Behavioral Therapy for depression and how they are moderated by the therapeutic alliance. *J Affect Disord*. 2021;279:662-70.
48. McCullough JP, Schramm, E., Penberthy, J.K. . CBASP as a Distinctive Treatment for Persistent Depressive Disorder. New York: Routledge; 2015.
49. Teo AR, Choi H, Valenstein M. Social relationships and depression: ten-year follow-up from a nationally representative study. *PLoS One*. 2013;8(4):e62396.
50. Hill CE, Knox, S. Training and supervision in psychotherapy. In: Lambert MJ, editor. *Bergin and Garfield's handbook of psychotherapy and behavior change*. 6. Hoboken, NJ: John Wiley & Sons; 2013. p. 775–811.
51. Horvath A, Del Re AC, Flückiger C, Symonds D. The alliance in adult psychotherapy [Special issue]. *Psychotherapy*. 2011;48:9-16.
52. Tschuschke V, Koemeda-Lutz M, von Wyl A, Cramer A, Schulthess P. The Impact of Patients' and Therapists' Views of the Therapeutic Alliance on Treatment Outcome in Psychotherapy. *The Journal of nervous and mental disease*. 2020;208(1):56-64.
53. Laws HB, Constantino MJ, Sayer AG, Klein DN, Kocsis JH, Manber R, et al. Convergence in patient-therapist therapeutic alliance ratings and its relation to outcome in chronic depression treatment. *Psychother Res*. 2017;27(4):410-24.
54. Kadur J, Lüdemann J, Andreas S. Effects of the therapist's statements on the patient's outcome and the therapeutic alliance: A systematic review. *Clin Psychol Psychother*. 2020;27(2):168-78.
55. Gumz A, Daubmann A, Erices R, Berger J, Reuter L, Kastner D. Associations between therapists' verbal techniques and patient-rated therapeutic alliance, insight, and problem solving. *Psychother Res*. 2024:1-12.
56. Heinonen E, Nissen-Lie HA. The professional and personal characteristics of effective psychotherapists: a systematic review. *Psychotherapy research*. 2020;30(4):417-32.
57. Patton J, Muran, J.C., Safran, J.D., Wachtel, P.L., & Winston, A. . Beth Israel Fidelity Scale (BIFS)1998.
58. Roick C, Kilian R, Matschinger H, Bernert S, Mory C, Angermeyer MC. [German adaptation of the client sociodemographic and service receipt inventory - an instrument for the cost of mental health care]. *Psychiatr Prax*. 2001;28 Suppl 2:84-90.
59. Beesdo-Baum K, Zaudig M, Wittchen HU. SCID-5-PD: Strukturiertes Klinisches Interview für DSM-5 – Persönlichkeitsstörungen. 1. Aufl. ed. Göttingen: Hogrefe; 2019.
60. Beesdo-Baum K, Zaudig M, Wittchen HU. SCID-5-CV: Strukturiertes Klinisches Interview für DSM-5 – Klinische Version. 1. Aufl. ed. Göttingen: Hogrefe; 2019.
61. Osório FL, Loureiro SR, Hallak JEC, Machado-de-Sousa JP, Ushirohira JM, Baes CVW, et al. Clinical validity and intrarater and test-retest reliability of the Structured Clinical Interview for DSM-5 - Clinician Version (SCID-5-CV). *Psychiatry Clin Neurosci*. 2019;73(12):754-60.
62. Somma A, Borroni S, Maffei C, Besson E, Garbini A, Granozio S, et al. Inter-rater reliability of the Italian translation of the structured clinical interview for DSM-5 personality disorders (SCID-5-PD): a study on consecutively admitted clinical adult participants. *J Psychopathol*. 2017;23:105-11.
63. Seggar Leslie B LM, J, Hansen Nathan, B. Assessing clinical significance: Application to the beck depression inventory. *Behavior Therapy*. 2002;33(2):253-69.
64. Schindler AC, Hiller W, Witthoft M. Benchmarking of cognitive-behavioral therapy for depression in efficacy and effectiveness studies--how do exclusion criteria affect treatment outcome? *Psychother Res*. 2011;21(6):644-57.
65. Kühner C, Bürger C, Keller F, Hautzinger M. Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II) [Reliability and validity of the Revised Beck Depression Inventory (BDI-II). Results from German samples]. *Der Nervenarzt*. 2007;78(6):651-6.
66. Krauth C, Stahmeyer JT, Petersen JJ, Freytag A, Gerlach FM, Gensichen J. Resource utilisation and costs of depressive patients in Germany: results from the primary care monitoring for depressive patients trial. *Depress Res Treat*. 2014;2014:730891.
67. Weightman MJ, Air TM, Baune BT. A review of the role of social cognition in major depressive disorder. *Front Psychiatry*. 2014;5:179.

68. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia–Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American journal of psychiatry*. 2011;168(12):1266-77.
69. Mundt JC, Greist JH, Jefferson JW, Federico M, Mann JJ, Posner K. Prediction of suicidal behavior in clinical research by lifetime suicidal ideation and behavior ascertained by the electronic Columbia-Suicide Severity Rating Scale. *The Journal of clinical psychiatry*. 2013;74(9):0-.
70. Gumz A. Trainieren und Messen therapeutischer Beziehungskompetenzen 2023; 22(3):[204-16 pp.]. Available from: <https://elibrary.klett-cotta.de/article/10.21706/pdp-22-3-204>.
71. Gumz A. Messen und Trainieren therapeutischer Kompetenzen. *Psychother Psychosom Med Psychol*. 2024(EFirst).
72. Berger JA, Kästner D, Gumz A. Therapeutische Eigenanteile in der Gegenübertragung. *Forum der Psychoanalyse*. 2024;40(2):233-52.
73. Gumz A. Messen und Trainieren therapeutischer Kompetenzen. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*. 2024.
74. Winnicott DW. Hate in the countertransference. *The International Journal of Psycho-Analysis*. 1949;30:69-75.
75. Gumz A, Longley M, Franken F, Janning B, Hosoya G, Derwahl L, et al. Who are the skilled therapists? Associations between personal characteristics and interpersonal skills of future psychotherapists. *Psychother Res*. 2024;34(6):817-27.
76. Muran JC, Safran JD, Samstag LW, Winston A. Evaluating an alliance-focused treatment for personality disorders. *Psychotherapy: Theory, Research, Practice, Training*. 2005;42(4):532-45.
77. Schramm E, Kriston L, Zobel I, Bailer J, Wambach K, Backenstrass M, et al. Effect of Disorder-Specific vs Nonspecific Psychotherapy for Chronic Depression: A Randomized Clinical Trial. *JAMA Psychiatry*. 2017;74(3):233-42.
78. European Medicines Agency. ICH E9 statistical principles for clinical trials. 1998.
79. Fonagy P, Rost F, Carlyle J-a, McPherson S, Thomas R, Pasco Fearon RM, et al. Pragmatic randomized controlled trial of long-term psychoanalytic psychotherapy for treatment-resistant depression: the Tavistock Adult Depression Study (TADS). *World Psychiatry*. 2015;14(3):312-21.
80. Hautzinger M, Keller F, Kühner C, Beck AT, Steer RA, Brown GK. Beck Depressions-Inventar : BDI-II ; Revision ; Manual. Frankfurt am Main: Pearson; 2009.
81. Schmitt A, Kulzer B, Hermanns N. German version of the GRID Hamilton Rating Scale for Depression (GRID-HAMD) 2015.
82. Whisman MA, Strosahl K, Fruzzetti AE, Schmalzing KB, Jacobson NS, Miller DM. A structured interview version of the Hamilton Rating Scale for Depression: Reliability and validity. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*. 1989;1(3):238.
83. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166(10):1092-7.
84. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom Med*. 2002;64(2):258-66.
85. Gräfe K, Zipfel S, Herzog W, Löwe B. Screening psychischer Störungen mit dem “Gesundheitsfragebogen für Patienten (PHQ-D)“. *Diagnostica*. 2004;50(4):171-81.
86. Ehrenthal JC, Dinger U, Schauenburg H, Horsch L, Dahlbender RW, Gierk B. [Development of a 12-item version of the OPD-Structure Questionnaire (OPD-SQS)]. *Z Psychosom Med Psychother*. 2015;61(3):262-74.
87. Horowitz LM, Strauß B, Kordy H. Inventar zur Erfassung interpersonaler Probleme. Deutsche Version. Manual (2. überarbeitete und neu normierte Auflage). Göttingen: Beltz Test; 2000.
88. Barkham M, Hardy GE, Startup M. The IIP-32: a short version of the Inventory of Interpersonal Problems. *Br J Clin Psychol*. 1996;35(1):21-35.
89. Angermeyer MC, Kilian R, Matschinger H. WHOQOL-100 und WHOQOL-BREF. Handbuch für die deutsche Version der WHO Instrumente zur Erfassung von Lebensqualität. Göttingen: Hogrefe; 2000.
90. Gumz A, Munder T, Schlipfenbacher C, Toussaint K, Warmuth M. Messen interpersoneller therapeutischer Kompetenz. FIS-Manual, deutsche Version. [Assessing interpersonal therapeutic competence. FIS-Manual, German Version]. 2018.

91. Munder T, Schlipfenbacher C, Toussaint K, Warmuth M, Anderson T, Gumz A. Facilitative interpersonal skills performance test: Psychometric analysis of a German language version. *J Clin Psychol.* 2019;75(12):2273-83.
92. Gumz A, Longley M, Schestag L, Hirschmeier C, Derwahl L, Weinreich J, et al. Die „Facilitative interpersonal skills“-Übung – Messen therapeutischer Kompetenz mit der deutschsprachigen Version [The facilitative interpersonal skills task – Assessment of therapeutic competence with the German language version]. *Psychotherapeut.* 2020;65(6):465-74.
93. Gumz A, Longley M, Franken F, Janning B, Hosoya G, Derwahl L, et al. Who are the skilled therapists? Associations between personal characteristics and interpersonal skills of future psychotherapists. *Psychother Res.* 2023:1-11.
94. Anderson T, Ogles BM, Patterson CL, Lambert MJ, Vermeersch DA. Therapist effects: facilitative interpersonal skills as a predictor of therapist success. *J Clin Psychol.* 2009;65(7):755-68.
95. Schmidt J, Lamprecht, F., Wittmann, W.W. . Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen [Satisfaction with inpatient care: Development of a questionnaire and first validity assessments]. *Psychother med Psychol.* 1989;39:248-55
96. Schmidt J, Nübling R. ZUF-8. Fragebogen zur Messung der Patientenzufriedenheit. In: Brähler E, Schumacher J, Strauß B, editors. *Diagnostische Verfahren in der Psychotherapie.* Göttingen: Hogrefe; 2002. p. 392-6.
97. Löffler-Stastka H, Gelo O, Pleschberger I, Schröder T, Orlinsky DE, Rønnestad MH, et al. Psychotherapy Training in Austria: Part 1: Baseline and Sociodemographic Background Data from the SPRISTAD Trial. *psychopraxis neuropraxis.* 2018;21:227-31.
98. Ablner B, Kessler H. Emotion Regulation Questionnaire – Eine deutschsprachige Fassung des ERQ von Gross und John. *Diagnostica.* 2009;55(3):144-52.
99. Bagby RM, Parker JDA, Taylor GJ. Twenty-five years with the 20-item Toronto Alexithymia Scale. *Journal of Psychosomatic Research.* 2020;131:109940.
100. Wilmers F, Munder T, Leonhart R, Herzog T, Plassmann R, Barth J, et al. Die deutschsprachige Version des Working Alliance Inventory-short revised (WAI-SR)-Ein schulübergreifendes, ökonomisches und empirisch validiertes Instrument zur Erfassung der therapeutischen Allianz. *Klinische Diagnostik und Evaluation.* 2008;1(3):343-58.
101. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clin Psychol Psychother.* 2010;17(3):231-9.
102. Hatcher RL, Gillaspay JA. Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research.* 2006;16:12-25.
103. Muran JC, Safran JD, Gorman BS, Samstag LW, Eubanks-Carter C, Winston A. The relationship of early alliance ruptures and their resolution to process and outcome in three time-limited psychotherapies for personality disorders. *Psychotherapy: Theory, Research, Practice, Training.* 2009;46(2):233-48.
104. Muran JC SJ, Samstag LW, Winston A. Patient and therapist post session questionnaires. In: Center BIM, editor. *New York* 1992.
105. Gumz A, Neubauer K, Horstkotte JK, Geyer M, Löwe B, Murray AM, et al. A bottom-up approach to assess verbal therapeutic techniques. Development of the Psychodynamic Interventions List (PIL). *PLoS One.* 2017;12(8):e0182949.
106. Partschefeld E SB, Geyer M, Philipp S. Simulationspatienten in der Psychotherapieausbildung. *Psychotherapeut.* 2013;58(5):438–45.
107. Kühne F, Heinze, P. E., Ay-Bryson, D. S., Maaß, U., Weck, F. Entwicklung einer Skala zur Erfassung psychotherapeutischer Basisfertigkeiten [Development of a scale for assessing basic psychotherapeutic skills]. *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis.* 2021;50(3-4):116–32.
108. Maaß U, Kühne F, Heinze PE, Ay-Bryson DS, Weck F. The concise measurement of clinical communication skills: Validation of a short scale. *Frontiers in psychiatry.* 2022;13:977324.
109. Zilcha-Mano S, Muran JC, Hungr C, Eubanks CF, Safran JD, Winston A. The relationship between alliance and outcome: Analysis of a two-person perspective on alliance and session outcome. *J Consult Clin Psychol.* 2016;84(6):484-96.

110. Gumz A HS, Marquardt G, Franken F, Wulf S, Hosoya G, Klapprott F, Esswein L, Frank B M, Breier M, Siegl J, Geyer M, Hauten L, Reuter L. Assessing Mediators of Change - Development and Psychometric Evaluation of the Mediators of Change in Psychotherapy-Inventory (MoCPI). subm.
111. Rumpold G, Doering S, Hofer S, Schussler G. [A German version of the measure of parental style (MOPS)]. *Z Psychosom Med Psychother.* 2002;48(1):59-74.
112. Parker G, Roussos J, Hadzi-Pavlovic D, Mitchell P, Wilhelm K, Austin MP. The development of a refined measure of dysfunctional parenting and assessment of its relevance in patients with affective disorders. *Psychol Med.* 1997;27(5):1193-203.
113. Rammstedt B KC, Klein MC, Beierlein C, Kovaleva A. Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit. 10 Item Big Five Inventory (BFI-10). *methoden, daten, analysen.* 2013;7 (2):233-49.
114. Rammstedt B KC, Klein MC, Beierlein C, Kovaleva A. Big Five Inventory (BFI-10). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS).* 2014.
115. Müller S WL, Schick Tanz P, Hopwood CJ, Zimmermann J. Development and validation of a German Interpersonal Sensitivities Circumplex (ISC-G): A self-report measure of perceived aversive behaviors of others. *European Journal of Psychological Assessment.* 2024;40(2):146–57.
116. Müller S, Wendt LP, Zimmermann J. Development and Validation of the Certainty About Mental States Questionnaire (CAMSQ): A Self-Report Measure of Mentalizing Oneself and Others. *Assessment.* 2023;30(3):651-74.
117. Watson D FR. Measurement of Social-evaluative Anxiety. *Journal of Consulting and Clinical Psychology.* 1969;33(4):448–57.
118. Kemper CJ LJ, Neuser J. SANB-5. Skala Angst vor negativer Bewertung-5. *Verfahrensdokumentation und Fragebogen.* Trier: Leibniz-Institut für Psychologie (ZPID). Open Test Archive. ; 2011.
119. Schulte D. Messung der Therapieerwartung und Therapieevaluation von Patienten (PATHEV). *Zeitschrift für Klinische Psychologie und Psychotherapie.* 2005;34(3):176-87.
120. Caspar F BT, Fingerle H, Werner M. Das deutsche IMI - Das Impact Message Inventory nach Kiesler und seine Circumplexstruktur. *Psychotherapie Im Dialog.* 2016;17:e1-e10.
121. Schmidt JA WC, Kiesler DJ. Psychometric and circumplex properties of the octant scale impact message inventory (IMI-C): A structural evaluation. *Journal of Counseling Psychology.* 1999;46(3):325-34.
122. Hafkenschied A. The Impact Message Inventory-Circumplex (IMI-C): Exploring the Aversiveness Component of Therapists' Covert Reactions to Patients' Interpersonal Communications. *Clin Psychol Psychother.* 2024;31(4):e3045.
123. Eubanks C, Muran J. RUPTURE RESOLUTION RATING SYSTEM (3RS): MANUAL VERSION 20222023.
124. Eubanks CF, Lubitz J, Muran JC, Safran JD. Rupture Resolution Rating System (3RS): Development and validation. *Psychother Res.* 2019;29(3):306-19.
125. Caspar F BT, Fingerle H, Werner M. IMI-D. Die deutsche Version des Impact Message Inventory (IMI). *Psychotherapie im Dialog.* 2016; 17(04):e1–e11.
126. Ladwig I, Rief W, Nestoriuc Y. Welche Risiken und Nebenwirkungen hat Psychotherapie? - Entwicklung des Inventars zur Erfassung Negativer Effekte von Psychotherapie (INEP) [What are the Risks and Side Effects to Psychotherapy? - Development of an Inventory for the Assessment of Negative Effects of Psychotherapy (INEP)]. *Verhaltenstherapie.* 2014;24(4):252-63.
127. Müller S WLP, Schick Tanz P, Hopwood C J, Zimmermann J. Development and validation of a German Interpersonal Sensitivities Circumplex (ISC-G): A self-report measure of perceived aversive behaviors of others. *European Journal of Psychological Assessment.* 2024;40(2):146–57.
128. Löffler-Stastka H GO, Pleschberger I, Schröder T, Orlinsky DE, Rønnestad MH, Willutzki U. Psychotherapieausbildung in Österreich. Teil 1: Basisdaten und soziodemographische Hintergrunddaten aus der SPRISTAD-Studie. *psychopraxis neuropraxis.* 2018 21:227–31.
129. Kemper C, Lutz J, Neuser J. Skala Angst vor negativer Bewertung-5 (SANB-5). (Hrsg.) L-ZfPIuZDZ, editor. Trier 2011.
130. Ahearn EP. The use of visual analog scales in mood disorders: A critical review. *Journal of Psychiatric Research.* 1997;31(5):569-79.

131. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
132. Babl A, Eubanks CF, Goldfried MR. A 21st Century Principle-Based Training Approach to Psychotherapy: A Contribution to the Momentum of Trans-theoretical Work. *Clin Psychol Eur*. 2024;6(Spec Issue):e11925.
133. Muran JC, Eubanks CF, Lipner LM, Bloch-Elkouby S. Renegotiating tasks or goals as rupture repair: A task analysis in a cognitive-behavioral therapy for personality disorder. *Psychother Res*. 2023;33(1):16-29.
134. Lewin A, Brondeel R, Benmarhnia T, Thomas F, Chaix B. Attrition Bias Related to Missing Outcome Data: A Longitudinal Simulation Study. *Epidemiology*. 2018;29(1):87-95.