AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Consulting and Clinical Psychology

ISSN: 0022-006X

2025, Vol. 93, No. 12, 814–828
https://doi.org/10.1037/ccp0000985

The Long-Term Effectiveness of Psychodynamic and Analytical Psychotherapy in Routine Care: Results From a Naturalistic Study Over 6 Years

Miriam Henkel[1], Johannes Zimmermann[1], Matthias Volz[1], Dorothea Huber[2], Hermann Staats[3], and Cord Benecke[1]

[1] Department of Psychology, University of Kassel

[2] Department for Clinical Psychology and Psychosomatics, International Psychoanalytic University Berlin

[3] Department of Education and Social Sciences, University of Applied Sciences Potsdam

Objective: This naturalistic study examines the long-term effectiveness of two psychoanalytically oriented therapies—psychodynamic psychotherapy (PP) and analytical psychotherapy (AP)—in routine care settings in Germany. ***Method:*** A total of 428 patients diagnosed with various mental disorders according to the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition, underwent treatment with either PP or AP. Diagnoses were reassessed at the end of treatment, and patients provided annual self-reports of their mental health over a 6-year period. We employed propensity score methods and piecewise linear latent growth curve models to estimate treatment effects on primary outcomes (i.e., number of diagnoses, symptom distress) and secondary outcomes (i.e., personality dysfunction, interpersonal problems, general life satisfaction). ***Results:*** Patients exhibited substantial and sustained improvements across all outcomes in both treatments. Notably, those receiving AP experienced greater (long-term) improvement in primary and most secondary outcomes compared to those in PP. Improvements in PP primarily occurred during the 1st year, while AP patients continued to show progress throughout the entire observation period. For some outcomes (symptom distress, personality dysfunction, and interpersonal problems), AP was particularly effective for patients with more severe problems at baseline. ***Conclusions:*** Psychoanalytically oriented treatments in routine care are associated with substantial and lasting improvements, even posttreatment. Our findings highlight an incremental benefit of AP over PP, particularly for complex cases. Future research should explore treatment mechanisms, cost-effectiveness, and implications for health policy.

What is the public health significance of this article?

This study suggests that psychoanalytically oriented therapies, such as psychodynamic psychotherapy and analytical psychotherapy, are associated with substantial improvements for a wide range of mental disorders, with changes that are sustained even after treatment ends. Analytical psychotherapy may offer greater long-term benefits than psychodynamic psychotherapy, particularly for individuals with more complex or severe mental health issues. This study highlights the importance of considering the type and intensity of psychotherapeutic treatment when shaping mental health policy, as different approaches may yield different long-term outcomes for different individuals in routine care settings.

Keywords: psychoanalytically oriented psychotherapy, naturalistic study, long-term psychotherapy, long-term effectiveness, routine care

Supplemental materials: https://doi.org/10.1037/ccp0000985.supp

Michael J. Constantino served as action editor.

Miriam Henkel https://orcid.org/0000-0002-6723-8529
Johannes Zimmermann https://orcid.org/0000-0001-6975-2356
Matthias Volz https://orcid.org/0000-0002-0946-1116
Dorothea Huber https://orcid.org/0009-0006-9007-5971
Hermann Staats https://orcid.org/0009-0007-3706-9852
Cord Benecke https://orcid.org/0000-0002-7977-5102

Miriam Henkel and Johannes Zimmermann shared first authorship. The authors confirm that there are no other studies (published, in press, or under review) that use the same data set as the one used in the submitted article. This study does not have a trial registration number. At the time of its inception in 2009, trial registration was not yet a standard requirement for naturalistic studies in Germany. However, the study protocol was published in a German-language peer-reviewed journal in 2011 (Benecke et al., 2011).

The authors thank the participating outpatient centers for their help with recruitment, all study therapists and research assistants who contributed to this study, the German Psychoanalytic Society for the funding, and all patients who participated in the study.

Miriam Henkel played a lead role in data curation and project administration, a supporting role in methodology and visualization, and an equal role in investigation and writing–original draft. Johannes Zimmermann

continued

Mental disorders represent a significant source of individual suffering and societal burden (e.g., Gustavsson et al., 2011; Vigo et al., 2016), and psychotherapy is a primary evidence-based approach for their treatment (Barkham & Lambert, 2021; Harrer et al., 2025; Leichsenring et al., 2022). Among the various therapeutic modalities, psychoanalytically oriented psychotherapies constitute a diverse family of treatments grounded in psychoanalytic theory, which emphasizes the impact of unconscious processes, motives, and conflicts on human behavior. Interventions typically focus on affect and emotion, interpersonal behaviors, past experiences, (unconscious) resistance, transference and countertransference phenomena, and the exploration of dreams, wishes, and fantasies (Blagys & Hilsenroth, 2000). Efforts to differentiate psychoanalytically oriented approaches show considerable international variation, yet distinctions are typically drawn with regard to treatment duration, intensity (in terms of session frequency), setting, and therapeutic technique. On one end of the spectrum are shorter term psychodynamic psychotherapies (PP), usually held once a week in a face-to-face setting, which often concentrate on a core conflict using more structured and supportive strategies. On the other end are intensive long-term approaches like analytical psychotherapy (AP), usually ranging from two to five weekly sessions. AP is closer to standard psychoanalysis, often employing the couch setting, and focuses on technical neutrality, regression, and in-depth biographical work to facilitate more fundamental changes in personality structure (Henkel et al., 2018, 2020, 2022; Ratzek et al., 2020; Ratzek et al., 2020). This study investigates the long-term effectiveness of these two approaches as delivered in routine clinical practice in Germany.

The efficacy of PP is well-established, as demonstrated in numerous randomized controlled trials (RCTs) and meta-analyses (e.g., Abbass et al., 2014; Leichsenring et al., 2023). There is also considerable evidence from RCTs supporting the effectiveness of psychoanalytically oriented long-term treatments (Leichsenring & Rabung, 2011; Woll & Schönbrodt, 2020). However, the treatments covered in these meta-analyses are highly diverse and do not fully represent AP as delivered in routine care. For instance, the inclusion criteria of 40–50 sessions and a treatment duration of at least 1 year tend to be more representative of PP than AP in health care practice (Rüger et al., 2015). Moreover, the treatments included in these meta-analyses often comprised modified, manualized approaches such as mentalization-based therapy (Bateman & Fonagy, 2016) and transference-focused psychotherapy (Yeomans et al., 2014), while more classical AP conducted in a couch setting was excluded by Woll and Schönbrodt (2020).

Investigating long-term and intensive treatments such as AP poses considerable challenges (Benecke et al., 2016). Applying the gold standard of RCTs, which emphasizes internal validity, is difficult in this context because AP is rarely delivered in manualized form for disorder-specific patient groups. Instead, these treatments follow a predominantly transdiagnostic therapeutic approach and are typically provided to heterogeneous patient populations with high rates of comorbidity. The long-term nature of AP further complicates study design: Maintaining waitlist control groups over extended periods is rarely feasible and raises ethical concerns, while the long follow-up intervals needed to capture sustained effects inevitably increase the risk of attrition and missing data. Consequently, only a limited number of RCTs on AP are available (e.g., Huber et al., 2013; Knekt et al., 2008; Leuzinger-Bohleber et al., 2019), and much of the evidence relies on naturalistic studies (e.g., Berghout et al., 2010; Brockmann et al., 2006; Grande et al., 2006; Leichsenring et al., 2005; Rudolf et al., 1994; Sandell et al., 2000). A meta-analysis of most of these studies found large pre–post effect sizes for AP but underscored the need for more controlled research with long-term follow-up data to draw firmer conclusions (de Maat et al., 2013). Furthermore, many of these studies are limited by small sample sizes or single-center designs, raising concerns about their generalizability.

Given the significant investment of time and resources required for intensive treatments like AP compared to PP, a key question concerns their relative effectiveness. The conceptual rationale for AP's potential superiority lies in its greater depth, duration, and intensity, hypothesized to produce more profound and sustainable structural change. Empirically, however, the evidence is sparse and requires careful interpretation. A handful of studies report faster initial improvements for shorter psychodynamic treatments (Knekt et al., 2008) while suggesting superior long-term outcomes for more intensive psychoanalytic approaches in domains such as symptom distress, personality functioning, and work ability (Berghout et al., 2010; Huber et al., 2013; Knekt et al., 2008; Knekt et al., 2016; Sandell et al., 2000). Other studies found no significant differences or advantages limited to specific outcomes like interpersonal problems (Grande et al., 2006). Crucially, some of these comparative studies were not randomized. This raises the possibility that their findings are confounded by selection bias, where, for instance, patients with specific characteristics (e.g., higher motivation; e.g., Rudolf et al., 1994) are more likely to be allocated to or choose the more intensive treatment, thus inflating its apparent effectiveness.

Beyond average treatment effects, a crucial question in contemporary psychotherapy research is "What works for whom?" (Norcross & Wampold, 2018). Identifying patient characteristics that moderate treatment effects is essential for evidence-based treatment selection, particularly for costly and intensive therapies like AP. Although meta-analyses suggest that long-term psychoanalytically oriented treatments are effective in treating complex mental disorders (Leichsenring & Rabung, 2011; Woll & Schönbrodt, 2020), it is unclear whether a specific approach, such as AP or PP, is more effective for patients with more or less severe impairments. Several statistical approaches have been developed to estimate individual differences in treatment effects and to examine possible moderating baseline variables (e.g., Cohen et al., 2020; Kaiser et al., 2022;

played a lead role in formal analysis, methodology, software, and visualization, a supporting role in data curation and project administration, and an equal role in investigation and writing–original draft. Matthias Volz played a supporting role in data curation and methodology and an equal role in writing–review and editing. Dorothea Huber played a supporting role in conceptualization and funding acquisition and an equal role in writing–review and editing. Hermann Staats played a supporting role in conceptualization and funding acquisition and an equal role in writing–review and editing. Cord Benecke played a lead role in conceptualization and funding acquisition, a supporting role in project administration and writing–review and editing, and an equal role in investigation.

Correspondence concerning this article should be addressed to Miriam Henkel, Department of Psychology, University of Kassel, Holländische Straße 36-38, 34127 Kassel, Germany. Email: miriam.henkel@uni-kassel.de

Mayer et al., 2020). In studies on treatment selection between PP and cognitive-behavioral therapy (CBT), initial symptom severity emerged as an important predictor for differential treatment response (Keefe et al., 2021; Mayer et al., 2020). Moreover, recent findings indicate that AP may be particularly effective for trauma-exposed patients in comparison to CBT (Krakau et al., 2024). However, empirical studies directly testing whether baseline impairment moderates the effectiveness of AP versus PP are missing.

In summary, while the reported studies on psychoanalytically oriented psychotherapies represent a substantial evidence base, several gaps remain. First, studies on AP that cover the necessary long follow-up periods and include large, heterogeneous samples of patients and therapists that are reasonably representative of treatments in routine clinical practice are still scarce. Second, the comparative effectiveness of different forms of psychoanalytically oriented therapies (e.g., AP vs. PP) under real-world conditions has been insufficiently examined. Third, there is limited knowledge about which forms of psychoanalytically oriented therapy work best for whom, although such knowledge is essential for individual treatment selection.

The German Psychoanalytic Society (DPG) Practice study (Benecke et al., 2011) was designed to address these gaps by comparing the outcomes of AP and PP as delivered in routine care in Germany over a 6-year period. Because randomization of patients or therapists was not feasible, we used propensity score (PS) methods (Austin, 2011) to mitigate the effects of selection bias. In line with the study protocol, we aimed to test three hypotheses (Benecke et al., 2011). First, we expected that both PP and AP would be associated with significant and lasting improvements across a range of outcomes, including symptom distress, number of diagnoses, personality dysfunction, interpersonal problems, and general life satisfaction. Second, based on the conceptual rationale of facilitating deeper structural change, we hypothesized that the longer and more intensive AP would prove more effective than PP in the long term. Last, and most importantly for clinical practice, we sought to test the hypothesis that the superiority of AP over PP would be most pronounced for patients with higher levels of impairment at baseline, thus providing evidence for a differential treatment indication.

Method

Procedure

The DPG Practice study is a naturalistic, prospective, long-term study inspecting psychoanalytically oriented psychotherapies in Germany (Benecke et al., 2011). The study protocol was reviewed and approved by the ethics committee of the medical faculty of the Otto von Guericke University of Magdeburg, Germany (Approval Number 32/09). Patients were recruited from 2009 to 2014 in 11 outpatient centers and by study therapists in private practices. After informed consent, patients completed a first self-report questionnaire ("basic documentation"), were interviewed by trained research assistants with Structured Clinical Interviews for *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (*DSM-IV*, SCID; see below), and completed a second self-report questionnaire after having been allocated to a cooperating therapist (T0). Therapists completed a questionnaire on patients' diagnoses and their treatment plans during or after some trial sessions. They also gave information about themselves concerning sociodemographic data and educational background at the beginning of their first study case. Subsequently, patients and therapists completed questionnaires once a year during treatment, and patients for at least 6 years in total irrespective of therapy duration (T1–T6). At the end of therapy, patients were again interviewed with the SCID by trained research assistants who were not informed about the treatment condition; however, there was no active blinding procedure. In general, patients and therapists received their questionnaires by post, together with a stamped addressed envelope, and sent them back to the study center. More detailed information on the study protocol and necessary adaptations is provided in Supplemental Note 1, Supplemental Tables 1 and 2.

Participants

Patients

A total of 887 patients provided informed consent to participate in the longitudinal study. Of these, only 577 patients were in contact with study therapists. We then excluded 149 patients for various reasons, resulting in a final sample of 428 patients who were confirmed to have started either AP ($n = 301$) or PP ($n = 127$) and thus form the basis of our current analyses.[1] Supplemental Figure 1 includes a flowchart detailing the study process and reasons for patient exclusions. Supplemental Table 3 summarizes differences between included and excluded patients in outcome variables at T0.

Table 1 presents sociodemographic data and *DSM-IV* diagnoses for the two treatment conditions. More than two thirds of patients received two or more *DSM-IV* diagnoses. The most common *DSM-IV* Axis I diagnoses were mood disorders and anxiety disorders. Additionally, over half of the patients met the criteria for at least one Axis II personality disorder (PD), with the most prevalent being obsessive–compulsive (18.7%), depressive (17.5%), and avoidant (16.6%) PDs, as detailed in Supplemental Table 4. There were few significant baseline differences between AP and PP patients, the only notable one being a higher level of school education in AP patients (see Henkel et al., 2019, for a detailed report).

Therapists

In total, 234 study therapists were involved in treating patients. Thus, some therapists treated more than one patient (i.e., 137 treated one patient, 60 treated two, and 37 treated three or more patients). In the study, 62.0% of therapists treated only AP cases, 20.1% treated only PP cases, and 17.9% applied both treatment approaches. A total of 189 of the therapists completed the questionnaire about their personal information: The mean age of therapists was $M = 50.27$ ($SD = 9.22$), and 73.5% were female. Therapists were either licensed analytical and psychodynamic therapists or they were in advanced stages of their therapeutic training in AP as well as PP[2]. In particular, 25.9% were in training, and 38.3% were supervisors or

[1] This number is slightly higher than in previous articles on the DPG Practice study (Henkel et al., 2019; Henkel et al., 2020). This is because previous articles investigated the therapy approaches in more detail (and therefore had to exclude patients with missing information on treatment details) and were limited to data at T0 and T1 (and therefore had to exclude patients who only provided data at later time points).

[2] In Germany, training in AP usually includes training in PP; however, you can also train in PP separately (without AP). The training ends with a state examination resulting in a license to practice psychotherapy ("Approbation"). Candidates who treat patients in outpatient care have passed an intermediate examination, and their cases are closely supervised.

Table 1
Sociodemographics and Diagnoses According to the DSM-IV of AP and PP Patients

Variable	AP (n = 301)	PP (n = 127)	All (N = 428)
Age, *M* (*SD*)	36.14 (10.35)	37.48 (12.25)	36.53 (10.95)
	0 missing	1 missing	1 missing
Sex = female	215 (71.67%)	96 (75.59%)	311 (72.83%)
	1 missing	0 missing	1 missing
Highest educational attainment			
Lower secondary	8 (2.80%)	4 (3.31%)	12 (2.95%)
Intermediate secondary	43 (15.03%)	37 (30.58%)	80 (19.66%)
High school diploma	98 (34.27%)	47 (38.84%)	145 (35.63%)
University degree	127 (44.41%)	29 (23.97%)	156 (38.33%)
Different certificate	10 (3.50%)	4 (3.31%)	14 (3.44%)
	15 missing	6 missing	21 missing
Marital status			
Married	81 (28.32%)	40 (32.79%)	121 (29.66%)
Stable relationship	52 (18.18%)	25 (20.49%)	77 (18.87%)
Single	121 (42.31%)	41 (33.61%)	162 (39.71%)
Divorced, separated, widowed	32 (11.19%)	16 (13.11%)	48 (11.76%)
	15 missing	5 missing	20 missing
Living situation			
Alone	101 (36.20%)	33 (27.97%)	134 (33.75%)
With partner	62 (22.22%)	35 (29.66%)	97 (24.43%)
With partner + children	62 (22.22%)	23 (19.49%)	85 (21.41%)
Only with children	16 (5.73%)	4 (3.39%)	20 (5.04%)
With parents	9 (3.23%)	2 (1.69%)	11 (2.77%)
Other living situations	29 (10.39%)	21 (17.80%)	50 (12.59%)
	22 missing	9 missing	31 missing
Number of *DSM-IV* diagnoses[a]			
0[b]	17 (5.82%)	10 (8.40%)	27 (6.57%)
1	65 (22.26%)	32 (26.89%)	97 (23.60%)
2	80 (27.40%)	24 (20.17%)	104 (25.30%)
3	42 (14.38%)	18 (15.13%)	60 (14.60%)
4 or more	88 (30.14%)	35 (29.41%)	123 (29.93%)
	9 missing	8 missing	17 missing
DSM-IV disorder classes[a]			
Mood disorders	188 (64.38%)	76 (63.87%)	264 (64.23%)
Anxiety disorders	145 (49.66%)	58 (48.74%)	203 (49.39%)
Somatization disorders	41 (14.04%)	16 (13.45%)	57 (13.87%)
Substance use	32 (10.96%)	14 (11.76%)	46 (11.19%)
Eating disorders	35 (11.99%)	11 (9.24%)	46 (11.19%)
Adjustment disorder	13 (4.45%)	10 (8.40%)	23 (5.60%)
Personality disorders	157 (53.77%)	59 (49.58%)	216 (52.55%)
	9 missing	8 missing	17 missing

Note. AP = analytical psychotherapy; PP = psychodynamic psychotherapy; *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition.
[a] Partially remitted disorders were also included. [b] Of these patients without a *DSM-IV* diagnosis, some had a recurrent disorder in full remission (often depression) or met several criteria across disorders. Others were noted by the interviewer as more burdened than reported, giving socially desirable responses or lacking emotional access to symptoms. Finally, some sought therapy due to interpersonal difficulties or stressful life situations (e.g., a partner's illness).

training analysts. Therapists also stated how many cases they had treated with specific treatment approaches in their clinical practice during the past year (prior to the study): The average percentage of AP treatments (37.2%) within the prior caseload of each therapist was similar to the average percentage of PP treatments (42.0%). In addition, therapists treated patients with other approaches (20.8%, e.g., crisis interventions, group therapy, family-centered therapy).

Treatments

In Germany, health insurances reimburse two types of psychoanalytically oriented individual psychotherapies, which we inspected in the DPG Practice study. The German Psychotherapy Guidelines (Rüger et al., 2015) describe indication criteria and treatment procedures for these approaches. For psychotherapies longer than 24 sessions (25 sessions prior to 2017 in our study), an independent expert receives a case report from the therapist and assesses whether the treatment indication is adequately determined. In the DPG Practice study, therapists and patients were free to choose a suitable therapy approach, including duration, setting, and frequency of sessions, without external interference. We determined whether a treatment was PP or AP by inspecting the applied and approved therapeutic approach of the health insurance application after around 1 year of treatment. We used this time point because some therapists used a short-term PP as a kind of test therapy and then applied for AP. These cases were assigned to AP. After 1 year, treatment approaches did not

change. Therapies adhered to the guidelines regarding setting and session frequency during the main treatment phase[3] as well as the technique used (according to therapists' self-reports; Henkel et al., 2020).

Analytical Psychotherapy

AP is a long-term psychotherapy with two to three (seldom four) sessions per week. Public health insurances cover 160–300 sessions; more sessions require a special authorization or a personal payment from the patient. AP can be carried out close to "standard psychoanalysis" in a couch setting and a technical orientation on regression, transference, and interpretation. The guidelines also describe a modified AP with setting modifications (face-to-face setting or lower frequency, e.g., two sessions a week) and/or technical modifications (e.g., supportive, structural interventions). In our study, we included both types of AP (with couch setting being slightly more prominent in the 1st year of treatment, 65.1%). Setting and frequency could change during therapies (e.g., from two sessions in the face-to-face setting to three sessions in the couch setting). Our APs had a mean duration of $M = 3.25$ years ($SD = 19.27$ months) and an average number of sessions of $M = 228.9$ ($SD = 108.77$). Therapists stated that treatments ended irregularly[4] in 24.7% of all AP treatments ($n = 219$, due to missing data at the end of treatment).

Psychodynamic Psychotherapy

PP is, in Germany, a generic term subsuming shorter forms of psychoanalytically oriented treatments, such as dynamic therapy or focal therapy (see Rüger, 2002, for a detailed description). Treatments are up to 100 sessions long and carried out in a face-to-face setting with one to two sessions per week. Technical emphasis is on a central focus or conflict; (counter)transference is observed but less frequently used for interpretation. In our study, PPs mainly took place once a week. The mean duration was $M = 1.58$ years ($SD = 15.56$ months), and the average number of sessions was $M = 47.82$ ($SD = 29.73$). Therapists stated that treatments ended irregularly in 29.5% of all PP treatments ($n = 95$, due to missing data at the end of treatment).

Measures

For the current analyses, we focused on transdiagnostic primary (i.e., symptom distress and number of *DSM-IV* diagnoses) and secondary outcomes (i.e., personality dysfunction, interpersonal problems, and general life satisfaction). Most outcomes were assessed from T0 to T6, encompassing seven measurement points. The SCIDs were only conducted at baseline and at the end of treatment (see Supplemental Note 1 for further details).

Symptom Checklist–90–Revised

The Symptom Checklist–90–Revised (Derogatis, 1977; Franke, 2002) is a self-report instrument assessing symptomatic burden on nine subscales (Somatization, Obsessive–Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism). Patients evaluate the impairment in the last 7 days on a 5-point Likert scale (0 = *not at all*, 4 = *extremely*). The mean score of all 90 items is a global severity index of symptom distress (Cronbach's $\alpha = .96$ at baseline).

SCID-I and SCID-II

SCIDs are semistructured interviews. SCID-I (First et al., 2002) assesses *DSM-IV* Axis I disorders (mood, psychotic, substance use, anxiety, somatoform, eating, and adjustment disorders), and SCID-II (First et al., 1997) assesses PDs of the *DSM-IV* Axis II (Cluster A: paranoid, schizoid, and schizotypal; Cluster B: borderline, narcissistic, histrionic, and antisocial; Cluster C: avoidant, dependent, and obsessive–compulsive; appendix categories: depressive, negativistic, and PD not otherwise specified). SCIDs were conducted by trained research assistants. As a deviation from the manual, research assistants did not use the SCID-II screening questionnaire but evaluated all SCID-II criteria. In addition, primary diagnoses were not used; a single patient could receive multiple diagnoses. We focused on the total number of *DSM-IV* diagnoses (Axis I and Axis II). Partially remitted disorders were weighted by 0.5.

Inventory of Personality Organization–16

The Inventory of Personality Organization–16 (Zimmermann et al., 2013, 2020) assesses self-reported personality dysfunction based on Kernberg's model of personality organization (Kernberg & Caligor, 2005; domains: identity diffusion, primitive defenses, and impaired reality testing). Patients use a 5-point Likert scale (1 = *never true*, 5 = *always true*) to evaluate the 16 items. The total mean score reflects overall impairments in personality functioning (Cronbach's $\alpha = .82$ at baseline in our sample).

Inventory of Interpersonal Problems–64

The Inventory of Interpersonal Problems–64 (Alden et al., 1990; Horowitz et al., 2016) assesses self-reported interpersonal difficulties based on the interpersonal circle model. Scales represent the octants of the interpersonal circle (domineering, vindictive, cold, socially inhibited, nonassertive, overly accommodating, self-sacrificing, and intrusive). The two orthogonal dimensions of agency and communion form the main axes of this circular arrangement. The 64 items include behaviors that are "hard for you to do" or that "you do too much." Patients evaluate how much they suffer from these behaviors on a 5-point Likert scale (0 = *not at all*, 4 = *extremely*). The dimensions agency and communion were only considered for multiple imputation and PS weights (see below). Outcome analyses focused on the total mean score measuring general interpersonal distress (Cronbach's $\alpha = .94$ at baseline in our sample).

[3] The average session frequency over the full course of treatment was 1.35 per week in AP and 0.58 in PP, lower than described in the guidelines. This was due to holidays, illness, and the common practice of session "fading out," especially in AP, where frequency gradually decreased toward the end of treatment (e.g., every 2 weeks, monthly).

[4] Therapists cited several reasons for irregular endings, mainly patients discontinuing treatment, but also that patients moved away or were already satisfied with treatment achievements.

Questions on Life Satisfaction

The Questions on Life Satisfaction (Henrich & Herschbach, 2000; Huber et al., 1988) is a questionnaire to evaluate generic aspects of quality of life by a general module, "General Life Satisfaction," and a disease-specific module, "Health." Here, we focus on the module "General Life Satisfaction." Patients evaluate the subjective importance of eight life domains (friends, hobbies, health, income/financial security, work, housing situation, family life/children, and partnership/sexuality) on a 5-point Likert scale (1 = *not important*, 5 = *very important*), as well as their satisfaction with each domain on another 5-point Likert scale (1 = *not satisfied*, 5 = *very satisfied*). Importance and satisfaction for each life domain are multiplied, and a mean score of all domains is calculated as a measure of general life satisfaction (Cronbach's $\alpha = .60$ at baseline in our sample).

Statistical Analyses

We analyzed data from all 428 patients who initiated one of the two treatments, corresponding to a modified intention-to-treat approach (Kahan et al., 2023). Our multistage analysis was designed to minimize attrition and selection bias before testing hypotheses, following a sequence of multiple imputation, PS weighting, and latent growth curve modeling. We conducted all analyses in R 4.2.3 (R Core Team, 2023) using the mice (van Buuren & Groothuis-Oudshoorn, 2011), lavaan (Rosseel, 2012), and lavaan.mi (https://github.com/TDJorgensen/lavaan.mi) packages.

Multiple Imputation

To address potential attrition bias and make the missing at random assumption more plausible, we used multiple imputation (D. B. Rubin, 1987). We employed a fully conditional specification approach, which facilitates flexible, variable-by-variable imputation by specifying a unique imputation model for each incomplete variable and iteratively generating imputations (van Buuren et al., 2006). Imputation models were developed drawing on information from all outcome variables and an additional 63 sociodemographic and clinical variables (see Supplemental Table 5). To ensure the imputation model for a target variable was compatible with subsequent analysis models, we included all other measurement points of an outcome variable as predictors when imputing that outcome at a specific time point. Additional potential predictors were screened based on their association with the variable's magnitude (using linear regression) and its missingness status (using logistic regression). These associations were computed separately within each treatment condition. A variable was included as a potential predictor in an imputation model if it showed a statistically significant effect ($p < .05$) in both regression models in at least one treatment condition.

Given the large number of potential predictors for some outcomes, we incorporated Lasso regression (Tibshirani, 1996) as an intermediate variable selection step to prevent unstable estimates. The Lasso penalty hyperparameter was tuned using 10-fold cross-validation. In the final imputation models, we only included predictors whose coefficients remained nonzero after the Lasso procedure. This strategy aligns with the indirect use of regularized regression in the context of multiple imputation (Zhao & Long, 2016).

We performed imputations using Bayesian linear, logistic, and polytomous regression models with standard noninformative priors. To align with potential treatment–covariate interactions in subsequent analysis models, imputations were conducted separately for the two treatment conditions. The imputation process was performed in cycles; after an initial 50 cycles to ensure stability, the imputed values were stored in a completed data set. This entire procedure was repeated to generate a total of 100 imputed data sets. The convergence of the imputation algorithm was confirmed via visual inspection of convergence plots.

PS Weighting

To minimize selection bias inherent in the nonrandomized study design, we used the PS method (Austin, 2011). We first identified baseline variables that were bivariately associated ($p < .10$) with treatment condition by estimating simple logistic regression models for each predictor in each of the 100 imputed data sets and subsequently pooling the results. We then computed a PS for each patient in each imputed data set—the predicted probability of receiving PP (vs. AP)—using a multiple logistic regression model that simultaneously included all previously selected predictors.

Rather than matching or using conventional inverse probability weights, we transformed the PS into overlap weights (Zhou et al., 2020). Overlap weights give higher weight to individuals with a moderate probability of receiving either treatment, effectively focusing the analysis on patients who could plausibly receive both treatments. This approach maximizes precision, reduces the variability of estimates compared to inverse probability weighting, and avoids excluding patients from the analysis. The resulting target estimand (for the second hypothesis) is the average treatment effect in the overlap population—patients for whom the choice of treatment was most uncertain.

To provide a check on whether covariate balancing was successful, we repeated the initial bivariate logistic regressions in the observed data set, this time applying the aggregated overlap weights. Successful adjustment is indicated when no measured baseline variable retains a significant association with the treatment condition in the weighted data.

Latent Growth Curve Modeling

We modeled the longitudinal trajectories for each outcome using multiple-group latent growth curve models (LGCMs), incorporating the overlap weights from the PS method. All models included latent intercepts (representing individual baseline levels) and latent slopes (representing individual rates of change). Treatment conditions were modeled as separate groups, allowing for the estimation of treatment-specific means and variances of the latent intercept and slope factors. This multiple-group approach also enabled us to model treatment-specific effects of latent intercepts on latent slopes. This parameterization directly tests for an interaction between a patient's baseline level and the treatment condition on the rate of change in the outcome. To stabilize estimation, residual variances were constrained to be equal across time points and treatment conditions.

For the four outcomes with seven measurement points, we compared four distinct models with varying assumptions about the

trajectory form: (a) a linear LGCM, assuming a constant rate of change; (b) a linear-log LGCM, where the rate of change decreases over time; (c) a piecewise linear LGCM, allowing a different rate of change in the 1st year compared to subsequent years; and (d) a flexible nonlinear LGCM or "latent basis model," where the form of change was freely estimated in each group (McNeish & Matta, 2018). For the outcome with only two measurement points, we used a linear LGCM with residual variances fixed to zero (Voelkle, 2007).

The optimal model was selected based on a combination of fit indices (comparative fit index > .95, root-mean-squared error of approximation < .06; Hu & Bentler, 1999) and information criteria. For the comparative fit index, we used a more restrictive baseline model appropriate for LGCMs (Widaman & Thompson, 2003). We did not interpret the standardized root-mean-squared residual, as its utility for LGCMs is questionable (e.g., McNeish & Matta, 2018). Models were estimated in each imputed data set using maximum likelihood. Test statistics were pooled across data sets using the D4 method (Grund et al., 2023) and robustified using the Yuan–Bentler method. To account for the nesting of patients within therapists, we computed cluster-robust standard errors, and the variance–covariance matrix was pooled as the weighted sum of within- and between-imputation components (Enders, 2010). Wald tests were used to determine p values and confidence intervals.

Hypothesis Testing and Effect Size Calculation

Using the estimated parameters from the selected LGCM for each outcome, we tested our three focal hypotheses: (a) the expected total change within each treatment condition; (b) the average treatment effect, defined as the expected difference in overall change between treatments for a person with an average baseline level; and (c) the interaction between treatment condition and baseline level. Standardized effect sizes (Cohen's d) were calculated using the standard deviation (*SD*) of the respective outcome in the general population (Alsalti et al., 2024), with the exception of the number of *DSM-IV* diagnoses, for which we used the pooled study *SD* at the end of treatment.

We employed an individual testing strategy with an α level of .05 for each of the five outcomes (M. Rubin, 2021). This approach was chosen over joint hypothesis testing (e.g., with Bonferroni correction) to provide greater statistical power and allow for outcome-specific conclusions. This was particularly appropriate for our study, given that limitations with certain outcomes made them less suitable for inclusion in a single, omnibus test. For example, data on the number of *DSM-IV* diagnoses were only available at the end of treatment and not at the intended 6-year follow-up (see Supplemental Note 1).

Transparency and Openness

The imputed datasets and R analysis code required to reproduce the findings of this study (specifically, the LGCMs) are permanently archived on the Open Science Framework at https://osf.io/4dcu6/. The full data containing additional sociodemographic and clinical variables that were used for multiple imputation and PI weighting are not publicly available in order to protect the privacy of study participants.

Results

Descriptive statistics for each time point in the two treatment conditions without imputation or PS weights can be found in Supplemental Table 6. The proportion of missing values was substantial and increased over time, ranging from around 36% at T1 to approximately 70% at T6. Applying PS weights successfully balanced the measured covariates between AP and PP (see Supplemental Note 2, Supplemental Table 7, and Supplemental Figure 2 for additional information). The piecewise linear LGCM was the optimal model for representing changes in the four outcomes with more than two measurement points (see Supplemental Table 8). Consequently, we used this type of model to test our hypotheses. The pooled model parameters for each outcome across the 100 imputations can be found in Supplemental Tables 9–13.

Table 2 presents the expected total changes and treatment effects for individuals with average baseline levels of the outcome variables. Figure 1 depicts the corresponding outcome trajectories (see Supplemental Figure 3 for individual outcome trajectories). Regarding the first hypothesis, significant improvements were found for all outcomes in both treatment conditions, with within-group effect sizes ranging from increased general life satisfaction of $d = 0.53$ in PP to reduced symptom distress of $d = -1.57$ in AP. Regarding the second hypothesis, AP demonstrated larger improvements than PP for all outcomes except general life satisfaction, with between-group effect sizes ranging from $d = -0.25$ for personality dysfunction to $d = -0.45$ for symptom distress.

Figure 1 reveals a distinct pattern across all outcomes modeled with the piecewise LGCM, characterized by stronger improvements in PP during the 1st year and in AP over the subsequent years. Expected annual rates of change over all periods are detailed in the first rows of Supplemental Tables 14–18. For instance, the expected improvement in PP was −0.381 points on the global severity index in the 1st year and then decreased to −0.012 per year, while in AP, it was −0.333 in the 1st year and then still −0.056 per year (see Panel A of Figure 1). This corresponds to a (nonsignificantly) greater improvement in PP in the 1st year ($B = 0.048$, $p = .379$) and a significantly greater annual improvement in AP in the following years ($B = -0.044$, $p < .001$). A similar pattern was observed in secondary outcomes, although statistical tests were only significant for personality dysfunction and interpersonal problems. It should also be noted that in AP, all outcomes continued to improve significantly from the 2nd year onwards, whereas this rate of change was only significant for personality dysfunction and interpersonal problems in PP.

Table 3 summarizes the interaction effects between treatment condition and outcome at baseline that are relevant for testing the third hypothesis. For symptom distress, personality dysfunction, and interpersonal problems, the interaction effect was significant; for number of *DSM-IV* diagnoses and general life satisfaction, however, it was not significant ($p > .20$). Figure 2 illustrates the expected difference between treatments over time for individuals with varying baseline levels. For example, for individuals with high baseline distress, the expected total change was −0.343 points greater with AP than with PP ($d = -0.88$, $p < .001$), whereas the difference was negligible for individuals with low baseline distress ($B = -0.007$, $d = -0.02$, $p = .937$; see Panel A in Figure 2 and Supplemental Table 14). Similar patterns emerged for personality dysfunction (see Panel C in

Table 2
Expected Total Changes and Treatment Effects for Individuals With Average Baseline Levels of the Respective Outcome

	AP				PP				Difference		
Outcome	*B* (*SE*)	95% CI	*p*	d_w	*B* (*SE*)	95% CI	*p*	d_w	*B* (*SE*)	95% CI	*p*	d_b
GSI	−0.614 (0.035)	[−0.683, −0.545]	<.001	−1.574	−0.439 (0.057)	[−0.551, −0.328]	<.001	−1.126	−0.175 (0.063)	[−0.297, −0.052]	.005	−0.449
SCID	−1.738 (0.161)	[−2.055, −1.421]	<.001	−1.086	−1.148 (0.253)	[−1.646, −0.651]	<.001	−0.717	−0.589 (0.283)	[−1.146, −0.033]	.038	−0.368
IPO-16	−0.559 (0.035)	[−0.627, −0.491]	<.001	−0.902	−0.403 (0.055)	[−0.510, −0.295]	<.001	−0.650	−0.156 (0.063)	[−0.281, −0.032]	.014	−0.252
IIP-64	−0.538 (0.048)	[−0.632, −0.444]	<.001	−1.055	−0.331 (0.083)	[−0.494, −0.169]	<.001	−0.649	−0.207 (0.094)	[−0.391, −0.023]	.028	−0.406
FLZ-M	3.162 (0.322)	[2.529, 3.795]	<.001	0.679	2.470 (0.568)	[1.354, 3.586]	<.001	0.530	0.692 (0.658)	[−0.600, 1.985]	.293	0.148

Note. $N = 428$. Estimates are derived from the latent growth curve models pooled across 100 imputed data sets. AP = analytical psychotherapy; PP = psychodynamic psychotherapy; *SE* = standard error; CI = confidence interval; GSI = Global Severity Index of the Symptom Checklist–90–Revised; SCID = number of diagnoses according to the Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders*; IPO-16 = Inventory of Personality Organization–16; IIP-64 = Inventory of Interpersonal Problems–64; FLZ-M = Questions on Life Satisfaction.

Figure 2 and Supplemental Table 16) and interpersonal problems (see Panel D in Figure 2 and Supplemental Table 17).

Discussion

Psychotherapy research has often focused on the efficacy of manualized short-term psychotherapies for specific disorders, prioritizing high internal validity. This approach can be unsatisfactory, especially for psychoanalytically oriented therapies, which are often nonmanualized, long-term, and transdiagnostic. This naturalistic study investigated AP and PP as they are delivered in routine care, surveying patients with various disorders over a 6-year period from the start of treatment. Below, we contextualize our findings within the existing literature, discuss study limitations, and raise questions for future research.

Effectiveness of Psychoanalytically Oriented Treatments in Routine Care

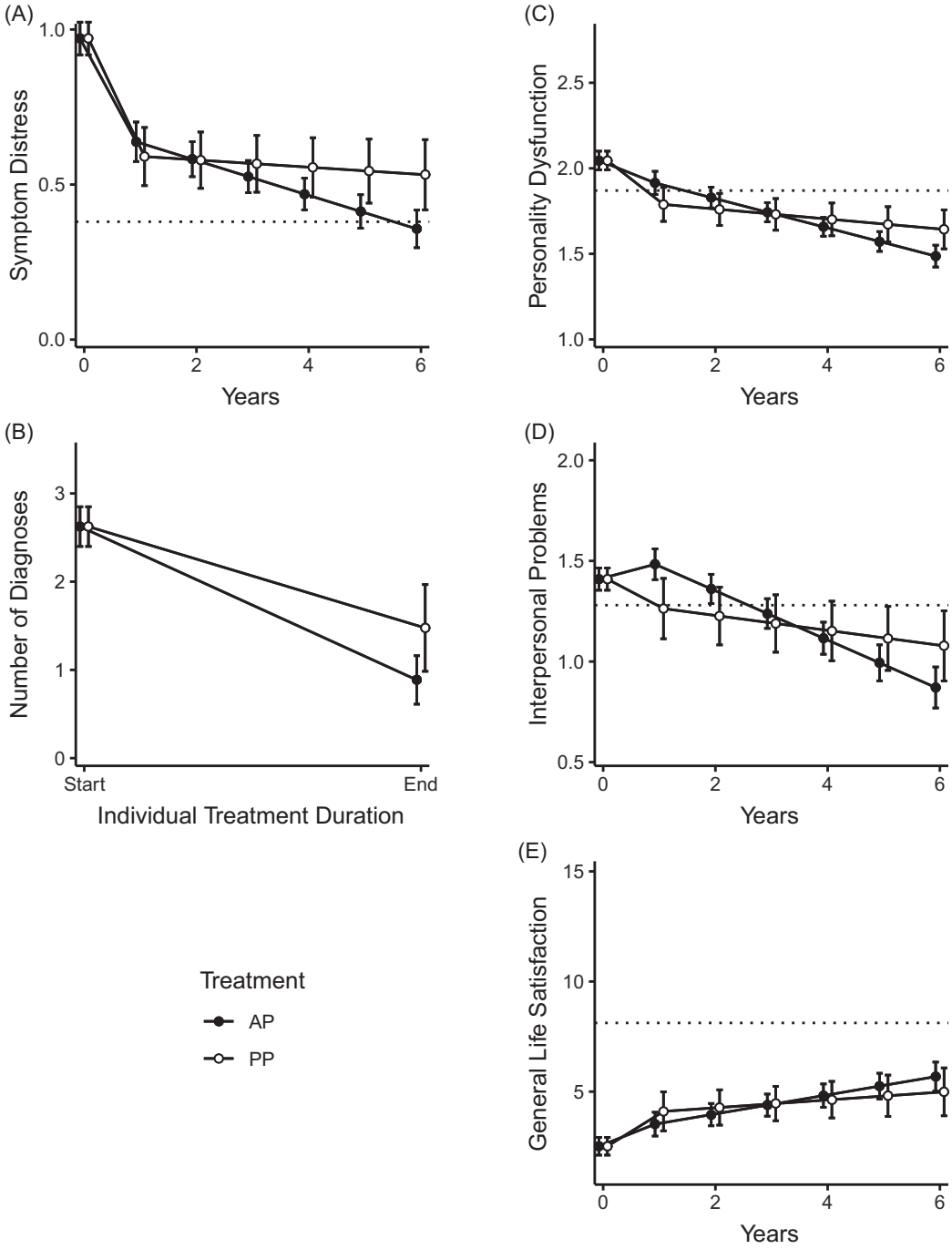
Overall, we observed mostly large within-group effect sizes for improvements in both treatment conditions. This aligns with the first hypothesis in the study protocol as well as with previous studies on psychoanalytically oriented therapies with long observation periods in Germany (Huber et al., 2013; Leuzinger-Bohleber et al., 2019) and in other countries (Knekt et al., 2011; Sandell et al., 2000). More specifically, the largest within-group effect sizes ($d > 1$) were found for symptom distress. Reductions in the number of *DSM-IV* diagnoses, personality dysfunction, and interpersonal problems were also large, with effect sizes in PP for interpersonal problems comparable to those reported for brief therapies of depression (McFarquhar et al., 2018). For general life satisfaction, only medium within-group effect sizes were found, consistent with previous studies on psychoanalytically oriented therapies (Huber et al., 2012). While the pattern of within-group effect sizes suggests a positive impact of AP and PP, it should be interpreted with caution, given that our study lacked a waitlist control or placebo condition. Consequently, it remains unclear what changes might occur in this patient group without treatment. Meta-analyses indicate that placebo treatments can also be associated with large within-group effect sizes (Bschor et al., 2024). However, in typical RCTs, high comorbidity rates—such as those in our sample—are uncommon, and the observation periods are generally much shorter than in our study.

Comparing the Effectiveness of AP and PP

Compared to PP, intensive long-term treatments such as AP are more costly—in terms of time, finances, and possibly emotional resources. Thus, it is crucial to critically evaluate the expected benefits of these treatments. Our second hypothesis posited that the long-term effects of AP would surpass those of PP. Our results indeed demonstrated that AP had an advantage over PP in reducing the number of *DSM-IV* diagnoses at the end of treatment. Additionally, AP showed long-term benefits in alleviating symptom distress, personality dysfunction, and interpersonal problems, emphasizing the importance of long-term follow-ups in this field of research. The between-group effect sizes ranged from small to medium. However, conventional interpretations of Cohen's effect sizes cannot be directly translated into clinical relevance, as this largely depends on the

Figure 1

Expected Outcome Trajectories of an Individual With Average Baseline Levels as a Function of Treatment Condition



Note. Primary outcomes (left side) include symptom distress (Panel A) and the number of diagnoses (Panel B); secondary outcomes (right side) include personality dysfunction (Panel C), interpersonal problems (Panel D), and general life satisfaction (Panel E). The dotted lines represent an average individual from the general population. The range of the *y*-axis corresponds to up to three standard deviations of the outcome variable. In Panels A, B, and C, the *y*-axis is truncated, with the lowest value corresponding to the theoretical minimum of the scale (i.e., absence of problems). AP = analytical psychotherapy; PP = psychodynamic psychotherapy.

Table 3
Interaction Effects Between Latent Intercept Factors and Treatment Condition on Change in Outcomes

	Total			S1			S2		
Outcome	*B* (*SE*)	95% CI	*p*	*B* (*SE*)	95% CI	*p*	*B* (*SE*)	95% CI	*p*
GSI	−0.351 (0.142)	[−0.630, −0.072]	.014	0.083 (0.131)	[−0.173, 0.340]	.524	−0.087 (0.026)	[−0.139, −0.035]	.001
SCID	−0.166 (0.148)	[−0.456, 0.124]	.260						
IPO-16	−0.370 (0.143)	[−0.651, −0.089]	.010	−0.021 (0.138)	[−0.292, 0.250]	.881	−0.070 (0.029)	[−0.127, −0.013]	.017
IIP-64	−0.529 (0.252)	[−1.023, −0.035]	.036	−0.053 (0.212)	[−0.468, 0.363]	.803	−0.095 (0.047)	[−0.188, −0.003]	.044
FLZ-M	0.026 (0.341)	[−0.644, 0.696]	.939	0.102 (0.312)	[−0.512, 0.715]	.745	−0.015 (0.066)	[−0.146, 0.116]	.820

Note. $N = 428$. Estimates are derived from the latent growth curve models pooled across 100 imputed data sets. Total = total change across the full period; S1 = latent slope factor representing the annual rate of change during the 1st year; S2 = latent slope factor representing the annual rate of change during the remaining years; *SE* = standard error; CI = confidence interval; GSI = Global Severity Index of the Symptom Checklist–90–Revised; SCID = number of diagnoses according to the Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders*; IPO-16 = Inventory of Personality Organization–16; IIP-64 = Inventory of Interpersonal Problems–64; FLZ-M = Questions on Life Satisfaction.

outcome measure (among other factors). To address this, the concept of minimal (clinically) important difference (Cuijpers et al., 2014; Angst et al., 2017) has been proposed. This refers to the smallest change in a treatment outcome that patients perceive as beneficial and that would justify a change in clinical management in the absence of negative side effects or excessive costs. Cuijpers et al. (2014) identified an effect size of 0.24 as a minimal important difference for depressive symptoms. The significant between-group effects in our study are in this range or higher, suggesting that—even though our outcome measures also targeted different problems—they may represent clinically important differences. However, the superiority of AP was not observed for general life satisfaction, meaning that our second hypothesis was only partially supported. For this outcome, it is important to consider that the heterogeneity of items resulted in an unreliable total score, which may have limited the detection of treatment effects.

When considering the timing of change, improvements in PP predominantly occurred during the 1st year of treatment, with symptoms generally remaining at low levels thereafter. For personality dysfunction and interpersonal problems, gains even continued beyond the 1st year, although less pronounced than in AP. In contrast, the relative effectiveness of AP increased over time, eventually surpassing that of PP after 6 years. This aligns with findings from Knekt et al. (2011), who noted that short-term therapies tend to yield quicker results, while psychoanalysis and long-term therapies prove more effective in the long run. Similar results have been observed in other studies (Huber et al., 2013; Sandell et al., 2000). Leuzinger-Bohleber et al. (2019) examined outcomes 3 years after the start of treatment, at a time when AP treatments had often just ended or were even still ongoing. They found no significant differences in symptom distress between AP and CBT. Our findings suggest that the absence of differences at that point could also be attributed to the timing of the assessments, as differences may only become apparent several years after the end of treatment.

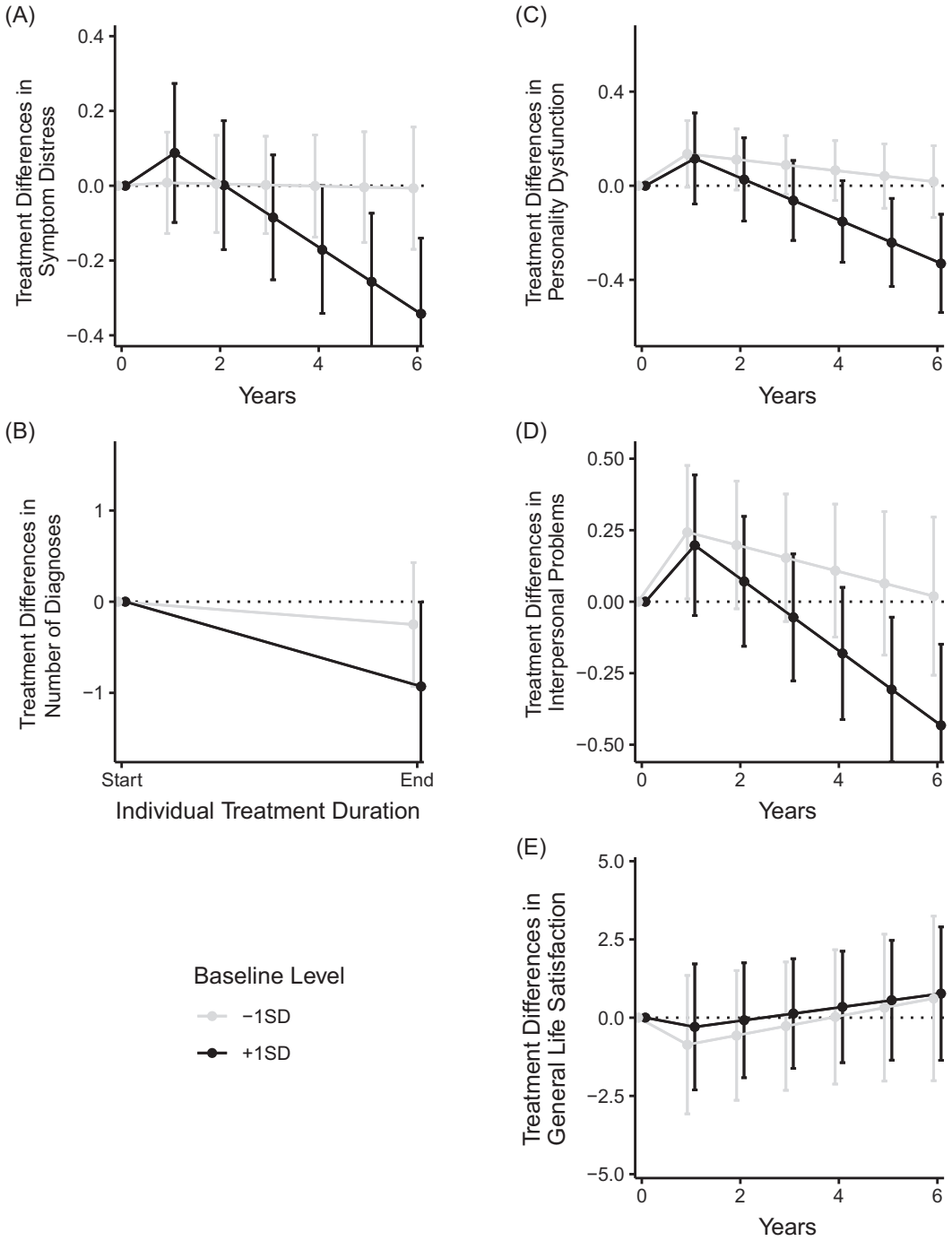
There are several factors that could account for the observed differences in treatment outcomes. The most prominent factors are treatment duration and session frequency. On average, AP treatments were much longer and included more sessions per week (3.25 years and 1.35 sessions per week) compared to PP treatments (1.58 years and 0.58 sessions per week). It is thus reasonable to assume that such differences were primary drivers of the treatment effects. For example, a meta-analysis of long-term treatments (Leichsenring & Rabung, 2011) showed that within-group effect sizes were positively correlated with the total number of sessions across treatment conditions. In addition, Bruijniks et al. (2020) found that therapies taking place more than once a week have greater effects, not only in psychoanalytically oriented therapies but also in CBT. The question remains whether PP treatments would have been equally effective had their duration and session frequency matched those of AP treatments. Another possible explanation for the differences in outcomes could be differences in setting and psychotherapeutic techniques. For example, a study found that the greater reduction in depressive symptoms during follow-up in AP compared with CBT was mediated by psychoanalytic techniques rather than a higher number of sessions (Zimmermann et al., 2015). Future analyses from the DPG Practice study will look more closely at the mechanisms of change underlying the treatments.

Exploring the Heterogeneity of Treatment Effects

Previous research on short-term treatments has examined several patient characteristics as moderators of treatment response, with baseline severity of symptoms being a key factor (Cohen et al., 2020; Keefe et al., 2021; Mayer et al., 2020). Our study replicated these findings in long-term treatments for symptom distress, personality dysfunction, and interpersonal problems, partially supporting the third hypothesis in the study protocol: While patients with less severe problems at baseline (−1 *SD*) did not show any differences in outcomes between AP and PP, the differences increased with greater baseline severity. Indeed, for those with elevated problems at baseline (+1 *SD*), the estimated advantage of AP over PP was moderate to strong, with effect sizes ranging from $d = 0.53$ for personality dysfunction to $d = 0.88$ for symptom distress. This result extends initial findings on treatment effect modifiers of AP (Krakau et al., 2024) and has specific implications for the application of AP in the health insurance system in Germany but also for long-term approaches close to standard psychoanalysis in general: For patients with less severe symptoms, impairments in personality functioning, or interpersonal problems, the added value of AP may be minor or even nonexistent. In these cases, the less costly PP yields comparable outcomes, with symptoms generally declining within a year and remaining at low levels after the end of treatment. However, for patients with severe symptoms, impairments in personality functioning, or interpersonal problems, AP may be indicated because stronger effectiveness can be expected.

Figure 2

Expected Outcome Differences Between Treatments as a Function of Baseline Level



Note. Primary outcomes (left side) include symptom distress (Panel A) and the number of diagnoses (Panel B); secondary outcomes (right side) include personality dysfunction (Panel C), interpersonal problems (Panel D), and general life satisfaction (Panel E). Positive values indicate an advantage for psychodynamic psychotherapy and negative values an advantage for analytical psychotherapy—except in Panel E, where the opposite is true. The range of the *y*-axis corresponds to two standard deviations of the outcome variable.

Limitations

The study has several limitations that mainly threaten the internal validity of results. First, the inclusion of patients in the study only depended on a therapist offering treatment and not on the patient fulfilling specific diagnostic criteria. This resulted in a heterogeneous group of individuals being included, ranging from those with no *DSM-IV* diagnoses to those with more than four *DSM-IV* diagnoses. Although this may seem unfavorable from the perspective of traditional RCTs of specific disorders, it reflects the reality of routine practice, contributes to the transferability of findings, and is consistent with recommendations for severity-based sampling from the perspective of dimensional models of psychopathology (Kramer et al., 2022). Greater heterogeneity may also be advantageous from a statistical point of view: For example, the inclusion of mild cases may reduce the likelihood of overestimating changes within treatment groups (by mitigating regression to the mean) and improve the power to detect differential treatment effects (by increasing the variance in baseline severity).

Second, we applied a modified intention-to-treat analysis, focusing only on patients who were confirmed to start AP or PP. This approach excluded patients who provided consent but were not in contact with study therapists or who had no confirmed AP or PP treatment (mainly because therapists did not provide information on the form of therapy; see Supplemental Figure 1). As shown in Supplemental Table 3, the latter patients were similar to our sample in terms of their baseline characteristics, whereas patients who were not in contact with therapists were more severely impaired. Notably, this suggests that our results are generalizable to patients who start sessions with therapists in private practices but may not be representative of those presenting in outpatient clinics and seeking psychotherapy.

Third, patients were not randomly assigned to the two treatment conditions, and in some cases, this assignment even occurred after therapy had already begun. For example, therapists sometimes used short-term PP (up to 25 sessions) as an initial test phase and decided for AP later during the 1st year of therapy. Although we attempted to mitigate these selection effects by using propensity weights to parallelize patients across treatment groups, this method can only address observed baseline imbalances. Unaccounted relevant variables may exist that could introduce bias into the estimation of causal treatment effects.

Fourth, we did not explicitly model differences in therapist effectiveness. For example, in the DPG Practice study, some therapists treated patients in both the PP and AP conditions (i.e., they were crossed over treatment conditions), whereas most therapists treated patients in only one condition (i.e., they were nested in treatment condition). Although it is possible to consider these design complexities in statistical models (e.g., Walwyn & Roberts, 2010), this is challenging with the multigroup LGCM approach we chose. Nevertheless, we sought to mitigate the risk of false-positive treatment effects by estimating model parameters with cluster-robust standard errors.

Fifth, therapists did not deliver the treatments according to a manual, and the integrity of the treatment was not ensured by independent raters because the sessions were not video- or audio-recorded. Nevertheless, therapists' self-reports indicated that the two treatment approaches differed in theoretically consistent ways with respect to settings and techniques (Henkel et al., 2019, 2020).

Sixth, organizational issues arose during the implementation of the DPG Practice study, resulting in a total sample that was smaller than anticipated, a relatively high rate of missing values, and several adaptations to the study protocol (see also Supplemental Note 1). For example, during the study, it became evident that therapists often failed to return their questionnaires or even to inform researchers that therapy had ended. As a result, SCIDs were frequently not conducted at the appropriate time, leading to more than 70% missing data at the end of treatment. Although we were able to minimize attrition bias by using multiple imputation, the precision and power of the analyses are considerably reduced by such a large number of missing values. At the same time, the number of patients included in this study was still larger than in most previous studies of psychoanalytically oriented therapies. Future long-term studies should improve the balance between ecological validity and research control (e.g., by using more structured follow-up mechanisms to reduce attrition rates).

Last, the study protocol (Benecke et al., 2011) lacked specificity, allowing for researcher degrees of freedom in the statistical analyses. To mitigate potential bias due to researcher allegiance, the analyses were conducted by a neutral methodologist (JZ) with no conflict of interest, such as membership in the DPG. Nevertheless, some methodological choices may remain open to debate. For instance, we used a widely accepted α level of .05 for individual tests. Alternatively, adjusting the α level to maintain a family-wise error rate of .05 could be considered. Note that using a conservative Bonferroni correction ($p < .05/5 = .01$) would still reject the joint null hypotheses (e.g., that AP and PP do not differ across all outcomes).

Conclusion

We presented results from the DPG Practice study, which examined the effectiveness of two psychoanalytically oriented therapies in routine care settings in Germany. The study is relatively unique as it is based on a large sample, represents everyday clinical practice, and includes a long follow-up period spanning 6 years. Both treatments were associated with substantial improvements for patients with various disorders, with benefits persisting after the end of treatment. Change in PP mostly occurred in the 1st year of treatment, and despite involving considerably fewer sessions, these gains were largely maintained over the long term. AP patients, in contrast, continued to improve for the entire observation period. At the end of treatment, AP demonstrated an advantage in reducing the number of *DSM-IV* diagnoses and, after 6 years, showed superiority in addressing symptom severity, personality dysfunction, and interpersonal problems. This advantage was particularly notable in patients with more severe problems at baseline, highlighting the importance of AP within the German health insurance system for those severely affected.

Future research based on the DPG Practice study should include disorder-specific analyses within diagnostic subgroups to clarify potential differences in treatment response. Moreover, interactions between baseline characteristics and outcomes could be examined using multivariate models to gain a more comprehensive understanding of their influence on treatment effects. Another important direction is the investigation of mechanisms of change in these therapies, including factors such as technique, treatment duration, and session frequency. To further deepen the understanding of these

processes, mixed method approaches could systematically integrate qualitative analyses of patients' and therapists' subjective experiences. A preliminary qualitative analysis of brief patient texts collected in this study has already yielded promising insights (Henkel et al., 2016). Finally, conducting cost–benefit analyses is essential to determine whether the benefits associated with AP translate into reduced health care costs (e.g., fewer days absent from work, decreased use of psychopharmacological treatments, and reduced inpatient admissions), potentially justifying the higher direct costs.

References

Abbass, A. A., Kisely, S. R., Town, J. M., Leichsenring, F., Driessen, E., De Maat, S., Gerber, A., Dekker, J., Rabung, S., Rusalovska, S., & Crowe, E. (2014). Short-term psychodynamic psychotherapies for common mental disorders. *Cochrane Database of Systematic Reviews*, *2014*(7), Article CD004687. https://doi.org/10.1002/14651858.CD004687.pub4

Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment*, *55*(3–4), 521–536. https://doi.org/10.1207/s15327752jpa5503&4_10

Alsalti, T., Protzko, J., Lakens, D., Elson, M., & Arslan, R. C. (2024). *From Ells to Metres: Population norms should supersede sample-local standardisation*. Advance online publication. https://doi.org/10.31234/osf.io/z34hg

Angst, F., Aeschlimann, A., & Angst, J. (2017). The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *Journal of Clinical Epidemiology*, *82*, 128–136. https://doi.org/10.1016/j.jclinepi.2016.11.016

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Barkham, M., & Lambert, M. J. (2021). The efficacy and effectiveness of psychological therapies. In M. Barkham, W. Lutz, L. G. Castonguay, A. E. Bergin, & S. L. Garfield (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change: 50th anniversary edition* (7th ed., pp. 135–189). Wiley.

Bateman, A., & Fonagy, P. (2016). *Mentalization-based treatment for personality disorders*. Oxford University Press. https://doi.org/10.1093/med:psych/9780199680375.001.0001

Benecke, C., Huber, D., Schauenburg, H., & Staats, H. (2016). Wie können Langzeittherapien mit kürzeren Behandlungen verglichen werden? Designprobleme und Lösungsvorschläge am Beispiel der APS-Studie [How can long-term therapies be compared with shorter term treatment? Design problems and solution proposals exemplified by the APD study]. *Psychotherapeut*, *61*(6), 476–483. https://doi.org/10.1007/s00278-016-0140-1

Benecke, C., Tschiesner, R., Boothe, B., Frommer, J., Huber, D., Krause, R., & Staats, H. (2011). Die DPG-Praxis-Studie: Vorstellung des Studiendesigns zur Untersuchung von Langzeiteffekten psychoanalytisch begründeter Psychotherapien [The DPG practice study: Presentation of the study design for investigation of long-term effects of psychoanalytically-based psychotherapy]. *Forum der Psychoanalyse*, *27*(2), 203–218. https://doi.org/10.1007/s00451-011-0072-x

Berghout, C. C., Zevalkink, J., & Hakkaart-van Roijen, L. (2010). A cost-utility analysis of psychoanalysis versus psychoanalytic psychotherapy. *International Journal of Technology Assessment in Health Care*, *26*(1), 3–10. https://doi.org/10.1017/S0266462309990791

Blagys, M. D., & Hilsenroth, M. J. (2000). Distinctive features of short-term psychodynamic-interpersonal psychotherapy: A review of the comparative psychotherapy process literature. *Clinical Psychology: Science and Practice*, *7*(2), 167–188. https://doi.org/10.1093/clipsy.7.2.167

Brockmann, J., Schlüter, T., & Eckert, J. (2006). Langzeitwirkungen psychoanalytischer und verhaltenstherapeutischer Langzeitpsychotherapien [Long-term effects of long-term psychoanalytic and long-term behavior therapy]. *Psychotherapeut*, *51*(1), 15–25. https://doi.org/10.1007/s00278-005-0454-x

Bruijniks, S. J. E., Lemmens, L. H. J. M., Hollon, S. D., Peeters, F. P. M. L., Cuijpers, P., Arntz, A., Dingemanse, P., Willems, L., van Oppen, P., Twisk, J. W. R., van den Boogaard, M., Spijker, J., Bosmans, J., & Huibers, M. J. H. (2020). The effects of once- versus twice-weekly sessions on psychotherapy outcomes in depressed patients. *The British Journal of Psychiatry*, *216*(4), 222–230. https://doi.org/10.1192/bjp.2019.265

Bschor, T., Nagel, L., Unger, J., Schwarzer, G., & Baethge, C. (2024). Differential outcomes of placebo treatment across 9 psychiatric disorders: A systematic review and meta-analysis. *JAMA Psychiatry*, *81*(8), 757–768. https://doi.org/10.1001/jamapsychiatry.2024.0994

Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J. M., & Driessen, E. (2020). A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive-behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychotherapy Research*, *30*(2), 137–150. https://doi.org/10.1080/10503307.2018.1563312

Cuijpers, P., Turner, E. H., Koole, S. L., van Dijke, A., & Smit, F. (2014). What is the threshold for a clinically relevant effect? The case of major depressive disorders. *Depression and Anxiety*, *31*(5), 374–378. https://doi.org/10.1002/da.22249

de Maat, S., de Jonghe, F., de Kraker, R., Leichsenring, F., Abbass, A., Luyten, P., Barber, J. P., Rien, V., Dekker, J., & Barber, J. P. (2013). The current state of the empirical evidence for psychoanalysis: A meta-analytic approach. *Harvard Review of Psychiatry*, *21*(3), 107–137. https://doi.org/10.1097/HRP.0b013e318294f5fd

Derogatis, L. R. (1977). *SCL-90-R: Administration, scoring, and procedures manual*. Clinical Psychometric Research.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B., & Benjamin, L. S. (1997). *Structured clinical interview for DSM–IV axis II personality disorders (SCID–II)*. American Psychiatric Press.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2002). *Structured clinical interview for DSM–IV–TR axis I disorders (SCID–I)*. Biometrics Research, New York State Psychiatric Institute.

Franke, G. H. (2002). *SCL-90-R. Die Symptom-Checkliste von Derogatis - Deutsche Version - Manual* [SCL-90-R. Derogatis' Symptom-Checklist—German version—manual]. Beltz.

Grande, T., Dilg, R., Jakobsen, T., Keller, W., Krawietz, B., Langer, M., Oberbracht, C., Stehle, S., Stennes, M., & Rudolf, G. (2006). Differential effects of two forms of psychoanalytic therapy: Results of the Heidelberg-Berlin study. *Psychotherapy Research*, *16*(4), 470–485. https://doi.org/10.1080/10503300600608082

Grund, S., Lüdtke, O., & Robitzsch, A. (2023). Pooling methods for likelihood ratio tests in multiply imputed data sets. *Psychological Methods*, *28*(5), 1207–1221. https://doi.org/10.1037/met0000556

Gustavsson, A., Svensson, M., Jacobi, F., Allgulander, C., Alonso, J., Beghi, E., Dodel, R., Ekman, M., Faravelli, C., Fratiglioni, L., Gannon, B., Jones, D. H., Jennum, P., Jordanova, A., Jönsson, L., Karampampa, K., Knapp, M., Kobelt, G., Kurth, T., … the CDBE2010Study Group. (2011). Cost of disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, *21*(10), 718–779. https://doi.org/10.1016/j.euroneuro.2011.08.008

Harrer, M., Miguel, C., van Ballegooijen, W., Ciharova, M., Plessen, C. Y., Kuper, P., Sprenger, A. A., Buntrock, C., Papola, D., Cristea, I. A., de Ponti, N., Bašić, Đ., Pauley, D., Driessen, E., Quero, S., Grimaldos, J., Buendía, S. F., Botella, C., Hamblen, J. L., … Cuijpers, P. (2025). Effectiveness of psychotherapy: Synthesis of a "meta-analytic research domain" across world regions and 12 mental health problems. *Psychological Bulletin*, *151*(5), 600–667. https://doi.org/10.1037/bul0000465

Henkel, M., Huber, D., Ratzek, M., Benecke, C., & Klug, G. (2022). Can we differentiate between psychoanalytic and psychodynamic psychotherapy?—An empirical investigation of therapists' self-reports. *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, *68*(4), 362–377. https://doi.org/10.13109/zptm.2022.68.4.362

Henkel, M., Wiens, M., Huber, D., Staats, H., Taubner, S., Wiegand-Grefe, S., Buchholz, M. B., Frommer, J., & Benecke, C. (2016). Was berichten Patienten und Therapeuten über psychoanalytische Langzeittherapie? Analyse von „Rückblicken" in der DPG-Praxisstudie [What do therapists and patients report about long-term psychoanalytical psychotherapy? Analysis of "reviews" in the DPG practice study]. *Psychotherapeut*, *61*(6), 484–490. https://doi.org/10.1007/s00278-016-0141-0

Henkel, M., Zimmermann, J., & Benecke, C. (2018). Psychoanalytic techniques in different psychoanalytically oriented psychotherapies—An analysis of therapists' self-reports. *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, *64*(3), 237–249. https://doi.org/10.13109/zptm.2018.64.3.237

Henkel, M., Zimmermann, J., Huber, D., Staats, H., & Benecke, C. (2020). Predictors of psychoanalytic technique: Results from a naturalistic longitudinal study. *Psychoanalytic Psychology*, *37*(2), 128–135. https://doi.org/10.1037/pap0000279

Henkel, M., Zimmermann, J., Huber, D., Staats, H., Wiegand-Grefe, S., Taubner, S., Frommer, J., & Benecke, C. (2019). Patient characteristics in psychodynamic psychotherapies. *Psychoanalytic Psychology*, *36*(1), 1–8. https://doi.org/10.1037/pap0000165

Henrich, G., & Herschbach, P. (2000). Questions on Life Satisfaction (FLZM)—A short questionnaire for assessing subjective quality of life. *European Journal of Psychological Assessment*, *16*(3), 150–159. https://doi.org/10.1027//1015-5759.16.3.150

Horowitz, L. M., Strauß, B., Thomas, A., & Kordy, H. (2016). *Inventar zur Erfassung interpersonaler Probleme: Deutsche Version (IIP-D)* [Inventory of Interpersonal Problems: German version (IIP-D)] (3rd ed.). Hogrefe. https://www.testzentrale.de/shop/manual-76664.html

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huber, D., Henrich, G., Clarkin, J., & Klug, G. (2013). Psychoanalytic versus psychodynamic therapy for depression: A three-year follow-up study. *Psychiatry: Interpersonal & Biological Processes*, *76*(2), 132–149. https://doi.org/10.1521/psyc.2013.76.2.132

Huber, D., Henrich, G., Gastner, J., & Klug, G. (2012). Must all have prices? The Munich psychotherapy study. In J. S. Ablon, R. Levy, & H. Kaechele (Eds.), *Psychodynamic psychotherapy research* (pp. 51–69). Humana Press. https://doi.org/10.1007/978-1-60761-792-1_4

Huber, D., Henrich, G., & Herschbach, P. (1988). Measuring the quality of life: A comparison between physically and mentally chronically ill patients and healthy persons. *Pharmacopsychiatry*, *21*(6), 453–455. https://doi.org/10.1055/s-2007-1017050

Kahan, B. C., White, I. R., Edwards, M., & Harhay, M. O. (2023). Using modified intention-to-treat as a principal stratum estimator for failure to initiate treatment. *Clinical Trials*, *20*(3), 269–275. https://doi.org/10.1177/17407745231160074

Kaiser, T., Volkmann, C., Volkmann, A., Karyotaki, E., Cuijpers, P., & Brakemeier, E.-L. (2022). Heterogeneity of treatment effects in trials on psychotherapy of depression. *Clinical Psychology: Science and Practice*, *29*(3), 294–303. https://doi.org/10.1037/cps0000079

Keefe, J. R., Kim, T. T., DeRubeis, R. J., Streiner, D. L., Links, P. S., & McMain, S. F. (2021). Treatment selection in borderline personality disorder between dialectical behavior therapy and psychodynamic psychiatric management. *Psychological Medicine*, *51*(11), 1829–1837. https://doi.org/10.1017/S0033291720000550

Kernberg, O. F., & Caligor, E. (2005). A psychoanalytic theory of personality disorders. In M. F. Lenzenweger & J. F. Clarkin (Eds.), *Major theories of personality disorder* (2nd ed., pp. 114–156). Guilford Press.

Knekt, P., Lindfors, O., Härkänen, T., Välikoski, M., Virtala, E., Laaksonen, M. A., Marttunen, M., Kaipainen, M., Renlund, C., & the Helsinki Psychotherapy Study Group. (2008). Randomized trial on the effectiveness of long-and short-term psychodynamic psychotherapy and solution-focused therapy on psychiatric symptoms during a 3-year follow-up. *Psychological Medicine*, *38*(5), 689–703. https://doi.org/10.1017/S003329170700164X

Knekt, P., Lindfors, O., Laaksonen, M. A., Renlund, C., Haaramo, P., Härkänen, T., Virtala, E., & the Helsinki Psychotherapy Study Group. (2011). Quasi-experimental study on the effectiveness of psychoanalysis, long-term and short-term psychotherapy on psychiatric symptoms, work ability and functional capacity during a 5-year follow-up. *Journal of Affective Disorders*, *132*(1–2), 37–47. https://doi.org/10.1016/j.jad.2011.01.014

Knekt, P., Virtala, E., Härkänen, T., Vaarama, M., Lehtonen, J., & Lindfors, O. (2016). The outcome of short- and long-term psychotherapy 10 years after start of treatment. *Psychological Medicine*, *46*(6), 1175–1188. https://doi.org/10.1017/S0033291715002718

Krakau, L., Ernst, M., Hautzinger, M., Beutel, M. E., & Leuzinger-Bohleber, M. (2024). Childhood trauma and differential response to long-term psychoanalytic versus cognitive-behavioural therapy for chronic depression in adults. *The British Journal of Psychiatry*, *225*(4), 446–453. https://doi.org/10.1192/bjp.2024.112

Kramer, U., Eubanks, C. F., Bertsch, K., Herpertz, S. C., McMain, S., Mehlum, L., Renneberg, B., & Zimmermann, J. (2022). Future challenges in psychotherapy research for personality disorders. *Current Psychiatry Reports*, *24*(11), 613–622. https://doi.org/10.1007/s11920-022-01379-4

Leichsenring, F., Abbass, A., Heim, N., Keefe, J. R., Kisely, S., Luyten, P., Rabung, S., & Steinert, C. (2023). The status of psychodynamic psychotherapy as an empirically supported treatment for common mental disorders—An umbrella review based on updated criteria. *World Psychiatry*, *22*(2), 286–304. https://doi.org/10.1002/wps.21104

Leichsenring, F., Biskup, J., Kreische, R., Staats, H., Rudolf, G., & Jakobsen, T. (2005). The Göttingen study of psychoanalytic therapy: First results. *The International Journal of Psychoanalysis*, *86*(2), 433–455. https://doi.org/10.1516/XX6F-AU0W-KWM3-G6LU

Leichsenring, F., & Rabung, S. (2011). Long-term psychodynamic psychotherapy in complex mental disorders: Update of a meta-analysis. *The British Journal of Psychiatry*, *199*(1), 15–22. https://doi.org/10.1192/bjp.bp.110.082776

Leichsenring, F., Steinert, C., Rabung, S., & Ioannidis, J. P. A. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: An umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry*, *21*(1), 133–145. https://doi.org/10.1002/wps.20941

Leuzinger-Bohleber, M., Hautzinger, M., Fiedler, G., Keller, W., Bahrke, U., Kallenbach, L., Kaufhold, J., Ernst, M., Negele, A., Schoett, M., Küchenhoff, H., Günther, F., Rüger, B., & Beutel, M. (2019). Outcome of psychoanalytic and cognitive-behavioural long-term therapy with chronically depressed patients: A controlled trial with preferential and randomized allocation. *The Canadian Journal of Psychiatry*, *64*(1), 47–58. https://doi.org/10.1177/0706743718780340

Mayer, A., Zimmermann, J., Hoyer, J., Salzer, S., Wiltink, J., Leibing, E., & Leichsenring, F. (2020). Interindividual differences in treatment effects based on structural equation models with latent variables: An EffectLiteR tutorial. *Structural Equation Modeling*, *27*(5), 798–816. https://doi.org/10.1080/10705511.2019.1671196

McFarquhar, T., Luyten, P., & Fonagy, P. (2018). Changes in interpersonal problems in the psychotherapeutic treatment of depression as measured by the Inventory of Interpersonal Problems: A systematic review and meta-analysis. *Journal of Affective Disorders*, *226*, 108–123. https://doi.org/10.1016/j.jad.2017.09.036

McNeish, D., & Matta, T. (2018). Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behavior Research Methods*, *50*(4), 1398–1414. https://doi.org/10.3758/s13428-017-0976-5

Norcross, J. C., & Wampold, B. E. (2018). A new therapy for each patient: Evidence-based relationships and responsiveness. *Journal of Clinical Psychology*, *74*(11), 1889–1906. https://doi.org/10.1002/jclp.22678

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ratzek, M., Huber, D., & Klug, G. (2020). Können analytische und tiefenpsychologisch fundierte Psychotherapie reliabel voneinander getrennt werden? [Can psychoanalytic and psychodynamic psychotherapy be reliably discriminated?]. *PPmP-Psychotherapie·Psychosomatik·Medizinische Psychologie*, *70*(2), 72–79. https://doi.org/10.1055/a-0901-7829

Rosseel, Y. (2012). lavaan: An R Package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. https://doi.org/10.1002/9780470316696

Rubin, M. (2021). When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. *Synthese*, *199*(3–4), 10969–11000. https://doi.org/10.1007/s11229-021-03276-4

Rudolf, G., Manz, R., & Öri, C. (1994). Ergebnisse psychoanalytischer Therapien [Outcome of psychoanalytic treatments]. *Zeitschrift für Psychosomatische Medizin und Psychoanalyse*, *40*(1), 25–40.

Rüger, U. (2002). Tiefenpsychologisch fundierte Psychotherapie [Depth psychologically founded psychotherapy]. *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, *48*(2), 117–138. https://doi.org/10.13109/zptm.2002.48.2.117

Rüger, U., Dahm, A., Dieckmann, M., & Neher, M. (2015). *Faber/Haarstrick: Kommentar Psychotherapie-Richtlinien* [Faber/Haarstrick: Comment on the Psychotherapy Guidelines] (10th ed.). Urban & Fischer.

Sandell, R., Blomberg, J., Lazar, A., Carlsson, J., Broberg, J., & Schubert, J. (2000). Varieties of long-term outcome among patients in psychoanalysis and long-term psychotherapy. A review of findings in the Stockholm Outcome of Psychoanalysis and Psychotherapy Project (STOPP). *The International Journal of Psychoanalysis*, *81*(5), 921–942. https://doi.org/10.1516/0020757001600291

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, *76*(12), 1049–1064. https://doi.org/10.1080/10629360600810434

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Vigo, D., Thornicroft, G., & Atun, R. (2016). Estimating the true global burden of mental illness. *The Lancet Psychiatry*, *3*(2), 171–178. https://doi.org/10.1016/S2215-0366(15)00505-2

Voelkle, M. C. (2007). Latent growth curve modeling as an integrative approach to the analysis of change. *Psychology Science*, *49*(4), 375–414.

Walwyn, R., & Roberts, C. (2010). Therapist variation within randomised trials of psychotherapy: Implications for precision, internal and external validity. *Statistical Methods in Medical Research*, *19*(3), 291–315. https://doi.org/10.1177/0962280209105017

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, *8*(1), 16–37. https://doi.org/10.1037/1082-989X.8.1.16

Woll, C. F. J., & Schönbrodt, F. D. (2020). A series of meta-analytic tests of the efficacy of long-term psychoanalytic psychotherapy. *European Psychologist*, *25*(1), 51–72. https://doi.org/10.1027/1016-9040/a000385

Yeomans, F. E., Clarkin, J. F., & Kernberg, O. F. (2014). *Transference-focused psychotherapy for borderline personality disorder: A clinical guide*. American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9781615371006

Zhao, Y., & Long, Q. (2016). Multiple imputation in the presence of high-dimensional data. *Statistical Methods in Medical Research*, *25*(5), 2021–2035. https://doi.org/10.1177/0962280213511027

Zhou, Y., Matsouaka, R. A., & Thomas, L. (2020). Propensity score weighting under limited overlap and model misspecification. *Statistical Methods in Medical Research*, *29*(12), 3721–3756. https://doi.org/10.1177/0962280220940334

Zimmermann, J., Benecke, C., Hörz, S., Rentrop, M., Peham, D., Bock, A., Wallner, T., Schauenburg, H., Frommer, J., Huber, D., Clarkin, J. F., & Dammann, G. (2013). Validierung einer deutschsprachigen 16-Item-Version des Inventars der Persönlichkeitsorganisation (IPO-16) [Validity of a German 16-item version of the Inventory of Personality Organization (IPO-16)]. *Diagnostica*, *59*(1), 3–16. https://doi.org/10.1026/0012-1924/a000076

Zimmermann, J., Löffler-Stastka, H., Huber, D., Klug, G., Alhabbo, S., Bock, A., & Benecke, C. (2015). Is it all about the higher dose? Why psychoanalytic therapy is an effective treatment for major depression. *Clinical Psychology & Psychotherapy*, *22*(6), 469–487. https://doi.org/10.1002/cpp.1917

Zimmermann, J., Müller, S., Bach, B., Hutsebaut, J., Hummelen, B., & Fischer, F. (2020). A common metric for self-reported severity of personality disorder. *Psychopathology*, *53*(3–4), 168–178. https://doi.org/10.1159/000507377

Received May 15, 2025
Revision received September 12, 2025
Accepted September 16, 2025 ■