AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Consulting and Clinical Psychology

© 2026 American Psychological Association
ISSN: 0022-006X

2026, Vol. 94, No. 1, 11–25
https://doi.org/10.1037/ccp0000983

Increasing the Effectiveness of Psychotherapy in Routine Care Through Transdiagnostic Online Modules? Randomized Controlled Trial Investigating Blended Care

Carmen Schaeuffele[1], Augustin Mutak[2, 3], Solveig Behr[1], Friederike Fenski[4], Leona Hammelrath[1], Marie Puetz[4], Steffi Pohl[2], Christine Knaevelsrud[1, 5], and Johanna Boettcher[4]

[1] Department of Education and Psychology, Clinical-Psychological Intervention, Freie Universität Berlin
[2] Department of Education and Psychology, Methods and Evaluation/Quality Assurance, Freie Universität Berlin
[3] Faculty of Humanities and Social Sciences, University of Zagreb
[4] Department of Clinical Psychology and Psychotherapy, Psychologische Hochschule Berlin
[5] German Center for Mental Health (DZPG), Partner Site Berlin-Potsdam, Berlin, Germany

Objective: Blended care (BC), the integration of Internet-based interventions into psychotherapy (PT), is thought of as a promising approach to enhance PT's effectiveness and efficiency. This randomized controlled trial aimed to investigate the effectiveness as well as the implementation and usage of BC with transdiagnostic online modules compared to PT in routine care in Germany. Routine outpatient PT is delivered by licensed psychotherapists across different therapeutic orientations (cognitive behavioral therapy, psychodynamic, systemic), with variable treatment lengths and procedures. ***Method:*** Psychotherapists in routine outpatient care recruited 1,159 patients who were randomized to BC or PT. The primary outcome was self-reported mental distress (the composite of anxiety and depression); secondary outcomes included self-reported satisfaction with life, level of functioning, eating pathology, and drug and alcohol use, as well as therapist-rated severity and changes. Outcomes were measured at baseline, 6 weeks, 12 weeks, 6 months, and 12 months. We examined whether BC and PT groups changed differently over time using linear mixed models. We also investigated differences in sessions and terminations and report usage metrics of the BC platform. ***Results:*** Contrary to our hypotheses, we did not find differences between BC and PT in outcomes, including anxiety, depression, satisfaction with life, level of functioning, eating pathology, alcohol and drug use, therapist-rated severity, and satisfaction with treatment at 6 months postrandomization (all $p > .05$). BC and PT did not differ in the number of sessions or terminations. Regarding usage of the BC platform, 534 patients (91.6%) received at least one online chapter, with $M = 7.26$ ($SD = 7.01$) of a total of 39 online chapters assigned on average, and patients logged in $M = 19.73$ ($SD = 24.66$) times and spent $M = 367.14$ ($SD = 338.27$) minutes on the platform. ***Conclusions:*** In this real-world application of BC, therapists had considerable flexibility in implementing BC and integrating Internet-based interventions with sessions. Our findings suggest that the benefits observed in more structured BC setups may not fully translate to a flexible and transdiagnostic BC setup in routine care, potentially due to variations in implementation and adherence.

What is the public health significance of this article?
This study suggests that a highly flexible blend of digital components and psychotherapy may not result in increased effectiveness of psychotherapy or reduce the number of sessions in routine care.

Keywords: blended care, blended therapy, routine care, transdiagnostic, personalization

Supplemental materials: https://doi.org/10.1037/ccp0000983.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Jessica Hamblen served as action editor.

Carmen Schaeuffele https://orcid.org/0000-0001-5902-0725

Christine Knaevelsrud and Johanna Boettcher are equal contributors. The data, analysis procedure, and code for all analyses are provided on the Open Science Framework at https://doi.org/10.17605/OSF.IO/PRE87. This article represents the main outcome article of the randomized controlled trial. Secondary analyses of this study population but relating to other research questions and measures are in preparation but have not been submitted or published yet. The authors also conducted several interview studies with subsamples of patients and psychotherapists. One qualitative interview study for which the authors have interviewed psychotherapists about their blended care inclusion process to gain insights into potential inclusion or exclusion criteria for blended care has been published elsewhere. The authors have no conflicts of interest to declare.

This study was funded by the Innovation Committee (Innovationsausschuss) of the Joint Federal Committee (Gemeinsamer Bundesausschuss, Grant 01VSF19041, awarded to Johanna Boettcher, Christine Knaevelsrud, and Steffi Pohl). The funders had no role in the study design, collection, analysis, and interpretation of data; writing of the article; or the decision to submit the article for publication.

Carmen Schaeuffele played a lead role in writing–original draft and writing–

continued

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Mental disorders are highly prevalent, with comorbidity being the rule rather than the exception (Jacobi et al., 2014). Mental disorders significantly impact the quality of life, social functioning, and even physical health (Arias et al., 2022). At the same time, less than 25% of patients with a mental disorder seek professional help (Heinig et al., 2021). For those who do, the average waiting time to start psychotherapy (PT) in German outpatient care is approximately 5 months, due to the current demand for PT exceeding available treatment capacities (Bundespsychotherapeutenkammer, 2021). Even among those who access PT, treatment response remains limited: A conservative estimate of response rates in routine specialized mental health care revealed that only a fifth of the patients respond to treatment (Cuijpers et al., 2024). A recent analysis of cognitive behavioral therapy (CBT) in a routine university outpatient setting in Germany showed a higher but still moderate recovery rate of 48% (for individuals exceeding cutoff scores before therapy; Velten et al., 2025). Integrating Internet-based interventions (IBI) into routine psychotherapeutic treatment (blended care [BC]) has the potential to address both challenges: the high strain on the health care system and the limited effectiveness of PT in routine settings. Broadly, BC models can be either integrated or sequential (Baumeister et al., 2018; Erbe et al., 2017). In sequential setups, the digital component is delivered in full before or after treatment. In integrated models, digital and face-to-face components are delivered as collaborating parts, with both contributing to therapeutic progress. The aim of the integrated BC model determines how the BC model is implemented: from integrating IBIs into face-to-face therapy with the aim to flexibilize and extend PT, to intensifying face-to-face PT with the aim to increase the effectiveness, or to replacing face-to-face sessions through IBI sessions with the aim to use resources more efficiently (Baumeister et al., 2018). Accordingly, when IBIs are used during PT, they may serve as a core element of treatment or as a supplementary tool to support face-to-face sessions (Ferrao Nunes-Zlotkowski et al., 2024). Digital and face-to-face components may be alternated or be tailored flexibly on a case-to-case basis (Ferrao Nunes-Zlotkowski et al., 2024).

When IBIs assume treatment responsibilities, such as providing psychoeducation, interactive therapeutic exercises, and symptom monitoring, BC may free up psychotherapeutic resources and reduce PT sessions. BC may also intensify the therapeutic process by increasing the overall treatment dosage, improving homework compliance, and strengthening key mechanisms of change, such as self-management. A recent systematic review and meta-analysis found that BC may indeed boost the effectiveness of PT in the context of depression, albeit not for anxiety (Ferrao Nunes-Zlotkowski et al., 2024). The review also provided initial evidence that supplementary BC models—where digital components support but do not replace core therapy content—are associated with higher effect sizes than core setups, in which digital tools deliver essential therapeutic content. They conclude that BC was found to be as effective as, or in some cases more effective than, standard PT. Drawing conclusions across BC studies remains challenging, as BC setups are so heterogeneous. It seems plausible that BC settings involving a reduction of face-to-face sessions would result in smaller effects than those where the IBI supplements, rather than substitutes, standard PT. Studies that do not reduce sessions report small to moderate between-group effects favoring BC over PT alone (Berger et al., 2018; Kalde et al., 2024; Schuster et al., 2020). For instance, a recent pilot trial using a modular, integrated BC approach for patients with depression—while keeping session numbers constant across groups—suggested small advantages of BC in reducing depressive and anxiety symptoms (Kalde et al., 2024). The majority of BC research (76%, according to Ferrao Nunes-Zlotkowski et al., 2024) to date has focused on depression, with limited evidence available for broader transdiagnostic applications. A recent pilot trial in a heterogeneous outpatient sample found that supplementing PT with an emotion regulation–focused IBI yielded outcomes comparable to PT alone (Bielinski et al., 2023). Overall, while only small effects of BC over PT may be expected (and only where PT sessions are not reduced), the existing studies are statistically underpowered to reliably detect effects of this magnitude. To increase the uptake of BC in routine care, the BC setup and IBI may also need to be adapted to the specific context: The German health care system has introduced Digital Health Applications that psychotherapists can prescribe to their patients (see L. Schmidt et al., 2024, for a synthesis of the evolution within the German health care system). Despite their potential, these interventions were typically developed as stand-alone tools and are not designed for integration into ongoing PT. However, studies on facilitators and barriers highlight that IBIs for BC should enable psychotherapists to flexibly use the online materials and tailor the blended treatment to each patient's specific needs (Titzler et al., 2018). Moreover, Digital Health Applications are disorder-specific, limiting their applicability for patients with comorbid or transdiagnostic presentations, which are the norm rather than the exception in routine care. For Germany specifically, a BC intervention should also be applicable for different therapeutic approaches beyond CBT, as a large proportion of psychotherapists in routine care are not CBT-based. To address these implementation barriers, we developed a transdiagnostic, transtheoretical, and modular BC intervention called TONI (*t*herapeutic *on*line *i*ntervention) that addresses shared mechanisms and symptoms across

review and editing, a supporting role in formal analysis, and an equal role in conceptualization and project administration. Augustin Mutak played a lead role in formal analysis and visualization and an equal role in writing–review and editing. Solveig Behr played an equal role in conceptualization, data curation, investigation, project administration, writing–original draft, and writing–review and editing. Friederike Fenski played an equal role in conceptualization, investigation, project administration, writing–original draft, and writing–review and editing. Leona Hammelrath played an equal role in conceptualization, data curation, investigation, writing–original draft, and writing–review and editing. Marie Puetz played a supporting role in project administration and an equal role in visualization, writing–original draft, and writing–review and editing. Steffi Pohl played an equal role in conceptualization, funding acquisition, investigation, methodology, supervision, and writing–review and editing. Christine Knaevelsrud played a lead role in conceptualization and an equal role in funding acquisition, methodology, project administration, supervision, and writing–review and editing. Johanna Boettcher played an equal role in conceptualization, funding acquisition, methodology, project administration, supervision, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Carmen Schaeuffele, Department of Education and Psychology, Clinical-Psychological Intervention, Freie Universität Berlin, Schwendener Strße 27, 14195 Berlin, Germany. Email: carmen.schaeuffele@fu-berlin.de

the whole spectrum of mental health and applies a common language of PT (Behr et al., 2024).

This study aimed to investigate the usage, implementation, and effectiveness of BC with TONI compared to regular face-to-face outpatient PT in routine care in Germany. We expected BC with TONI to be superior in alleviating mental distress (the composite of anxiety and depression) to routine PT with a small effect at 6 months postrandomization, based on initial evidence in German routine care (Berger et al., 2018). Beyond that, we analyzed the differences between BC and PT in secondary outcomes, including level of functioning, satisfaction with life, eating pathology, alcohol and drug use, number of sessions, and therapist-rated symptom distress.

Method

Trial Design

In a pragmatic parallel-group, two-arm randomized controlled superiority trial conducted in routine outpatient care in Germany, this study aimed to assess the comparative effectiveness of BC to PT. The design is depicted in Figure 1. We reported the study procedure in detail in the study protocol (Schaeuffele et al., 2022). The study was preregistered (DRKS00028536) and approved by the ethics committee of Psychologische Hochschule Berlin (EK202121). Deidentified data and analysis code are available at https://doi.org/10.17605/OSF.IO/PRE87.

Procedure

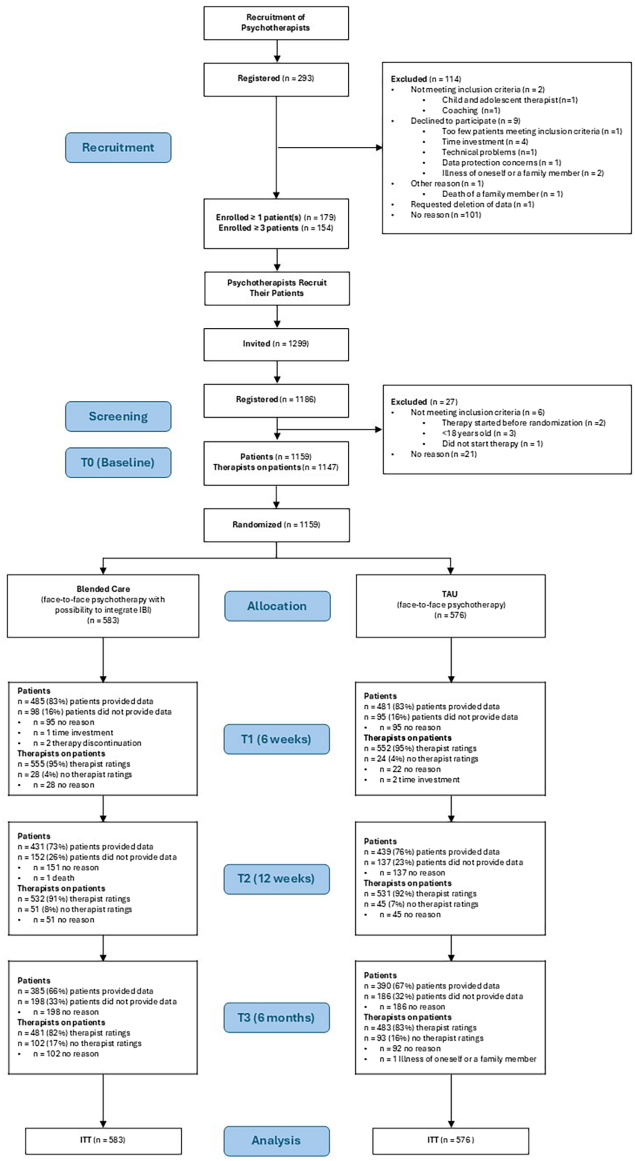
In order to investigate the research question on a broad set of the population, we included psychotherapists and patients with various backgrounds. All psychotherapists, irrespective of psychotherapeutic approach, participating in routine outpatient PT within the public health insurance system in Germany with Internet access in their practice were eligible to participate. After giving electronic informed consent, they were screened for eligibility online on the intervention platform. Once psychotherapists participated in the study, they could invite their patients to the study according to these inclusion criteria: Patients were over 18 years old, had Internet access, had sufficient German language skills, and were able to read and write. There were no exclusion criteria based on diagnoses or symptom severity due to the naturalistic nature of the trial. After receiving an invitation to participate by their therapist (printed personalized QR code or personalized invitation link via email), patients gave electronic informed consent to participate online and confirmed eligibility criteria. Patients needed to be enrolled during the initial consultation sessions with a therapist that are used to establish diagnosis and rapport before applying for coverage through their public health insurance (so-called probatory sessions). All assessments were collected online on the intervention platform. Patients and therapists were reimbursed for their study participation.

Randomization

We implemented randomization at the patient level and thus chose block randomization to ensure that psychotherapists treat patients of both groups. After patients provided their consent and baseline data, the TONI platform automatically randomly allocated patients to either BC or routine PT in a 1:1 ratio, applying a prespecified, concealed randomization algorithm with a block size of four. Due to the nature of the intervention, neither participants nor clinical personnel were blinded to study conditions.

Figure 1

Consolidated Standards of Reporting Trials Statement Flowchart



Note. IBI = Internet-based intervention; TAU = treatment as usual (psychotherapy); ITT = intention-to-treat. See the online article for the color version of this figure.

Interventions

Routine PT

All patients received routine outpatient PT. Psychotherapists were either psychologists or physicians who underwent postgraduate

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

training in PT and had a license to practice as psychotherapists. Psychotherapists could be licensed in CBT, psychodynamic (including psychoanalysis), systemic therapy, or a combination. Routine outpatient PT comprises heterogeneous treatment lengths and procedures. Due to the naturalistic and pragmatic nature of this trial, additional (mental) health care utilization was not limited.

BC (Routine PT and TONI)

In the BC group, psychotherapists could use TONI in addition to routine outpatient PT. TONI is a comprehensive digital toolkit that includes 12 transdiagnostic and transtheoretical modules (see Table 1), as well as a range of self-monitoring tools. TONI incorporates a range of behavior change techniques, including multimedia psychoeducation, interactive exercises, reflection prompts, and case examples (see exemplary screenshots in Supplemental Figure S1 for illustration). The multistage participatory intervention development involving psychotherapists of different orientations and patients with diverse symptomatology is described in detail elsewhere (Behr et al., 2024). The rationale for the selection and structure of the online modules was grounded in transdiagnostic treatment principles (Schaeuffele et al., 2021). Module topics were chosen to reflect core psychological processes known to contribute to the development and maintenance of a wide range of mental disorders, such as difficulties in emotion regulation. In addition, modules covered issues relevant across disorders (e.g., self-worth, communication), as well as select disorder-specific issues. For example, the Body and Well-Being module includes chapters on eating behavior and body image to address aspects of eating pathology, while the Addictive Substances and Behaviors module addresses substance use and behavioral addictions. Psychotherapists could individualize BC by selecting and ordering specific modules or chapters and by implementing TONI flexibly into the therapy process. Due to the naturalistic and pragmatic structure of the trial, we did not instruct psychotherapists to apply a specific implementation of TONI; that is, they could choose the frequency, duration, and format of TONI. In addition, patients could request access to modules. To allow for a high blend of sessions and the IBI, psychotherapists were able to view patients' entries in the modules' record forms. Patients were free to share additional TONI components, like symptom monitoring. Psychotherapists could enable messaging to allow for written communication with their patients through the platform. At the start of the study, all participating psychotherapists were required to watch brief instructional videos introducing the core functions of the TONI platform. In addition, they were invited to voluntarily participate in a live online workshop, where platform features were demonstrated in greater detail and strategies for implementing BC into routine practice were discussed.

Outcomes

All outcomes were assessed at baseline, 6 weeks, 12 weeks, 6 months, and 12 months postrandomization, unless otherwise specified. While all participants were invited to take part in the first four assessment points (baseline through 6 months), only a subset was invited to the 12-month follow-up due to logistical constraints. Specifically, only those who enrolled early enough for their 12-month follow-up to occur before the end of the study period were contacted for this final time point.

Table 1
Transdiagnostic Online Modules in Therapeutic Online Intervention

Module	Description of major concept and component
Development	Chapter 1: Biopsychosocial model of distress; Chapter 2: Biographical history of symptoms; Chapter 3: Genogram and intergenerational mental health
Values and goals	Chapter 1: Motivation and barriers to change; Chapter 2: Identifying personal values; Chapter 3: Therapy goals
Mindfulness	Chapter 1: Introduction to mindfulness; Chapter 2: Mindfulness-based exercises (internal and external)
Getting active	Chapter 1: Relationship between activity and mood; Chapter 2: Identifying fulfilling and enjoyable activities; Chapter 3: Implementing activities in daily life
Thoughts	Chapter 1: Relationship between thoughts and emotions; Chapter 2: Life story shaping thoughts and beliefs; Chapter 3: Mentalization and different perspectives
Understanding feelings	Chapter 1: Understanding feelings and emotional stance; Chapter 2: Messages behind feelings; Chapter 3: Life story shaping emotional experience
Dealing with feelings	Chapter 1: Tolerating emotions; Chapter 2: Avoidance/defense mechanisms and new reactions
Self-worth and strengths	Chapter 1: Concept of self-worth and personal standards; Chapter 2: Kindness and self-compassion; Chapter 3: Identifying strengths and resources
Communication	Chapter 1: Meanings of messages; Chapter 2: Communication patterns and reflection of relationships; Chapter 3: Nonviolent communication of needs and wishes
Body and well-being	Chapter 1: Relationship between body and mind; Chapter 2: Eating and body image; Chapter 3: Sleep; Chapter 4: Movement; Chapter 5: Pain and bodily complaints; Chapter 6: Sexuality; Chapter 7: Stress
Addictive substances and behaviors	Chapter 1: Core concepts of substance and nonsubstance use; Chapter 2: Pros and cons of substance use/addictive behavior; Chapter 3: Dealing with cravings; Chapter 4: Relapse prevention and dealing with relapses
Collaboration	This module is targeted at a patient's significant other Chapter 1: Significant other's own mental well-being; Chapter 2: Basics of therapy and emotional well-being; Chapter 3: Joint exercises for patients and significant others

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Primary Outcome

The primary outcome was mental distress, as the composite of self-reported anxiety (Generalized Anxiety Disorder–7 [GAD-7]) and depression (Patient Health Questionnaire–8 [PHQ-8]). The GAD-7 is a seven-item measure of general anxiety (Löwe et al., 2008; Spitzer et al., 2006). Participants rate anxiety symptoms over the past 2 weeks on a 4-point scale (0 = *not at all* to 3 = *nearly every day*). The GAD-7 is unidimensional and internally reliable (α = .89). The internal consistency in our sample at baseline was α = .85. PHQ-8 is an eight-item tool for depressive symptoms (Gräfe et al., 2004; Kroenke et al., 2009; Kroenke & Spitzer, 2002) that omits the suicidal ideation item from the Patient Health Questionnaire–9. Participants rate their emotional well-being over the past 2 weeks on a 4-point scale (0 = *not at all* to 3 = *nearly every day*). Scores range from 0 to 24, with higher scores indicating more severe depression. The PHQ-8 is internally reliable (α = .88) and valid in clinical and nonclinical settings (Wu et al., 2020). The internal consistency in our sample at baseline was α = .84. In addition to mean score changes, we examined the proportion of participants showing reliable change on the PHQ-8 and GAD-7, based on the Reliable Change Index (RCI; see the Statistical Analyses section for details).

Secondary Outcomes

Secondary self-reported outcomes included satisfaction with life (Satisfaction with Life Scale), level of functioning (Assessment of Quality of Life–8D [AQoL-8D]), eating pathology (Eating Disorder Examination Questionnaire 8 [EDE-Q8]), and alcohol/drug use (Alcohol Use Disorders Identification Test [AUDIT]/Drug Use Disorders Identification Test [DUDIT]). The Satisfaction with Life Scale measures overall life satisfaction (Diener et al., 1985; Glaesmer et al., 2011). Participants rate five items on a 7-point Likert scale (1 = *strongly disagree* to 7 = *strongly agree*). Scores range from 5 to 35, with higher scores indicating greater satisfaction. The Satisfaction with Life Scale is internally reliable (α = .92; Glaesmer et al., 2011). The internal consistency in our sample at baseline was α = .85. The AQoL-8D is a 35-item instrument assessing overall quality of life (Richardson et al., 2014). It covers eight domains: independent living, pain, senses, mental health, happiness, coping, relationships, and self-worth. The AQoL-8D is internally reliable (α = .80; Richardson et al., 2014). Internal consistency at baseline for the AQoL-8D total score in our sample was α = .93, and for the subscales, it was α = .72 for coping, α = .77 for relationships, α = .66 for independent living, α = .80 for mental health, α = .88 for pain, α = .71 for self-worth, α = .82 for happiness, and α = .44 for senses. Each item is rated on a 4- to 6-point Likert scale, depending on the item. We used the sum score for the subscales and total scale (total range: 35–175), with higher scores indicating greater impairment in quality of life. Sum scores were calculated by summing responses to items without applying the utility algorithm to allow for straightforward interpretation of raw changes over time. The EDE-Q8 is a brief, eight-item self-report measure designed to assess core eating disorder psychopathology (Kliem et al., 2016). It captures key aspects such as restraint, eating concern, shape concern, and weight concern. Items 1–5 apply a 7-point frequency scale (0–6) reflecting the number of days a behavior occurred in the past 28 days, while Items 6–8 apply a 7-point scale (0–6) reflecting severity or concern. The EDE-Q8 has demonstrated good psychometric properties. For consistency with other measures in our study, we computed a sum score for the EDE-Q8 by adding responses across the eight items (range: 0–48), rather than using the mean score as originally described by Kliem et al. (2016). Internal consistency at baseline for the EDE-Q8 in our sample was α = .92. Alcohol use was assessed using the AUDIT–Consumption, a three-item screening tool that measures frequency of drinking, typical quantity consumed, and frequency of heavy drinking episodes (Bush et al., 1998). Each item is scored on a 0–4 scale, with higher scores indicating greater alcohol consumption (total range: 0–12). Internal consistency for the three-item AUDIT at baseline was modest (α = .57). Interitem correlations ranged from −.01 to .56, indicating no to moderate associations between items. Internal consistency improved when Item 1 was excluded, with α = .72 for the two remaining items. Drug use was measured using the DUDIT–Consumption, a four-item screening tool designed to assess frequency and quantity of illicit drug use (Hildebrand, 2015). Each item is rated on a 0–4 scale, with higher scores indicating greater levels of drug use (total range: 0–16). Internal consistency for the DUDIT at baseline was modest (α = .56). Item correlations ranged from .16 to .38, indicating small to moderate associations between items. As the AUDIT and DUDIT measures were only assessed when participants indicated in a filter question that they consume alcohol or drugs, these outcomes are only reported descriptively.

We also measured satisfaction with treatment at 6 months post-treatment with an adaptation of the Client Satisfaction Questionnaire (CSQ; Attkisson & Zwick, 1982; J. Schmidt et al., 1989). Eight items relating to various aspects of patient contentment, such as overall satisfaction, perceived benefits, and willingness to recommend the treatment to others, are rated on a 4-point Likert scale (ranging from 1 to 4), with total scores being calculated as sum over the item scores (ranging from 8 to 32). The CSQ has demonstrated good reliability (α = .92). Internal consistency for the CSQ in our sample was α = .91.

Rating of Symptoms and Change

Therapists rated symptom distress with the Clinical Global Impression (CGI) Scale. The CGI Scale evaluates illness severity and improvement over time (Berk et al., 2008). It includes two components: The one-item CGI–Severity rates current severity on a 7-point scale from *normal* to *among the most extremely ill patients* (range 1–7). The one-item CGI–Improvement rates improvement or deterioration on a 7-point scale from *very much improved* to *very much worse* (range 1–7) and was accordingly not measured at baseline. In addition, patients also indicated their subjective change with an adapted version of the CGI–Change. Therapists also provided information on the number of sessions at 6 months. Since TONI was intended for use during active PT, they were also asked to indicate on the platform or during the 6-month assessment whether therapy had concluded. If so, they were prompted to indicate whether the PT ended prematurely and who initiated the termination.

Engagement With TONI

We also assessed participants' engagement with TONI based on platform-collected usage data. Engagement was operationalized through several metrics: the number of chapters assigned by therapists, the number of patient-initiated module requests, frequency and type of modules assigned, platform features used (e.g.,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

pinned pages, tracker entries, diary use), and overall platform usage (logins and time spent). Messaging activity between therapists and patients was also recorded for dyads where psychotherapists enabled the feature.

Sample Size

We hypothesized that BC would be superior to routine PT in reducing mental distress at 6 months postrandomization, with an incremental effect of $d = 0.2$ (Berger et al., 2018). A total of 400 patients per group are needed to detect this effect at 6 months postrandomization in an randomized controlled trial with a t test and α set to .05 (two-sided) with a power of 80%. The nesting of patients in therapists was taken into account by considering therapists' effects: An assumed average caseload of five new patients per year and an intraclass correlation of 0.05 (Schiefele et al., 2017) result in a design effect of 1.2 [design effect $= 1 + (nC - 1) \times$ intraclass correlations $= 1 + (5 - 1) \times 0.05 = 1.20$]. In addition, a dropout rate of 20% was assumed. This results in a sample size of at least ($400 \times$ Design Effect $\times$ Missing Values $= 400 \times 1.2 \times 1.2$) 1,152 patients (~231 psychotherapists) to detect the assumed effect of $d = 0.2$.[1]

Statistical Analyses

All analyses were performed using R 4.4.1 (R Core Team, 2013). For the primary analyses, we included data from all randomized patients (intention-to-treat analysis) and conducted the analyses blindly with respect to the intervention group label. Analyses on the usage of TONI were conducted separately once the main analyses were completed. The data, analysis procedure, and code for all analyses are provided on the Open Science Framework https://doi.org/10.17605/OSF.IO/PRE87.

First, we tested for independence of observations with regard to clustering in therapists in a mixture null model with therapist and patient random effects. The intraclass correlations (see Supplemental Table S3) for most of the outcomes were negligible (below 0.05); only for therapist-rated severity (CGI–Severity) was the clustering effect slightly higher (=0.127). As it was overall very small, we did not consider accounting for this effect in the further analyses. We multiply imputed the missing values of the primary and secondary outcomes using predictive mean matching in the mice package for R (van Buuren & Groothuis-Oudshoorn, 2011). In the performed imputation, the multilevel structure of the data was taken into account with the patient ID as the cluster variable. The variables used for the imputation can be divided into two main groups. The first group consists of questionnaires measuring clinically relevant variables. The full overview of the battery is available at the Open Science Framework. We used patients' questionnaire data for imputation, with the exception of the Working Alliance Inventory–Short Revised Questionnaire, where therapists' data were used. Regarding the item-level data imputation, we used item-level data to impute other items from the same scale or subscale, but only scale- or subscale-level data to impute items from other scales or subscales to avoid collinearity issues. Regarding the scale- or subscale-level data imputation, scale and subscale scores were passively imputed. The second group of variables used for imputation consists of certain demographic and medical characteristics of the patients and of study-related variables: patient age, gender, location, socioeconomic status, employment status, relationship status, education, distance, preferred and applied PT method, existence of previous diseases, family history of diseases, medication, therapy termination, whether the patient experiences discrimination, and whether they attended therapy. The variables were chosen on a theoretical basis, with the exception of diseases, for which we found significant correlations with the outcomes. We imputed 30 values for each missing value. We used predictive mean matching as the imputation method, both for the time point– and patient-level variables. The imputation procedure converged normally with a potential scale reduction factor <1.1. Imputation was performed using Freie Universität Berlin's High Performance Cluster (Bennett et al., 2020).

Although we randomized patients to intervention groups, we tested for group differences in all baseline variables. Variables with group differences of at least a small effect size were controlled for in the subsequent analyses. In order to evaluate the sensitivity of the results to this adjustment, we also reported the unadjusted results. For evaluating the effectiveness of blended versus routine PT, we first analyzed the data by linear multilevel regression using the lme4 package (Bates et al., 2014). We conducted a separate analysis for every primary and secondary outcome. The predictors included in the multilevel models were time point (with 0 indicating the baseline and 4 indicating the last follow-up), study group, and their interaction. With regard to time point and study group, the model was fully saturated. We also included education level as a covariate, as we found group differences in this variable between the two study groups at baseline. We fitted a random intercept model with a random intercept for every patient. The model assumes that the within-person residuals at different time points are independent and identically distributed. We assumed an unstructured covariance structure for the mixed models. The results were pooled across imputations according to the pooling guidelines in van Buuren and Groothuis-Oudshoorn (2011; Section 5.3; see also analysis code). When pooling the analyses results in mice, estimated results are averaged across imputations, and the total variance over the repeated analyses is calculated according to Rubin's (1987, p. 76) rules. To calculate the effect size after the performed analyses, we first calculated the estimated means on the pooled analyses results using the emmeans package (Lenth, 2025). The emmeans package was further utilized to perform t tests for testing mean differences, based on which the Cohen's d was computed (see also analysis code on the Open Science Framework). The t tests were performed (a) within each time point to compare the two groups and (b) within each of the two groups to compare each of the time points. For the significance tests, we applied the multivariate t-distribution correction (Kotz & Nadarajah, 2004). When calculating the statistical significance of effects from mixed models, we utilized a Kenward–Roger approximation to calculate the degrees of freedom. Following Jacobson and Truax (1991), we combined cutoff values at posttreatment with the RCI between baseline and 6 months to determine the patient's response status based on completer data. We considered patients recovered if they demonstrated clinically significant improvement by RCI > 1.96 from baseline to 6 months and their 6-month scores were below the clinical cutoff. We considered patients reliably improved if they showed improvement by RCI > 1.96 but their 6-month scores are

[1] Note that the study was preregistered with this rather conservative power analysis. In addition to the a priori power analysis, in the actual analysis we included covariates in the analysis and made use of partially observed data using multiple imputation, thus more likely achieving a higher power.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

above the clinical cutoff of 10. We considered patients reliably deteriorated if their RCI < 1.95 from baseline to 6 months. All other patients were considered unchanged. We applied internal consistencies as reported in the psychometric evaluation studies (GAD-7: $\alpha = .89$ and PHQ-8: $\alpha = .88$). We used the baseline standard deviations for the GAD-7 and PHQ-8 observed in our sample.

Transparency and Openness

We report how we determined our sample size. Deidentified data and analysis code are available at https://doi.org/10.17605/OSF.IO/PRE87, and the study procedures were detailed in a study protocol (Schaeuffele et al., 2021).

Results

The study's flow is depicted in Figure 1. Table 2 shows baseline characteristics for patients randomized to BC versus PT. Participants were mostly female, in their mid-30s, in a relationship, educated, and employed. A total of 39.54% lived in a rural area, and 32.01% were on medication. Additionally, 42.02% reported at least one health issue, and 47.89% had undergone PT in the past. The most common diagnoses were affective, anxiety, and adjustment disorders, with half the sample having at least one comorbid diagnosis. There were significant differences with at least a small effect size between BC and routine PT regarding education. We controlled for these in the analyses. To compare education's influence, we repeated the analyses without education as a predictor (see Supplemental Tables S4 and S5). There were no considerable or significant differences in baseline values of the outcomes between the two groups (Tables 3 and 4). Table 5 depicts therapist characteristics. Most therapists were licensed in CBT (82.68%), followed by psychodynamic therapy (19.55%). On average, therapists included $M = 6.23$ patients in the trial ($SD = 4.60$; range: 1–32).

Supplemental Figure S15 shows missingness rates. Due to design aspects, not all patients were invited to the 12-month post-randomization assessment, and there is a high number of missing values for the respective outcomes. As the choice of patients to be invited depended on the time, we do not expect strong systematics of missing values. Systematic missings in outcomes were considered by multiple imputation using a strong set of baseline and later outcome measures, making the assumption of missing at random rather plausible.

Intervention Effects on Primary and Secondary Outcomes

Estimated mean scores and standard errors for all primary and secondary outcome measures based on multiply imputed data are depicted in Table 3 (observed means and standard deviations are given in Supplemental Table S1). Findings from the linear mixed models and Cohen's *d* between-group effect sizes are presented in Table 4. Time was a significant predictor of all outcomes. There was no significant difference between BC and PT nor any indication that the two groups changed differently over time in primary or secondary outcomes (see Table 4). Between-group effect sizes at 6 months postrandomization were negligibly small and nonsignificant.

Within-group effect sizes over time are presented in Supplemental Table S2 and were small to moderate. Sensitivity analyses showed that not adjusting for imbalances in education had a negligible effect on the results (see Supplemental Tables S4 and S5). Similarly, repeating the analyses with completer data resulted in comparable means and effects (see Supplemental Table S6). Diagnostic plots of residuals versus fitted values, the distribution of the residuals, and the distribution of the random intercepts for the main analyses model can be found in Supplemental Figures S1–S14. For most of the outcomes, there are no deviations from normally distributed errors. Also, the assumption of normally distributed random effects is met for most of the outcomes. Deviations from a normal distribution only occur for the Pain and Independent Living subscales of the AQoL-8D and eating pathology measured by the EDE-Q8. These are mainly due to nonnormally distributed outcome variables. Residual versus fitted plots indicate slight violations to model specification for all variables; these are however only minor.

Therapy Duration and Satisfaction With Treatment

Patients receiving BC had, on average, 16.6 ($SD = 8.75$) face-to-face sessions, and patients receiving conventional PT had 16.4 ($SD = 7.03$) sessions in 6 months. There was no significant group difference in the number of sessions, $t(914.29) = 0.57$, $p = .566$. For 25.26% ($n = 147$) of patients in BC and 26.22% ($n = 151$) of patients in PT, therapists indicated that treatment was terminated—including dropouts as well as therapy completions—during the 6-month study period. Of these, in the BC group it ended earlier as planned for 67.35% ($n = 99$) of patients, and was terminated as planned for 32.65% ($n = 48$). In conventional PT, it ended earlier than planned for 59.60% ($n = 90$) of patients, and for 40.40% ($n = 61$), it ended as planned. There was no significant difference in termination rates between study groups, $\chi^2(1) = 1.61$, $p = .205$. Treatment satisfaction was high in both groups, with a mean CSQ score (maximum of 32 points) of 27.2 ($SD = 3.98$) in BC and 26.9 ($SD = 4.38$) in routine PT. There was no significant difference in overall satisfaction with treatment between groups, $t(755.73) = 0.88$, $p = .379$.

Reliable Change and Subjective Rating of Change

Table 6 presents the responder status based on the RCI and clinical cutoff criteria for the GAD-7 and PHQ-8 at the 6-month assessment following Jacobson and Truax's (1991) method. In the total sample, most participants remained unchanged on the GAD-7 (67.3%–71.6% across groups) and PHQ-8 (69.6%–76.8%), while a smaller proportion were classified as recovered (18.7%–19.5% for the GAD-7; 19.9%–20.3% for the PHQ-8) or reliably improved (3.1%–8.5%). Reliable deterioration was rare, occurring in fewer than 3% of participants. In the sample of patients who exceeded clinical cutoffs of 10 at baseline (59.10% on the GAD-7 and 68.77% on the PHQ-8), slightly higher recovery rates (24.3%–29.4%) were found compared to the total sample. Figure 2 supplements these results by illustrating subjective global change ratings at 6 months from patient and psychotherapist perspectives. The majority of ratings fell into the "much improved" or "minimally improved" categories (35.34%–43.06%), while few ratings indicated worsening.

Platform Usage

The platform asks therapists to assign modules or chapters within modules to their patients. In addition, patients can request their

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 2
Sociodemographic Characteristics of Patients at Baseline

Sociodemographic characteristic	BC (n = 583)		PT (n = 576)		Total (n = 1,159)		Statistical test of group difference		
	n	%	n	%	n	%	df	χ^{2a}	p
Gender									20[b]
Female	400	68.61	417	72.40	817	70.49			
Male	172	29.50	155	26.91	327	28.21			
Nonbinary	4	0.68	2	0.35	6	0.52			
Gender-fluid	4	0.68	0	0.00	4	0.35			
Other	1	0.17	0	0.00	1	0.09			
No answer	2	0.34	2	0.35	4	0.35			
Age									
M (SD)	35.82 (12.57)		37.43 (12.96)		36.62 (12.79)		1,154	−2.15	**.03**[c]
Range	18–89		18–83		18–89				
Relationship							1	2.38	.12
Single	205	35.16	177	30.73	382	32.69			
In a relationship	378	64.84	399	69.27	777	67.04			
Highest education									**<.01**[b]
No degree	2	0.34	2	0.35	4	0.35			1.00[b]
Nine years of education	40	6.86	25	4.34	65	5.61			.07[b]
Secondary school certificate	79	13.55	107	18.58	186	16.05			**.02**[b]
College entrance qualification	322	55.23	268	46.53	590	50.91			**<.01**[b]
Vocational education	136	23.33	162	28.13	298	25.71			.07[b]
Other	4	0.67	12	2.08	16	1.38			**.05**[b]
Employment status							5	7.61	.18
Employed	333	57.12	350	60.76	683	58.93			
Self-employed	22	3.77	36	6.25	58	5.00			
Unemployed	51	8.75	38	6.60	89	7.68			
Pensioned	24	4.12	21	3.65	45	3.88			
In education	108	18.52	91	15.80	199	17.17			
Other	45	7.72	40	6.94	85	7.33			
Place of residence							2	1.44	.49
Rural area	221	37.91	237	41.15	458	39.52			
Medium-sized city	235	40.31	225	39.06	460	39.69			
Large city	127	21.78	114	19.79	241	20.80			
Migration									
Foreign country of birth	43	7.38	31	5.38	74	6.38	1	1.61	.20
At least one parent born in foreign country	130	22.30	102	17.71	232	20.02	1	3.53	.06
Subjective socioeconomic status[d]									
M (SD)	5.30 (1.87)		5.20 (1.84)		5.25 (1.85)		1,157	0.89	.37[c]
Range	1–10		1–10		1–10				
Currently on medication	172	29.50	199	34.55	371	32.01	1	3.16	.08
Physical health issues									
Yes (at least one)[e]	245	42.02	242	42.01	487	42.02	1	4.81	1.00
Body mass index^3 $\geq$ 30	141	24.19	149	25.87	290	25.02	1	0.35	.56
Cardiac disease	26	4.46	29	5.03	55	4.75	1	0.10	.75
Asthma	74	12.69	45	7.81	119	10.27	1	6.97	**.01**
Cancer	9	1.54	13	2.25	22	1.90	1	0.45	.50
Diabetes	9	1.54	13	2.25	22	1.90	1	0.45	.50
Autoimmune disorder	49	8.40	61	10.59	110	9.49	1	1.37	.24
Prior psychotherapy	268	45.97	287	49.83	555	47.89	1	1.58	.21
Mean number of treatments (SD)[f]	1.76 (1.04)		1.87 (1.28)		1.81 (1.17)		533	−1.14	.25[c]
Treatment within the last 5 years	186	31.90	212	36.80	398	34.34	1	2.87	.09
Distance to therapist									
Mean number of minutes (SD)	23.93 (44.64)		22.28 (28.30)		23.11 (37.42)		1,157	0.75	.45[c]
Mean treatment preference[g]									
For BC (SD)	7.47 (2.40)		7.49 (2.52)		7.48 (2.46)		1,157	−0.10	.92[c]
For PT (SD)	6.20 (2.86)		6.10 (2.94)		6.15 (2.90)		1,157	0.61	.54[c]
Primary diagnosis[e]									
Mental and behavioral disorders due to substance use	8	1.37	5	0.87	13	1.12	1	0.29	.59
Schizophrenia, schizotypal and delusional disorders	3	0.51	3	0.52	6	0.52			1.00[b]
Affective disorders	311	53.34	305	52.95	616	53.15	1	0.01	.94
Anxiety disorders	134	22.98	138	23.96	272	23.47	1	0.10	.75
Obsessive–compulsive disorder	15	2.57	14	2.43	29	2.50	1	<0.01	1.00

(*table continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 2 (*continued*)

Sociodemographic characteristic	BC (n = 583)		PT (n = 576)		Total (n = 1,159)		Statistical test of group difference		
	n	%	n	%	n	%	df	χ^{2a}	p
Acute stress reaction and posttraumatic stress disorder	55	9.43	33	5.73	88	7.59	1	5.15	**.02**
Adjustment disorder	93	15.95	98	17.01	191	16.48	1	0.17	.68
Somatoform disorders	30	5.15	31	5.38	61	5.26	1	0.02	.96
Eating disorders	24	4.12	17	2.95	41	3.54	1	0.84	.36
Personality disorders	26	4.46	30	5.21	56	4.83	1	0.21	.64
Hyperkinetic and attention-deficit disorders	15	2.57	12	2.08	27	2.33	1	0.13	.72
Other disorders[h]	21	3.64	25	4.39	46	4.01	1	0.24	.62
Missing diagnosis	8	1.37	6	1.04	14	1.21	1	0.06	.81
At least one comorbid diagnosis	333	57.12	307	53.30	640	55.22	1	1.56	.21
Family history of similar mental health issues	301	51.63	290	50.35	591	50.99	1	0.14	.71

Note. Significant values appear in bold. BC = blended care; PT = psychotherapy.
[a] For samples with $n < 40$, Yates' continuity correction was applied. [b] Fisher's exact test. [c] t test for independent samples. [d] Based on the German version of the MacArthur Scale. [e] Multiple answers possible. [f] Unplausible values were removed. [g] On a scale from 0 (*not at all*) to 10 (*very much*), patients were asked how much both treatment options (BC vs. PT) appealed to them. [h] Dissociative disorders, other and unspecified reactions to severe stress, other neurotic disorders, sexual dysfunction, enduring personality change after catastrophic experience, habit and impulse disorders, pervasive developmental disorders, emotional disorders with onset specific to childhood, and nonorganic sleep disorders.

therapist to assign a module. Therapists assigned patients, on average, $M = 7.26$ ($SD = 7.01$) out of 39 online chapters in TONI. A total of 534 patients (91.6%) received at least one chapter, while 273 (46.83%) patients requested at least one module. They requested $M = 1.44$ ($SD = 2.3$) modules on average. Supplemental Figure S16 shows frequencies of assigned online modules (operationalized as receiving at least one chapter). The most frequently assigned modules were "Development" and "Values and Goals," with 227 patients (38.94%) assigned respectively. Therapists assigned the Collaboration module least frequently (33 patients, 5.66%). $N = 174$ patients (29.85%) pinned at least one page, with $M = 1.36$ ($SD = 4.5$) of pinned pages. $N = 152$ patients (26.07%) had at least one entry in any tracker. Usage of trackers varied widely ($M = 11.23$, $SD = 85.11$, with a range from 0 to 1907). $N = 73$ patients (12.52%) completed at least one diary entry. Regarding overall platform usage, patients logged in $M = 19.73$ ($SD = 24.66$) times and spent $M = 367.14$ ($SD = 338.27$) minutes on the platform. A total of 117 therapists enabled messaging. Patients who had messaging enabled sent $M = 0.42$ ($SD = 1.2$) messages to their therapist and received $M = 0.94$ ($SD = 1.56$) messages from their therapist. Supplemental Figure S17 depicts histograms of key platform engagement metrics, highlighting the heterogeneity of engagement.

Discussion

This study investigated the effects of BC in a naturalistic setting and compared its effects to routine PT. Contrary to our hypothesis that BC will improve the effectiveness of PT, we did not find differences between BC and PT in outcomes, including anxiety, depression, satisfaction with life, level of functioning, eating pathology, drug use, alcohol use, therapist-rated severity, and satisfaction with treatment. We also did not find a reduced number of therapeutic sessions or differences in therapy termination.

We hypothesized that integrating TONI, a transdiagnostic and transtheoretical modular digital intervention, would enhance the effects of PT in routine outpatient care. This assumption was grounded on (a) evidence suggesting that BC can improve PT effectiveness, at least in the context of depression (Erbe et al., 2017; Ferrao Nunes-Zlotkowski et al., 2024), (b) the ability for therapists to outsource certain materials, such as psychoeducation, to TONI while adapting face-to-face sessions flexibly to patient needs, which may result in (c) a potentially higher therapy dosage, and (d) improved patient empowerment.

Although our trial was robustly powered to detect small effects, we found no evidence for the superiority of BC over PT. The lack of observed effects may be interpreted in light of the high degree of flexibility in the implementation of BC, the limited engagement with the TONI platform, and the relatively low initial symptom severity within the patient population.

Most BC setups are standardized without many possibilities for patients and therapists to individualize the IBI/BC setup (Ferrao Nunes-Zlotkowski et al., 2024). With TONI, we encouraged a personalized and flexible use of online modules. While this autonomy aligns with the structure of German routine care and TONI's goal of addressing a wide range of mental health concerns and therapy trajectories, it may have introduced complexity that was overwhelming for some users. Additionally, the implementation of TONI was ultimately guided by the psychotherapists themselves, who may have served as a bottleneck in determining which modules to select and how BC was delivered in practice. We did not instruct psychotherapists to prescribe a minimum number of modules. Thus, we did not define the *intended use* for patients or therapists. Our goal was to explore the real-world uptake of BC in this setting. In modular interventions, a standard for adherence has not been established as total modules may differ between patients. In a secondary analysis of this trial, we compared different operationalizations of adherence identified through a scoping review (Eicher et al., 2025). In line with the general IBI literature (e.g., Moshe et al., 2021)—and regardless of the dropout operationalization applied—treatment completion was associated with better outcomes in our secondary analysis (Eicher et al., 2025). Thus, fostering more consistent engagement and adherence to the intervention may have increased BC's effects. Combining the flexibility of personalized module selection with a more structured, linear approach, featuring deadlines, reminders, and a clearer timeline, may enhance accountability and improve engagement. Upcoming findings from a

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 3

Estimated Means and SE for Blended Care and Psychotherapy by Measurement Points Based on Multiply Imputed Data

Outcome	Group	Baseline		6 weeks		12 weeks		6 months		12 months	
		M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Mental distress (GAD-7 + PHQ-8)	BC	23.63	0.63	19.72	0.63	18.45	0.64	17.22	0.64	17.41	0.68
Mental distress (GAD-7 + PHQ-8)	PT	23.71	0.62	19.57	0.63	18.19	0.62	16.82	0.64	17.02	0.66
Anxiety (GAD-7)	BC	10.85	0.33	9.03	0.34	8.47	0.33	7.87	0.35	7.95	0.36
Anxiety (GAD-7)	PT	10.91	0.33	9.07	0.33	8.42	0.33	7.73	0.33	7.8	0.36
Depression (PHQ-8)	BC	12.79	0.36	10.69	0.36	9.98	0.37	9.35	0.37	9.45	0.39
Depression (PHQ-8)	PT	12.8	0.35	10.5	0.36	9.77	0.36	9.1	0.37	9.22	0.37
Satisfaction with life (SWLS)	BC	18.37	0.47	18.42	0.48	19.61	0.48	19.65	0.49	19.6	0.51
Satisfaction with life (SWLS)	PT	18.65	0.46	18.57	0.47	19.92	0.47	20.05	0.47	19.92	0.49
Functioning (AQoL-8D)	BC	92.07	1.33	89.28	1.34	87.41	1.34	85.65	1.34	86.35	1.35
Functioning (AQoL-8D)	PT	92.08	1.32	89.16	1.32	86.3	1.33	84.46	1.34	85.34	1.35
Happiness subscale (AQoL-8D)	BC	11.66	0.21	11.43	0.21	11.18	0.21	10.93	0.21	11.03	0.23
Happiness subscale (AQoL-8D)	PT	11.57	0.2	11.41	0.21	11.08	0.21	10.81	0.21	10.9	0.21
Mental Health subscale (AQoL-8D)	BC	24.66	0.37	23.49	0.38	22.88	0.38	22.39	0.39	22.52	0.39
Mental Health subscale (AQoL-8D)	PT	24.8	0.37	23.67	0.37	22.7	0.36	22.12	0.38	22.34	0.39
Pain subscale (AQoL-8D)	BC	5.76	0.2	5.63	0.2	5.69	0.2	5.6	0.2	5.67	0.22
Pain subscale (AQoL-8D)	PT	5.92	0.2	5.73	0.2	5.65	0.2	5.6	0.2	5.56	0.22
Relationship subscale (AQoL-8D)	BC	17.47	0.35	17.17	0.36	16.88	0.36	16.67	0.36	16.67	0.37
Relationship subscale (AQoL-8D)	PT	17.42	0.34	16.99	0.34	16.58	0.35	16.36	0.35	16.62	0.35
Self-Worth subscale (AQoL-8D)	BC	9.19	0.19	8.86	0.19	8.62	0.19	8.31	0.2	8.42	0.21
Self-Worth subscale (AQoL-8D)	PT	9.26	0.19	8.85	0.19	8.49	0.19	8.21	0.19	8.34	0.19
Senses subscale (AQoL-8D)	BC	5.62	0.12	5.56	0.12	5.51	0.12	5.48	0.12	5.56	0.13
Senses subscale (AQoL-8D)	PT	5.61	0.12	5.61	0.12	5.54	0.12	5.47	0.13	5.48	0.13
Eating pathology (EDE-Q8)	BC	17.13	0.96					16.47	1	16.14	1.18
Eating pathology (EDE-Q8)	PT	17.32	0.97					16.54	1.01	16.32	0.99
Alcohol use (AUDIT)	BC	3.70	0.10					3.27	0.12	3.26	0.25
Alcohol use (AUDIT)	PT	3.44	0.10					3.35	0.13	3.26	0.21
Drug use (DUDIT-C)	BC	3.54	0.26					3.65	0.28	3.82	0.78
Drug use (DUDIT-C)	PT	3.44	0.23					4.03	0.43	4.86	0.74
Therapist-rated severity (CGI-S)	BC	4.62	0.09	4.4	0.09	4.21	0.1	3.99	0.1	3.96	0.11
Therapist-rated severity (CGI-S)	PT	4.62	0.09	4.41	0.09	4.27	0.09	4.02	0.1	4.02	0.1

Note. Means and *SE*s as predicted by the linear mixed models, based on the imputed data. EDE-Q8, AUDIT, and Drug Use Disorders Identification Test (DUDIT) were not measured at 6 and 12 weeks. Results are averaged over the levels of education. For AUDIT and DUDIT, the observed means are presented. Their data were not imputed or analyzed with linear mixed models as they were only collected after patients indicated on a filter question that they consume alcohol or drugs. *SE* = standard error; PT = psychotherapy; GAD-7 = Generalized Anxiety Disorder–7; PHQ-8 = Patient Health Questionnaire–8; BC = blended care; SWLS = Satisfaction with Life Scale; AQoL-8D = Assessment of Quality of Life–8D; EDE-Q8 = Eating Disorder Examination Questionnaire 8; AUDIT = Alcohol Use Disorders Identification Test; DUDIT-C = Drug Use Disorders Identification Test–Consumption; CGI-S = Clinical Global Impression–Severity.

comparison of fixed and flexible BC formats in German routine care will help clarify whether greater structure leads to better outcomes (Baumeister et al., 2021). Usage data in our trial showed that the first three modules (Values and Goals, Development, and Mindfulness) were assigned most often. This could reflect a particularly good fit for this content and the digital format as well as a broad applicability across patients and therapeutic schools (Behr et al., 2024). It could, however, also reflect an inability or reluctance of psychotherapists to make individual choices for their patients, with the first three modules being a safe and convenient choice for all their patients. In BC, as an emerging field, nearly all therapists are novices with no to very limited experience in how to integrate digital tools in their PT. Psychotherapists repeatedly identified this lack of knowledge and confidence as a barrier to implementing BC in qualitative studies (Dech et al., 2022). Although targeted training might improve therapists' familiarity and confidence and improve module activation, implementation research suggests that training alone is unlikely to be sufficient (Lyon & Koerner, 2016). More hands-on experience, ongoing support, and structural changes, such as integration into workflows and organizational incentives, may also be necessary to meaningfully change practice. The high caseload in the routine health care system and the limited scope of the study (on average, therapists enrolled six patients in the study, which were then randomized to the two groups, ultimately resulting in an average of three patients treated with BC) may have also discouraged transformative changes to their practice. While qualification and confidence in using digital tools will likely increase in the future, it still seems plausible that not all therapists want to actively engage with online materials. Future studies may also focus on investigating the rationale and motivation of therapists to include online material in their therapy.

In our sample, 31%–41% of participants scored below the clinical cutoffs for anxiety or depression at baseline, suggesting that a substantial portion of those receiving PT may present with subclinical symptomatology. The low initial severity may contribute to the lower response rates observed in our trial compared to those typically reported for PT. When only considering patients who had clinical symptoms at baseline, response rates did improve but remained below what is reported in the literature (see, e.g., Velten et al., 2025, for the effectiveness of CBT in mixed diagnoses in

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 4

Results of the Linear Mixed Models and Between-Group Effect Sizes (Cohen's d) Based on Multiply Imputed Data

						Baseline		6 weeks		12 weeks		6 months		12 months	
Outcome	Time	Group	Education	Time × Group	Contrast	*d*	[95% CI]	*d*	[95% CI]	*d*	[95% CI]	*d*	[95% CI]	*d*	[95% CI]
Mental distress (GAD-7 + PHQ-8)	$F(4, 4624) = 225.94$, $<.001$	$F(1,1151) = 0.43$, $p = 0.511$	$F(5,1151) = 3.33$, $p = 0.005$	$F(4, 4624) = 0.28$, $p = .894$	BC—PT	-0.01	[−0.07, 0.06]	-0.01	[−0.06, 0.07]	0.02	[−0.05, 0.09]	0.03	[−0.04, 0.1]	0.02	[−0.04, 0.09]
Anxiety (GAD-7)	$F(4, 4624) = 146.67$, $<.001$	$F(1,1151) = 0.07$, $p = 0.796$	$F(5,1151) = 1.86$, $p = 0.099$	$F(4, 4624) = 0.2$, $p = .941$	BC—PT	-0.01	[−0.07, 0.06]	-0.01	[−0.07, 0.06]	0.01	[−0.06, 0.07]	0.02	[−0.05, 0.08]	0.02	[−0.05, 0.08]
Depression (PHQ-8)	$F(4, 4624) = 190.83$, $<.001$	$F(1,1151) = 0.85$, $p = 0.356$	$F(5,1151) = 4.04$, $p = 0.001$	$F(4, 4624) = 0.28$, $p = .891$	BC—PT	0	[−0.07, 0.07]	0	[−0.04, 0.09]	0.03	[−0.04, 0.1]	0.03	[−0.04, 0.1]	0.03	[−0.04, 0.1]
Satisfaction with life (SWLS)	$F(4, 4624) = 29.62$, $<.001$	$F(1,1151) = 1.22$, $p = 0.269$	$F(5,1151) = 3.69$, $p = 0.003$	$F(4, 4624) = 0.14$, $p = .967$	BC—PT	-0.03	[−0.11, 0.04]	-0.03	[−0.09, 0.06]	-0.03	[−0.11, 0.04]	-0.04	[−0.12, 0.03]	-0.03	[−0.1, 0.04]
Functioning (AQoL-8D)	$F(4, 4624) = 82.17$, $<.001$	$F(1,1151) = 0.87$, $p = 0.352$	$F(5,1151) = 7.39$, $< .001$	$F(4, 4624) = 0.86$, $p = .49$	BC—PT	0	[−0.08, 0.08]	0	[−0.07, 0.08]	0.05	[−0.03, 0.12]	0.05	[−0.03, 0.13]	0.04	[−0.04, 0.12]
Coping subscale (AQoL-8D)	$F(4, 4624) = 52.41$, $<.001$	$F(1,1151) = 3.49$, $p = 0.062$	$F(5,1151) = 2.07$, $p = 0.067$	$F(4, 4624) = 0.52$, $p = .722$	BC—PT	0.03	[−0.04, 0.09]	0.03	[−0.02, 0.12]	0.07	[0, 0.14]	0.03	[−0.04, 0.1]	0.06	[−0.01, 0.12]
Happiness subscale (AQoL-8D)	$F(4, 4624) = 25.79$, $<.001$	$F(1,1151) = 0.7$, $p = 0.401$	$F(5,1151) = 2.54$, $p = 0.027$	$F(4, 4624) = 0.13$, $p = .972$	BC—PT	0.02	[−0.05, 0.09]	0.02	[−0.06, 0.07]	0.02	[−0.05, 0.09]	0.03	[−0.04, 0.1]	0.03	[−0.04, 0.1]
Mental Health subscale (AQoL-8D)	$F(4, 4624) = 94.61$, $<.001$	$F(1,1151) = 0.09$, $p = 0.760$	$F(5,1151) = 2.95$, $p = 0.012$	$F(4, 4624) = 0.93$, $p = .446$	BC—PT	-0.02	[−0.09, 0.05]	-0.02	[−0.09, 0.05]	0.02	[−0.05, 0.09]	0.03	[−0.04, 0.1]	0.02	[−0.05, 0.09]
Pain subscale (AQoL-8D)	$F(4, 4624) = 2.78$, $p = .0252$	$F(1,1151) = 0.03$, $p = 0.867$	$F(5,1151) = 14.08$, $< .001$	$F(4, 4624) = 0.77$, $p = .545$	BC—PT	-0.04	[−0.11, 0.03]	-0.04	[−0.1, 0.05]	0.01	[−0.06, 0.08]	0	[−0.07, 0.07]	0.02	[−0.05, 0.1]
Relationship subscale (AQoL-8D)	$F(4, 4624) = 15.75$, $<.001$	$F(1,1151) = 0.91$, $p = 0.339$	$F(5,1151) = 3.71$, $p = 0.002$	$F(4, 4624) = 0.5$, $p = .735$	BC—PT	0.01	[−0.06, 0.08]	0.01	[−0.05, 0.1]	0.04	[−0.03, 0.12]	0.05	[−0.03, 0.12]	0.01	[−0.07, 0.08]
Self-Worth subscale (AQoL-8D)	$F(4, 4624) = 58.48$, $<.001$	$F(1,1151) = 0.26$, $p = 0.610$	$F(5,1151) = 1.25$, $p = 0.283$	$F(4, 4624) = 0.65$, $p = .63$	BC—PT	-0.02	[−0.09, 0.05]	-0.02	[−0.07, 0.07]	0.03	[−0.04, 0.11]	0.03	[−0.04, 0.1]	0.02	[−0.05, 0.09]
Senses subscale (AQoL-8D)	$F(4, 4624) = 2.38$, $p = .0497$	$F(1,1151) = 0$, $p = 0.965$	$F(5,1151) = 6.49$, $< .001$	$F(4, 4624) = 0.4$, $p = .807$	BC—PT	0	[−0.06, 0.07]	0	[−0.09, 0.05]	-0.01	[−0.08, 0.06]	0	[−0.06, 0.07]	0.03	[−0.04, 0.09]
Eating pathology (EDE-Q8)	$F(4, 4624) = 3.44$, $p = .008$	$F(1,1151) = 0.03$, $p = 0.854$	$F(5,1151) = 4.13$, $p = 0.001$	$F(4, 4624) = 0.01$, $p = 1$	BC—PT	-0.01	[−0.07, 0.05]					0	[−0.07, 0.06]	-0.01	[−0.07, 0.06]
Therapist-rated severity (CGI-S)	$F(4, 4624) = 74.91$, $<.001$	$F(1,1151) = 0.41$, $p = 0.520$	$F(5,1151) = 3.71$, $p = 0.002$	$F(4, 4624) = 0.23$, $p = .922$	BC—PT	0	[−0.07, 0.07]	0	[−0.07, 0.07]	-0.03	[−0.1, 0.04]	-0.02	[−0.08, 0.05]	-0.02	[−0.09, 0.04]

Note. EDE-Q8 was not collected at 6 and 12 weeks. CI = confidence interval; PT = psychotherapy; GAD-7 = Generalized Anxiety Disorder–7; PHQ-8 = Patient Health Questionnaire–8; BC = blended care; SWLS = Satisfaction with Life Scale; AQoL-8D = Assessment of Quality of Life–8D; EDE-Q8 = Eating Disorder Examination Questionnaire 8; CGI-S = Clinical Global Impression–Severity.

Table 5
Sociodemographic Characteristics of Therapists at Baseline

Demographic variable	n = 178 (%)
Gender	
Female	138 (77.53)
Male	38 (21.35)
No answer	2 (1.12)
Age[a]	
M (*SD*)	45.54 (9.42)
Range	27–74
Years of professional experience	
M (*SD*)	12.03 (8.16)
Range	1–44
Licensed in[b]	
Cognitive-behavioral therapy	148 (83.15)
Psychodynamic therapy	35 (19.66)
Psychanalytic therapy	6 (3.37)
Systemic therapy	7 (3.93)
More than one approach	11 (6.18)
Application of[b]	
Cognitive-behavioral therapy	147 (82.58)
Psychodynamic therapy	37 (20.79)
Psychoanalytic therapy	5 (2.81)
Systemic therapy	22 (12.36)
More than one approach	30 (16.85)
Location of practice	
Rural area	60 (33.71)
Medium-sized city	61 (34.27)
Large city	57 (32.02)

Note. Demographic characteristics are available for n = 178 therapists. N = 179 therapists recruited patients and are included in later analysis.
[a] Unplausible values were removed. [b] Multiple answers possible.

university outpatient centers across Germany). In their subjective perspective, the majority of patients and therapists indicated that patients were very or very much improved at 6 months post-randomization. Future research should explore factors that may influence treatment outcomes, including potential moderators that could account for variability in effects. While the present study focused on overall group differences, understanding for whom and under what conditions treatments are most effective is an important question for further investigation.

In more structured setups, BC may involve alternating sessions with online components and thus reduce the number of sessions while maintaining effectiveness (Kemmeren et al., 2023). Although session reduction was not explicitly targeted, we investigated whether BC would incidentally impact the number of sessions delivered and did not find any indication that BC reduces the number of sessions. In routine care in Germany, therapists have little incentive to reduce the number of sessions, as their compensation is session-based. Our findings suggest that BC, when applied with high flexibility and without structural incentives, does not leverage its potential for reducing the use of psychotherapeutic resources.

Strengths and Limitations

We developed and tested a flexible, modular, and transtheoretical BC tool that allows psychotherapists of all orientations to integrate digital components across the whole spectrum of mental health. Main strengths of the trial are the naturalistic setting and the transdiagnostic application and evaluation. As such, it offers evidence for implementation in practice for a broad clientele. Other strengths include the trial's sufficient power to detect small differences, statistical analyses based on masked data, and the use of self- as well as therapist-rated measures. The integration of all therapeutic orientations as well as all diagnoses mirrors routine practice and increases external validity. As we consider individual therapies, spillover effects are less likely, and the "no interferences" assumption of the stable-unit-treatment-assumption is likely met. Randomization to treatment allowed for a causal interpretation of the results. We found some small imbalances between groups in baseline demographic variables, which we investigated in sensitivity analyses. To deal with systematic dropout, we collected a large number of possible variables that may explain the missing data mechanism and controlled for it via multiple imputation. However, there are still limitations to the generalizability of our findings: Despite extensive efforts to recruit therapists from diverse orientations, most therapists were CBT therapists. The largest proportion of patients was diagnosed by their therapist with an affective disorder, followed by anxiety and adjustment disorders. The study team did not conduct additional clinical interviews to corroborate diagnoses, and the diagnostic status at 6-month randomization is unknown. Psychotherapists were responsible for recruiting their patients, which might have constrained the scope of patient recruitment. While qualitative interviews with psychotherapists in

Table 6
Reliable Change for GAD-7 and PHQ-8 by Group and Clinical Status at Baseline

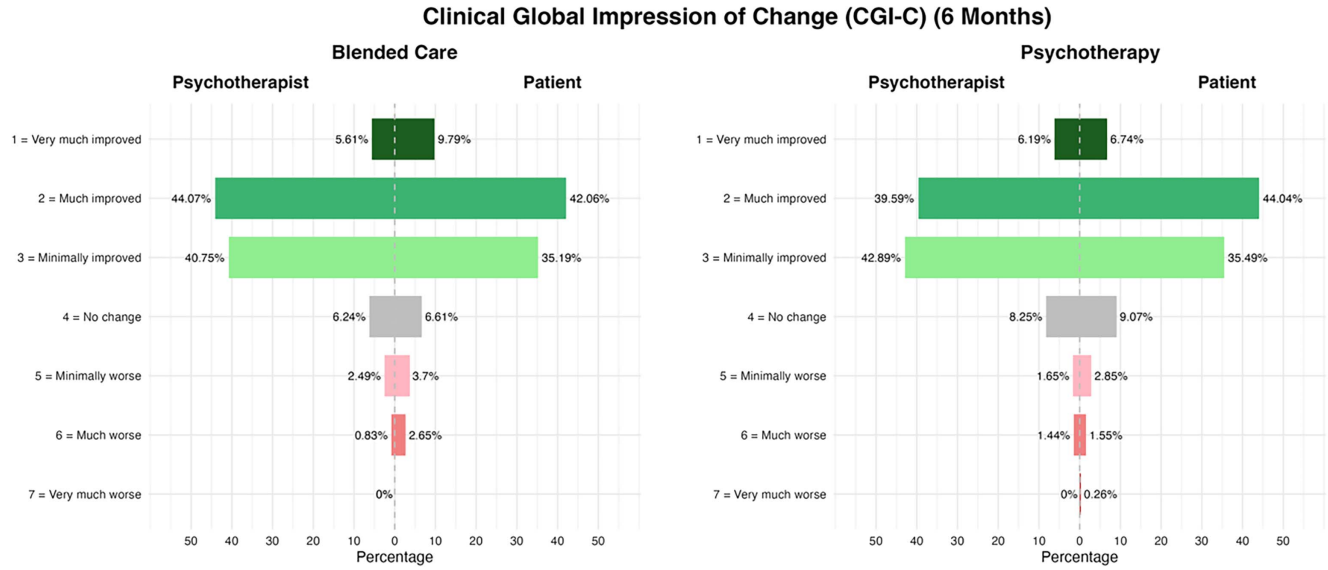
		Total sample				Sample above clinical cutoff			
		BC		PT		BC		PT	
Outcome	Status	n	%	n	%	n	%	n	%
GAD-7	Recovered	114	18.7	122	19.5	98	28.4	100	29.4
	Reliably improved	19	3.1	23	3.7	19	5.5	23	6.8
	Reliably deteriorated	13	2.1	10	1.6	3	0.9	3	0.9
	Unchanged	437	71.6	421	67.3	225	65.2	214	62.9
PHQ-8	Recovered	112	20.3	110	19.9	103	26.1	98	24.3
	Reliably improved	47	8.5	39	7.1	47	11.9	39	9.7
	Reliably deteriorated	9	1.6	7	1.3	0	0.0	3	0.7
	Unchanged	415	69.6	420	76.8	244	61.9	263	65.3

Note. Responder status is based on completer data. GAD-7 = Generalized Anxiety Disorder–7; PHQ-8 = Patient Health Questionnaire–8; BC = blended care; PT = psychotherapy.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 2

Clinical Global Impression of Change (Psychotherapist and Patient Rating) at 6 Months Postrandomization



Note. The figure shows the Clinical Global Impression of Change (CGI-C) ratings at the 6-month assessment by psychotherapists and patients. The figure is based on completer data. See the online article for the color version of this figure.

our study indicated that psychotherapists did not restrict patient inclusion into the trial (Jordan et al., 2025), this approach may still limit generalizability. We did not instruct therapists or patients to use a certain number of modules but rather provided the option to use them. While our decision is in line with the study goal to examine how the online tool is used under "real-life conditions," it makes it more difficult to interpret the null findings: Did we not find differences because the tool was not sufficiently used? Or was usage adequate and just not associated with an increase in effectiveness? As stated above, future studies should examine more structured versions of BC and investigate how therapists decide when and how to integrate online tools.

Conclusion

BC holds the potential to increase therapy effectiveness and reduce therapist burden. Our findings suggest that the benefits observed in more structured BC setups may not fully translate to a flexible and transdiagnostic BC setup in routine care, potentially due to variations in implementation and adherence.

References

Arias, D., Saxena, S., & Verguet, S. (2022). Quantifying the global burden of mental disorders and their economic value. *EClinicalMedicine*, *54*, Article 101675. https://doi.org/10.1016/j.eclinm.2022.101675

Attkisson, C. C., & Zwick, R. (1982). The Client Satisfaction Questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning*, *5*(3), 233–237. https://doi.org/10.1016/0149-7189(82)90074-X

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. arXiv:1406.5823 [Stat]. https://arxiv.org/abs/1406.5823

Baumeister, H., Bauereiss, N., Zarski, A.-C., Braun, L., Buntrock, C., Hoherz, C., Idrees, A. R., Kraft, R., Meyer, P., Nguyen, T. B. D., Pryss, R., Reichert, M., Sextl, T., Steinhoff, M., Stenzel, L., Steubl, L., Terhorst, Y., Titzler, I., & Ebert, D. D. (2021). Clinical and cost-effectiveness of PSYCHOnlineTHERAPY: Study protocol of a multicenter blended outpatient psychotherapy cluster randomized controlled trial for patients with depressive and anxiety disorders. *Frontiers in Psychiatry*, *12*, Article 660534. https://doi.org/10.3389/fpsyt.2021.660534

Baumeister, H., Grässle, C., Ebert, D. D., & Krämer, L. V. (2018). Blended Psychotherapy—Verzahnte Psychotherapie: Das Beste aus zwei Welten? *PiD—Psychotherapie im Dialog*, *19*(04), 33–38. https://doi.org/10.1055/a-0592-0264

Behr, S., Fenski, F., Boettcher, J., Knaevelsrud, C., Hammelrath, L., Kovacs, G., Schirmer, W., Petrick, H., Becker, P., & Schaeuffele, C. (2024). TONI—One for all? Participatory development of a transtheoretic and transdiagnostic online intervention for blended care. *Internet Interventions*, *35*, Article 100723. https://doi.org/10.1016/j.invent.2024.100723

Bennett, L., Melchers, B., & Proppe, B. (2020). *Curta: A general-purpose high-performance computer at ZEDAT, Freie Universität Berlin*. Freie Universität Berlin. https://doi.org/10.17169/REFUBIUM-26754

Berger, T., Krieger, T., Sude, K., Meyer, B., & Maercker, A. (2018). Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: Results of a pragmatic randomized controlled trial. *Journal of Affective Disorders*, *227*, 455–462. https://doi.org/10.1016/j.jad.2017.11.021

Berk, M., Ng, F., Dodd, S., Callaly, T., Campbell, S., Bernardo, M., & Trauer, T. (2008). The validity of the CGI severity and improvement scales as measures of clinical effectiveness suitable for routine clinical use. *Journal of Evaluation in Clinical Practice*, *14*(6), 979–983. https://doi.org/10.1111/j.1365-2753.2007.00921.x

Bielinski, L. L., Krieger, T., Kley, M. A., Moggi, F., & Berger, T. (2023). Blending an internet-based emotion regulation intervention with face-to-face psychotherapy: Findings from a pilot randomized controlled trial. *Internet Interventions*, *33*, Article 100650. https://doi.org/10.1016/j.invent.2023.100650

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Bundespsychotherapeutenkammer. (2021). *BPtK-Auswertung: Monatelange Wartezeiten bei Psychotherapeut*innen*. https://www.bptk.de/wp-content/uploads/2021/03/20210329_pm_bptk_monatelange-Wartezeiten.pdf

Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D., & Bradley, K. A. (1998). The AUDIT alcohol consumption questions (AUDIT-C): An effective brief screening test for problem drinking. *Archives of Internal Medicine*, *158*(16), 1789–1795. https://doi.org/10.1001/archinte.158.16.1789

Cuijpers, P., Miguel, C., Harrer, M., Ciharova, M., & Karyotaki, E. (2024). The outcomes of mental health care for depression over time: A meta-regression analysis of response rates in usual care. *Journal of Affective Disorders*, *358*, 89–96. https://doi.org/10.1016/j.jad.2024.05.019

Dech, M., Klawohn, J., & Romanczuk-Seiferth, N. (2022). Das Beste aus zwei Welten: Eine systematische Übersicht zu Faktoren der Implementierung einer "Blended Therapy" (Gemischte Therapie) in der Psychotherapeutischen Routineversorgung [The best of both worlds: A systematic overview of implementation factors of blended therapy (mixed therapy) in psychotherapeutic routine care]. *Verhaltenstherapie*, *32*(4), 153–164. https://doi.org/10.1159/000524332

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Eicher, S., Fenski, F., Behr, S., Hammelrath, L., Boettcher, J., Schaeuffele, C., & Knaevelsrud, C. (2025). *How to define and report dropout in blended therapy for mental health: A scoping review and comparative secondary analysis of operational definitions*. Open Science Framework. https://doi.org/10.31219/osf.io/9hnmk_v1

Erbe, D., Eichert, H.-C., Riper, H., & Ebert, D. D. (2017). Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: Systematic review. *Journal of Medical Internet Research*, *19*(9), Article e306. https://doi.org/10.2196/jmir.6588

Ferrao Nunes-Zlotkowski, K., Shepherd, H. L., Beatty, L., Butow, P., & Shaw, J. M. (2024). Blended psychological therapy for the treatment of psychological disorders in adult patients: Systematic review and meta-analysis. *Interactive Journal of Medical Research*, *13*(1), Article e49660. https://doi.org/10.2196/49660

Glaesmer, H., Grande, G., Braehler, E., & Roth, M. (2011). The German version of the satisfaction with Life Scale (SWLS). *European Journal of Psychological Assessment*, *27*(2), 127–132. https://doi.org/10.1027/1015-5759/a000058

Gräfe, K., Zipfel, S., Herzog, W., & Löwe, B. (2004). Screening psychischer Störungen mit dem "Gesundheitsfragebogen für Patienten (PHQ-D)" [Screening of mental disorders with the Patient Health Questionnaire (PHQ-D)]. *Diagnostica*, *50*(4), 171–181. https://doi.org/10.1026/0012-1924.50.4.171

Heinig, I., Wittchen, H.-U., & Knappe, S. (2021). Help-seeking behavior and treatment barriers in anxiety disorders: Results from a representative German Community Survey. *Community Mental Health Journal*, *57*(8), 1505–1517. https://doi.org/10.1007/s10597-020-00767-5

Hildebrand, M. (2015). The psychometric properties of the drug use disorders identification test (DUDIT): A review of recent research. *Journal of Substance Abuse Treatment*, *53*, 52–59. https://doi.org/10.1016/j.jsat.2015.01.008

Jacobi, F., Höfler, M., Siegert, J., Mack, S., Gerschler, A., Scholl, L., Busch, M. A., Hapke, U., Maske, U., Seiffert, I., Gaebel, W., Maier, W., Wagner, M., Zielasek, J., & Wittchen, H.-U. (2014). Twelve-month prevalence, comorbidity and correlates of mental disorders in Germany: The Mental Health Module of the German Health Interview and Examination Survey for Adults (DEGS1-MH). *International Journal of Methods in Psychiatric Research*, *23*(3), 304–319. https://doi.org/10.1002/mpr.1439

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Jordan, S., Becker, P., Behr, S., Fenski, F., Knaevelsrud, C., Boettcher, J., & Schaeuffele, C. (2025). "Who blends in and why (not)?" A qualitative study on psychotherapists' patient inclusion in blended care. *Internet Interventions*, *41*, Article 100847. https://doi.org/10.1016/j.invent.2025.100847

Kalde, J., Atik, E., Stricker, J., Schückes, M., Neudeck, P., Pittig, A., & Pietrowsky, R. (2024). Enhancing the effectiveness of CBT for patients with unipolar depression by integrating digital interventions into treatment: A pilot randomized controlled trial. *Psychotherapy Research*, *34*(8), 1131–1146. https://doi.org/10.1080/10503307.2023.2277866

Kemmeren, L. L., van Schaik, A., Draisma, S., Kleiboer, A., Riper, H., & Smit, J. H. (2023). Effectiveness of blended cognitive behavioral therapy versus treatment as usual for depression in routine specialized mental healthcare: E-COMPARED trial in the Netherlands. *Cognitive Therapy and Research*, *47*(3), 386–398. https://doi.org/10.1007/s10608-023-10363-y

Kliem, S., Mößle, T., Zenger, M., Strauß, B., Brähler, E., & Hilbert, A. (2016). The Eating Disorder Examination–Questionnaire 8: A brief measure of eating disorder psychopathology (EDE-Q8). *International Journal of Eating Disorders*, *49*(6), 613–616. https://doi.org/10.1002/eat.22487

Kotz, S., & Nadarajah, S. (2004). *Multivariate t-distributions and their applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511550683

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, *32*(9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B. W., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, *114*(1–3), 163–173. https://doi.org/10.1016/j.jad.2008.06.026

Lenth, R. (2025). *emmeans: Estimated marginal means, aka least-squares means*. R package (Version 1.11.0-003) [Computer software]. https://rvlenth.github.io/emmeans/

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, *46*(3), 266–274. https://doi.org/10.1097/MLR.0b013e318160d093

Lyon, A. R., & Koerner, K. (2016). User-centered design for psychosocial intervention development and implementation. *Clinical Psychology: Science and Practice*, *23*(2), 180–200. https://doi.org/10.1111/cpsp.12154

Moshe, I., Terhorst, Y., Philippi, P., Domhardt, M., Cuijpers, P., Cristea, I., Pulkki-Råback, L., Baumeister, H., & Sander, L. B. (2021). Digital interventions for the treatment of depression: A meta-analytic review. *Psychological Bulletin*, *147*(8), 749–786. https://doi.org/10.1037/bul0000334

R Core Team. (2013). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Richardson, J., Iezzi, A., Khan, M. A., & Maxwell, A. (2014). Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. *The Patient–Patient-Centered Outcomes Research*, *7*(1), 85–96. https://doi.org/10.1007/s40271-013-0036-x

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

Schaeuffele, C., Heinrich, M., Behr, S., Fenski, F., Hammelrath, L., Zagorscak, P., Jansen, A., Pohl, S., Boettcher, J., & Knaevelsrud, C. (2022). Increasing the effectiveness of psychotherapy in routine care through blended therapy with transdiagnostic online modules (PsyTOM): Study protocol for a randomized controlled trial. *Trials*, *23*(1), Article 830. https://doi.org/10.1186/s13063-022-06757-0

Schaeuffele, C., Schulz, A., Knaevelsrud, C., Renneberg, B., & Boettcher, J. (2021). CBT at the crossroads: The rise of transdiagnostic treatments. *International Journal of Cognitive Therapy*, *14*(1), 86–113. https://doi.org/10.1007/s41811-020-00095-2

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Schiefele, A.-K., Lutz, W., Barkham, M., Rubel, J., Böhnke, J., Delgadillo, J., Kopta, M., Schulte, D., Saxon, D., Nielsen, S. L., & Lambert, M. J. (2017). reliability of therapist effects in practice-based psychotherapy research: A guide for the planning of future studies. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(5), 598–613. https://doi.org/10.1007/s10488-016-0736-3

Schmidt, J., Lamprecht, F., & Wittmann, W. W. (1989). Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen [Satisfaction with inpatient care: Development of a questionnaire and first validity assessments]. *PPmP: Psychotherapie Psychosomatik Medizinische Psychologie*, *39*(7), 248–255.

Schmidt, L., Pawlitzki, M., Renard, B. Y., Meuth, S. G., & Masanneck, L. (2024). The three-year evolution of Germany's digital therapeutics reimbursement program and its path forward. *NPJ Digital Medicine*, *7*(1), Article 139. https://doi.org/10.1038/s41746-024-01137-1

Schuster, R., Laireiter, A.-R., Berger, T., Moritz, S., Meyer, B., Hohagen, F., & Klein, J. P. (2020). Immediate and long-term effectiveness of adding an Internet intervention for depression to routine outpatient psychotherapy: Subgroup analysis of the EVIDENT trial. *Journal of Affective Disorders*, *274*, 643–651. https://doi.org/10.1016/j.jad.2020.05.122

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Titzler, I., Saruhanjan, K., Berking, M., Riper, H., & Ebert, D. D. (2018). Barriers and facilitators for the implementation of blended psychotherapy for depression: A qualitative pilot study of therapists' perspective. *Internet Interventions*, *12*, 150–164. https://doi.org/10.1016/j.invent.2018.01.002

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

Velten, J., Christiansen, H., Hoyer, J., In-Albon, T., Lincoln, T., Lutz, W., Margraf, J., Schöttke, H., Stark, R., Werheid, K., Willutzki, U., Alpers, G. W., Bartholdy, S., Berger, E.-M., Brakemeier, E.-L., Bräscher, A.-K., Brockmeyer, T., Dziobek, I., Fehm, L., … Rubel, J. A. (2025). Effectiveness of cognitive behavioral therapy for adult mental disorders: A large-scale naturalistic study across 29 university outpatient clinics. *Behaviour Research and Therapy*, *186*, Article 104691. https://doi.org/10.1016/j.brat.2025.104691

Wu, Y., Levis, B., Riehm, K. E., Saadat, N., Levis, A. W., Azar, M., Rice, D. B., Boruff, J., Cuijpers, P., Gilbody, S., Ioannidis, J. P. A., Kloda, L. A., McMillan, D., Patten, S. B., Shrier, I., Ziegelstein, R. C., Akena, D. H., Arroll, B., Ayalon, L., … Thombs, B. D. (2020). Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: A systematic review and individual participant data meta-analysis. *Psychological Medicine*, *50*(8), 1368–1380. https://doi.org/10.1017/S0033291719001314

Received November 20, 2024
Revision received August 21, 2025
Accepted August 26, 2025 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.