Behaviour Research and Therapy 186 (2025) 104691

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Effectiveness of cognitive behavioral therapy for adult mental disorders: A large-scale naturalistic study across 29 university outpatient clinics

Julia Velten [a,*], Hanna Christiansen [b,c], Jürgen Hoyer [d], Tina In-Albon [e], Tania Lincoln [f], Wolfgang Lutz [g], Jürgen Margraf [a,c], Henning Schöttke [h], Rudolf Stark [i], Katja Werheid [j], Ulrike Willutzki [k], Georg W. Alpers [l], Stephan Bartholdy [m], Elisa-Maria Berger [n], Eva-Lotta Brakemeier [m], Anne-Kathrin Bräscher [n], Timo Brockmeyer [o,p], Isabel Dziobek [q], Lydia Fehm [r], Thomas Forkmann [s], Julia Glombiewski [e], Sylvia Helbig-Lang [t], Andrea Hermann [i], Anke Kirsch [u], Tim Klucken [v], Patrizia Odyniec [k], Anya Pedersen [w], Babette Renneberg [x], Almut Rudolph [y], Brian Schwartz [g], Tobias Teismann [a], Gabriele Wilz [y], Julian A. Rubel [h]

[a] *Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr University Bochum, Germany*
[b] *Clinical Child and Adolescent Psychology, Philipps University Marburg, Germany*
[c] *German Center for Mental Health (DZPG), Partner Site Bochum/Marburg, Germany*
[d] *Clinical Psychology and Psychotherapy, Technische Universität Dresden, Germany*
[e] *Department for Psychology, University of Kaiserslautern-Landau, Germany*
[f] *Clinical Psychology and Psychotherapy, University of Hamburg, Germany*
[g] *Clinical Psychology and Psychotherapy, Department of Psychology, University of Trier, Germany*
[h] *Clinical Psychology and Psychotherapy, Institute of Psychology, University of Osnabrück, Germany*
[i] *Psychotherapy and Systems Neuroscience, Justus-Liebig University Gießen, Germany*
[j] *Clinical Neuropsychology and Psychotherapy, Department of Psychology, Bielefeld University, Germany*
[k] *Clinical Psychology and Psychotherapy, Department of Psychology, Faculty of Health, University of Witten/Herdecke, Germany*
[l] *School of Social Sciences, Department of Psychology, University of Mannheim, Germany*
[m] *Clinical Psychology and Psychotherapy, Institute of Psychology, University of Greifswald, Germany*
[n] *Clinical Psychology, Psychotherapy, and Experimental Psychopathology, Institute of Psychology, Johannes Gutenberg University Mainz, Germany*
[o] *Clinical Psychology and Psychotherapy, Institute of Psychology, University of Göttingen, Germany*
[p] *Clinical Psychology and Translational Psychotherapy, Institute of Psychology, University of Münster, Germany*
[q] *Clinical Psychology of Social Interaction, Berlin School of Mind and Brain, Institute of Psychology, Humboldt Universität zu Berlin, Germany*
[r] *Center for Psychotherapy, Institute of Psychology, Humboldt-Universität zu Berlin, Germany*
[s] *Clinical Psychology and Psychotherapy, Department of Psychology, University of Duisburg-Essen, Germany*
[t] *Psychotherapy Training Program, Clinical Psychology and Psychotherapy, University of Hamburg, Germany*
[u] *Department of Clinical Psychology and Psychotherapy, University of the Saarland, Saarbrücken, Germany*
[v] *Clinical Psychology and Psychotherapy, Department of Psychology, University of Siegen, Germany*
[w] *Clinical Psychology and Psychotherapy, University of Kiel, Germany*
[x] *Clinical Psychology and Psychotherapy, Freie Universität Berlin, Germany*
[y] *Department of Counseling and Clinical Intervention, Institute of Psychology, Friedrich-Schiller Universität Jena, Germany*

ARTICLE INFO

Keywords:
Health services research
Transdiagnostic effectiveness
Clinical global impression scale
Reliable improvement

ABSTRACT

Practice-based evidence has emerged as an important complementary paradigm to studies in controlled trials. This paper presents results of a large research-practice network at German university outpatient clinics; the KODAP initiative. Pre-post effect sizes, direct assessments of change, and rates of clinically significant and reliable improvement are reported in a heterogeneous clinical sample of 6624 adult patients treated between 2023 and 2014 in 29 psychotherapeutic outpatient clinics. Clinical diagnoses, determined with structured diagnostic clinical interviews at baseline across all clinics, encompassed a wide range of psychopathology. Effectiveness was comparable to other studies in naturalistic settings ($d \approx 0.75$–0.95) and somewhat lower than changes reported in disorder-specific CBT efficacy trials. In direct assessments of change, only 1.9% of the patients reported symptom worsening and 3.4% reported no change during treatment. Overall, the results show the

* Corresponding author.
E-mail address: julia.velten@rub.de (J. Velten).

https://doi.org/10.1016/j.brat.2025.104691
Received 24 July 2024; Received in revised form 16 January 2025; Accepted 17 January 2025
Available online 23 January 2025

potential of multi-site naturalistic research initiatives in general and the effectiveness of outpatient CBT at German university outpatient clinics in particular.

The efficacy of cognitive behavioral therapy (CBT) as a treatment for a wide range of mental health conditions (Driessen & Hollon, 2010; Hofmann et al., 2012; Kroenke, 2007; Olatunji et al., 2010; Tolin, 2010; Turner et al., 2014) is supported by robust evidence from meta-analyses of randomized controlled trials (RCTs). Despite empirical support, narrow inclusion/exclusion criteria, limited age ranges, and a focus on specific disorders raise concerns about the comparability of treatments in research with other care settings (Lutz et al., 2016; Shadish et al., 2000). Recognizing this gap, complementary naturalistic studies on real-world effectiveness have gained prominence. These studies assess the feasibility and benefit of treatments in diverse populations beyond the controlled confines of RCTs. While early analyses of non-randomized data suggested limited effectiveness (Seligman, 1995), growing evidence supports routine care CBT (Hans & Hiller, 2013b), with large effects of disorder-specific treatments (e.g., $d = 0.94$; Lutz et al., 2016). Effectiveness of CBT is often comparable to efficacy found in RCTs, especially if comparable patient groups are investigated (Hans & Hiller, 2013a; Lincoln et al., 2012; Lutz et al., 2016).

However, because naturalistic research has traditionally drawn evidence from small-scale, clinic-specific studies of selected mental disorders, their narrow scope and lack of diversity limit the broader applicability of their findings (Kathmann et al., 2022; Lutz et al., 2016; Westbrook & Kirk, 2005). This is particularly true for less prevalent mental disorders, where the number of patients seeking treatment at individual clinics is too small to permit systematic evaluation of treatment effects.

To address these limitations, a paradigm shift toward collaborative efforts between clinicians and researchers has been recommended (Borkovec, 2002; Lutz et al., 2019). A prominent example is the Improving Access to Psychological Therapies (IAPT) program from the UK (Clark, 2011, 2018), which generates nation-wide practice-based evidence through the obligatory use of session-by-session patient--reported outcomes. The IAPT initiative helped to collect an impressive amount of data which has been used to uncover promising findings in relation to treatment success (Gyani et al., 2013), costs (Radhakrishnan et al., 2013), and cost-effectiveness (Zala et al., 2019).

The implementation of research-practice cooperation projects is highly dependent on the organizational framework of the country-specific healthcare system for mental disorders. While the provision and scientific evaluation of outpatient care for people with common mental disorders such as depression and anxiety in the UK is managed centrally, the majority of outpatient mental health care in other countries, such as Germany, is provided by individual service providers (e.g., therapists in private practice, outpatient clinics at hospitals or universities). As these providers are under independent scientific and therapeutic management, priorities (e.g., focus on certain disorders or treatment modalities) differ, as do the number of patients and disorders treated. Thus, a particular challenge in assessing the effectiveness of treatments outside of RCTs in countries with independent mental health care providers is the huge heterogeneity regarding which variables are collected at which time-points, and which assessment tools are used? (Hoyer et al., 2015). National collaborations between clinics have the potential to overcome this obstacle.

1. The current study

In 2013, an initiative group began laying the groundwork for a collaboration of university outpatient clinics in Germany. The aim of this joined effort was to facilitate a systematic coordination of naturalistic effectiveness research across independent clinics in Germany. The product of this initiative was a core battery of measures that focus on general psychological distress and symptoms relevant to many patient groups, such as depressive symptoms along with a number of socio-demographic and contextual variables. In 2024, 37 university outpatient clinics for adults and 16 clinics for children and adolescents are part of this network. After a series of feasibility studies (Hoyer et al., 2015; Velten et al., 2017) and studies utilizing baseline data (Velten et al., 2018, Velten et al., 2021, In-Albon et al., 2019), a first international paper described the state of the project and outlined how the research structure might be utilized in the future (Margraf et al., 2021).

The aim of the current study was to provide a description of the effectiveness of CBT to reduce patients' psychological distress and depressive symptoms for adult patients. Towards this goal, data from 29 outpatient clinics were analyzed. The effectiveness of disorder-specific CBT to improve the primary outcome of general psychological distress was considered based on the effect sizes of pre-post changes in the Brief Symptom Inventory (Derogatis, 1975). As secondary outcomes, depressive symptoms measured with the Beck Depression Inventory (Beck et al., 1996), and a direct assessment of symptom change measured with the clinical Global Impression Scales (CGI-I; Kadouri et al., 2007) answered by both patients and therapists at post-treatment were obtained. A measure of the clinical significance of change was calculated for the complete sample, patients scoring in the clinical range of the respective measures at baseline, as well as subsamples of patients with different index diagnoses.

2. Methods

2.1. Participants

The data set included $N = 6624$ patients, who were ≥18 years, received disorder-specific outpatient CBT in the participating clinics, and provided pre- and post-treatment data for either the BSI and/or BDI. Data were collected between April 2014 and August 2023 and most

Table 1
Sample description.

	Total ($N = 6624$)
Age, M (SD)	36.49 (13.5)
	n (valid %)
Gender	
Women	4269 (64.5)
Men	2349 (35.5)
Marital status	
Single	3566 (54.9)
Married	1619 (24.9)
Divorced	436 (6.7)
Other incl. separated, widowed	874 (13.5)
Partnership	
No	3055 (49.6)
Yes	3103 (50.4)
School education	
Still in school	70 (1.1)
No degree	58 (0.9)
Basic school certificate (German „Hauptschulabschluss")	672 (10.8)
Intermediate school certificate (German "Realschulabschluss")	1524 (24.4)
High School Degree (German „Abitur")	3851 (61.7)
Other	66 (1.1)
Education	
Working towards degree	1082 (17.6)
No degree	342 (5.6)
Vocational training	2007 (32.6)
College or university degree	1507 (24.5)
Other	1213 (19.7)

treatments ($n = 6,111$, 92.3%) started after January 2018. Table 1 shows the sociodemographic characteristics.

2.2. Instruments and measures

2.2.1. Treatments

Treatments were provided by licensed clinical psychologists or therapists in postgraduate training at outpatient clinics located at psychology departments in 19 cities across Germany. Therapists in training had at least a bachelor's degree in psychology, a master's degree in clinical psychology, one and a half year of clinical experience from working in an inpatient setting, and regular meetings (every forth session) with a licensed CBT supervisor. Treatments generally followed published CBT guidelines for each disorder, but were in most cases less standardized than in RCTs. Treatments were paid for by the German health insurance system, which routinely covers 24 to 60 sessions.

2.2.2. Brief Symptom Inventory

The Brief Symptom Inventory (BSI; Derogatis, 1975) is a widely used questionnaire designed to measure psychological distress and symptoms of mental disorders. The BSI consists of 53 items that are rated on a 5-point Likert scale from 0 (not at all) to 4 (extremely). The Global Severity Index (GSI) is an indicator of psychological distress level that combines information about the number of symptoms and the intensity of distress. In the present study, internal consistency at baseline was excellent ($\alpha = .96$).

2.2.3. Beck Depression Inventory

The Beck Depression Inventory II (BDI; Beck et al., 1996) assesses depressive symptoms with 21 items rated on item-specific 4-point scales ranging from 0 to 3, yielding summary scores that range from 0 to 63. Due to their high clinical relevance, levels of depressive symptoms (e.g., changes in sleep, lack of concentration) are commonly considered secondary outcomes in treatment studies of anxiety, posttraumatic stress (Hans & Hiller, 2013a), or somatoform disorders (Bleichhardt et al., 2004). In the present study, internal consistency at baseline was excellent ($\alpha = .91$).

2.2.4. Clinical Global Impression Scale

The Clinical Global Impression Scale (CGI; Guy, 1976; Kadouri et al., 2007) allows for ratings of severity of symptoms (CGI-S) and global improvement (CGI-I) in mental health care settings. At the beginning of the treatment, therapists rated illness severity (i.e., CGI-S) on a 7-point scale ranging from 1 (not ill at all) to 7 (among the most extremely ill patients). At the end of treatment, both patients and therapists rated global improvement of symptoms (i.e., the CGI-I) on a 7-point scale ranging from 1 (very much improved) to 7 (very much worse). As a direct measure of change, the CGI-I is a face-valid tool for monitoring therapeutic outcome, regardless of the disorder being treated (Busner & Targum, 2007).

2.2.5. Clinical diagnoses

Diagnoses at baseline were determined based on structured clinical interviews, assessing ICD-10, DSM-IV, or DSM-5 criteria, whose validity and reliability exceed clinical judgment and non-standardized diagnostic procedures (Margraf et al., 2017). Interviews included the Structured Clinical Interview (First et al., 2016; Wittchen et al., 1997) and the Diagnostic Interview for Mental Disorders (Margraf & Cwik, 2017).

2.3. Procedure

Patients provided written informed consent for their anonymized data to be used in research projects. In line with German regulations, a limited number of sessions are reserved for diagnostic procedures including case history. Pre-treatment data were assessed during this initial phase, post-treatment data were assessed at the end of treatment, however, in cases where no data assessment was conducted at the end of treatment (e.g., because treatment was terminated early by the patient), the last available assessment after baseline was used. Study procedures have been approved by the Ethics Committee of the Faculty of Psychology of [blinded for peer-review].

2.4. Data analysis

Descriptive variables for patients, therapists, and clinics were considered to characterize the sample. To assess treatment effectiveness, psychological distress, and depressive symptoms assessed at baseline and post-treatment were analyzed with paired t-tests and effect sizes were reported as Cohen's d with 95% confidence intervals. Effect sizes were calculated by subtracting the post-treatment score from the initial assessment score and dividing by the pooled standard deviation. Clinical significance of the improvements was scrutinized focusing on subsamples of participants who scored in the clinical range of the respective questionnaire, meaning they had baseline scores above 0.67 for the GSI of the BSI (Kliem & Brähler, 2017) or above 14.5 on the BDI (von Glischinski et al., 2019). Using the methodology developed by Jacobson and Truax (1991), we combined these cut-off values at post-treatment with the Reliable Change Index (RCI) between baseline and post-assessments to create the following categories: (1) Recovered: Clinically significant improvement as shown by reliable improvement (RCI >1.96) with post-treatment scores in the healthy range, (2) Reliably improved: Reliable improvement but post-treatment scores still in the clinical range, and (3) Reliably deteriorated: Any reliable deterioration (RCI < −1.95). Internal consistencies reported in the respective treatment manuals of the GSI ($\alpha = .96$; Kliem & Brähler, 2017) and BDI ($\alpha = .91$; Hautzinger et al., 2023) as well as the standard deviations at baseline for all participants with valid change-scores ($SD = 0.62$ for GSI and $SD = 11.23$ for BDI) were used to calculate the RCIs. Pre-post change-scores $> |0.34|$ and $> |9|$ (Keller & Alexandrowicz, 2024) were considered as indicating reliable change for GSI and BDI, respectively. Further, as a direct measure of change, CGI-I responses by patients and therapists at post-assessment were analyzed.

To provide information on the differential effectiveness of treatments across mental disorders and comorbidities, analyses were conducted for the complete sample as well as for the most common diagnostic categories. The list of diagnoses and diagnostic categories was determined based on ICD-10-CM codes with individual categories created for disorders with 50 or more patients.

3. Results

3.1. Descriptive information

At baseline, information on mental health diagnoses was available for 6570 patients (99.2%). Of these, 97.7% ($n = 6422$) received at least one diagnosis. The mean number of mental disorders at baseline was 1.55 with 57.4% ($n = 3772$) of patients receiving one, 27.8% ($n = 1827$) receiving two, 9.3% ($n = 608$) receiving three, 2.5% ($n = 164$) receiving four, and 0.8% ($n = 51$) receiving five or more diagnoses. Almost half (49%; $n = 2849$) of patients did not receive prior treatment for their mental health problem, 22.8% ($n = 1326$) had received outpatient treatment, 11.9% ($n = 692$) had received inpatient treatment, and 14.8% ($n = 863$) had received both. At the start of treatment, 74.3% ($n = 4313$) of patients were able to work. In addition to the 17.3% ($n = 1007$) of patients on sick leave, 4.0% ($n = 233$) received a retirement pension, 2.1% ($n = 122$) an invalidity pension, and 2.2% indicated "other" ($n = 130$). According to their therapists, 0.1% ($n = 4$) of the patients had no clinical disorder, 1.9% ($n = 93$) were borderline cases of mental disorder, 11.6% (n = 575) were characterized by mild, 31.4% ($n = 1555$) by moderate, 42.1% ($n = 2083$) by marked, 12.0% ($n = 595$) by severe, and 0.8% ($n = 39$) by extreme symptomatology. The five most

common index diagnoses or diagnostic categories at baseline were Recurrent depressive disorder, not currently in remission (ICD-10: F33 without F33.4) with 21.7% ($n = 1436$), Single depressive episode (ICD-10: F32) with 12.3% ($n = 813$), Acute stress reaction, adjustment disorders, and other reactions to severe stress (ICD-10: F43.0/2/8/9) with 9.0% ($n = 596$), Social phobias (ICD-10: F40.1) with 7.0% ($n = 462$), and Agoraphobia and Panic disorder (ICD-10: F40.0 and F41.0) with 6.9% ($n = 455$). A complete list with categories of index diagnoses can be found in Supplementary Table S1.

In total, treatments were provided by 1110 therapists who, on average, treated 5.97 patients of the present sample (range = 1 to 60, median = 5). Most therapists were women ($n = 918$, 82.7%) and at the beginning of the treatments 31.88 years old (range = 23 to 59). Many treatments (67.2%) were provided by psychotherapists in training.

Twenty-nine outpatient clinics located at 23 institutions (e.g., university departments) provided data. On average, data of 228 patients ($SD = 318.52$, range = 1 to 1384) per clinic were included. The average number of regular psychotherapy sessions after the diagnostic phase was 40.32 ($SD = 22.32$, range = 0 to 134). While most treatments were individual therapy (93.6%; $n = 6158$), some included a combination of group and individual therapy (6.3%, $n = 417$). As many treatments were provided during the COVID-19 pandemic, 43.4% ($n = 2602$) of treatments included at least one session of online therapy.

3.2. Treatment effectiveness

Patients ($n = 6460$) showed significant improvements in the primary outcome of psychological distress as measured with the GSI, M_{Pre} (SD) = 1.04 (0.62), M_{Post} (SD) = 0.63 (0.57), $t(6459) = 59.94$, $p < .001$, $d = 0.75$ [0.72; 0.77]. Depressive symptoms ($n = 6160$) also improved significantly, M_{Pre} (SD) = 20.97 (11.23), M_{Post} (SD) = 10.80 (10.33), $t(6159) = 74.75$, $p < .001$, $d = 0.95$ [0.92; 0.98].

Both psychological distress and depressive symptoms significantly improved across all categories of index diagnoses, with smallest effects found for psychological distress in patients diagnosed with disorders that began in childhood and adolescence, and largest effects for depressive symptoms in patients with Major depressive disorder, single episode. Table 2 shows the treatment effectiveness for each diagnostic category.

Considering the number of baseline diagnoses, improvements in psychological distress, $F(4, 6256) = 17.74$, $p < .001$, $\eta^2 = 0.011$, and depressive symptoms, $F(4, 5981) = 17.08$, $p < .001$, $\eta^2 = 0.011$, were larger for patients with more comorbidity (See Supplementary Table S2 for more information).

3.3. Clinical significance

Focusing only on patients with values above the clinical cut-off at baseline, patients showed large reductions in psychological distress as measured with the GSI ($n = 4407$), M_{Pre} (SD) = 1.34 (0.52), M_{Post} (SD) = 0.78 (0.60), $t(4406) = 63.70$, $p < .001$, $d = 0.96$ [0.92; 1.00], with 48.0% of these patients ($n = 2114$) recovering and additional 18.1% ($n = 798$) showing reliable improvement at post-treatment. Further, depressive symptoms ($n = 4251$) also improved significantly in patients starting in the clinical range of the BDI, M_{Pre} (SD) = 26.61 (8.54), M_{Post} (SD) = 13.19 (11.04), $t(4250) = 81.62$, $p < .001$, $d = 1.25$ [1.21; 1.29],

Table 2
Changes in psychological distress and depressive symptoms in the complete sample.

	Global Severity Index				Beck Depression Inventory			
Index diagnoses or diagnostic categories	*N*	Pre *M (SD)*	Post *M (SD)*	ES	N	Pre *M (SD)*	Post *M (SD)*	ES
Disorders due to psychoactive substance use (F10 to F19)	51	0.92 (0.61)	0.56 (0.56)	0.69 [0.38; 1.00]	50	18.48 (11.78)	9.48 (9.82)	0.94 [0.60; 1.27]
Schizophrenia and other psychotic disorders (F20 to F29)	120	0.90 (0.54)	0.66 (0.52)	0.54 [0.35; 0.73]	118	17.50 (10.31)	11.24 (9.21)	0.69 [0.49; 0.89]
Bipolar disorder (F31)	67	1.00 (0.62)	0.68 (0.64)	0.49 [0.24; 0.75]	68	20.69 (11.49)	12.38 (11.42)	0.76 [0.48; 1.02]
Major depressive disorder, single episode (F32)	793	1.12 (0.59)	0.61 (0.56)	0.92 [0.84; 1.00]	757	24.49 (9.94)	10.69 (9.98)	1.24 [1.15; 1.34]
Major depressive disorder, recurrent (F33)	1495	1.18 (0.61)	0.72 (0.57)	0.81 [0.75; 0.87]	1443	25.09 (10.73)	12.77 (10.89)	1.13 [1.06; 1.19]
Persistent mood (affective) disorders (F34)	221	1.06 (0.60)	0.73 (0.90)	0.65 [0.51; 0.80]	221	22.24 (10.19)	12.47 (10.18)	0.88 [0.72; 1.03]
Agoraphobia and Panic disorder (F40.0. F41.0)	448	0.96 (0.57)	0.54 (0.52)	0.76 [0.66; 0.87]	406	16.30 (9.76)	8.40 (9.03)	0.86 [0.74; 0.97]
Social phobias (F40.1)	443	1.10 (0.57)	0.64 (0.54)	0.84 [0.74; 0.95]	440	20.08 (10.61)	9.78 (9.48)	1.04 [0.92; 1.16]
Specific (isolated) phobias (F40.2)	176	0.65 (0.53)	0.38 (0.39)	0.67 [0.50; 0.83]	164	11.77 (9.82)	5.84 (6.31)	0.65 [0.48; 0.82]
Generalized anxiety disorder (F41.1)	182	1.08 (0.58)	0.65 (0.59)	0.76 [0.60; 0.93]	172	18.95 (9.05)	9.82 (9.53)	0.91 [0.74; 1.09]
Obsessive-compulsive disorder (F42)	218	0.96 (0.57)	0.59 (0.55)	0.70 [0.55; 0.84]	210	17.60 (10.57)	9.53 (9.62)	0.81 [0.66; 0.97]
Adjustment disorders (F43.2)	583	0.76 (0.51)	0.43 (0.42)	0.68 [0.59; 0.77]	534	16.26 (8.93)	7.54 (7.60)	0.94 [0.84; 1.04]
Post-traumatic stress disorder (F43.1)	208	1.46 (0.73)	0.88 (0.75)	0.84 [0.68; 0.99]	200	27.07 (11.88)	14.08 (13.57)	1.05 [0.88; 1.23]
Somatoform disorders (F45)	360	0.91 (0.54)	0.56 (0.47)	0.75 [0.64; 0.87]	334	17.78 (9.71)	9.89 (8.91)	0.89 [0.77; 1.02]
Eating disorders (F50)	246	1.05 (0.62)	0.65 (0.52)	0.70 [0.56; 0.84]	238	22.38 (11.32)	11.90 (10.50)	0.94 [0.78; 1.09]
Borderline personality disorder (F60.3)	160	1.57 (0.72)	1.06 (0.80)	0.78 [0.60; 0.95]	157	28.72 (13.84)	18.00 (15.15)	0.86 [0.68; 1.04]
Other personality disorders (F60 except F 60.3)	60	1.01 (0.62)	0.67 (0.52)	0.60 [0.33; 0.88]	58	21.14 (12.07)	11.59 (10.24)	0.88 [0.57; 1.18]
Disorders with onset in childhood and adolescence (F90-F98)	76	0.96 (0.54)	0.74 (0.65)	0.39 [0.15; 0.62]	78	16.01 (10.20)	10.81 (10.62)	0.50 [0.26; 0.73]
Other	354	0.81 (0.57)	0.53 (0.52)	0.56 [0.45; 0.67]	338	15.40 (10.65)	9.29 (9.55)	0.61 [0.49; 0.72]

with 54.7% of these patients (n = 2325) recovering and additional 11.0% (n = 468) reliably improving. In the complete sample, rates of reliable deterioration were 5.6% (n = 364) and 2.3% (n = 142) for psychological distress (GSI) and depressive symptoms (BDI), respectively. Across different diagnostic categories, recovery rates for psychological distress (See Table 3) ranged from 30.7% (n = 43) among patients with Borderline personality disorder to 62.9% (n = 44) among patients with Specific (isolated) phobias.

For depressive symptoms (See Table 4), 39.2% (n = 51) among patients with Borderline personality disorder recovered, while 67.9 % (n = 38) of patients with Specific phobias recovered.

Considering the number of baseline diagnoses, recovery rates for psychological distress ranged from 31.3% (n = 36) in patients with four baseline diagnoses to 51.1% (n = 1152) in patients with one baseline diagnosis. For depressive symptoms, recovery rates ranged from 40.8% (n = 51) in patients with four diagnoses to 56.9% (n = 1247) in patients with one baseline diagnosis (See Supplementary Table S3 for more information).

3.4. Direct assessment of change

Of 5193 patients with available data, 1.9% rated their symptoms as worse, 3.4% as unchanged, 17.5% as minimally improved, 47.2% as much improved, and 30.0% as very much improved at post-treatment. Clinicians (n = 5364) rated 2.1% of their patients' symptoms as worse, 5.7% as unchanged, 25.7% as minimally improved, 44.4% as much improved, and 22.1% as very much improved. Figs. 1 and 2 show patients' and therapists' CGI-I ratings across different diagnostic categories.

Considering different baseline diagnoses, most patients rated their symptoms as either minimally, much, or very much improved after treatment, with improvement ratings ranging from 90.3% for patients with posttraumatic stress disorder to 98.1% for patients with schizophrenia and other psychotic disorders as the index diagnosis. In contrast, only a minority of patients reported a worsening of overall symptoms with a maximum of 4.6% for patients with posttraumatic stress disorder. Therapists' CGI-I ratings also showed high rates of improvement ranging from 83.1% for patients with borderline personality disorder to 96.4% for patients with specific phobias.

Concerning the degree of comorbidity, rates of improvement as rated by patients were 95.4%, 93.7%, 93.2%, 92.3%, and 96.8% for patients with one, two, three, four, five and more baseline diagnoses, respectively. Similarly, therapists rated 92.9%, 92.4%, 87.4%, 87.2%, and 90.6% of patients with one, two, three, four, five and more baseline diagnoses as improved. Both patients' and therapists' assessments of symptom change showed a small negative association with the number of baseline diagnoses in that patients with more diagnoses, $r(5094) = 0.042$, $p = .003$, as well as their therapists, $r(5217) = 0.06$, $p < .001$, rated improvement of symptoms slightly less positively (see Supplementary Table S4 for more details). Associations between CGI and post as well as change scores of psychometric questionnaires are presented in Supplementary File S5.

Table 3
Clinically significant changes in psychological distress for patients scoring in the clinical range at baseline.

Index diagnoses or diagnostic categories	*N* (%)	Pre *M (SD)*	Post *M (SD)*	ES	Recovered, *N* (%)	Reliably improved, *N* (%)	Reliably deteriorated, *N* (%)*
Disorders due to psychoactive substance use (F10 to F19)	31 (60.8)	1.27 (0.54)	0.77 (0.61)	0.84 [0.41; 1.21]	13 (41.9)	4 (12.9)	2 (3.9)
Schizophrenia and other psychotic disorders (F20 to F29)	68 (56.7)	1.25 (0.46)	0.84 (0.57)	0.85 [0.57; 1.24]	21 (30.9)	13 (19.1)	10 (8.3)
Bipolar disorder (F31)	42 (62.7)	1.35 (0.50)	0.90 (0.69)	0.59 [0.26; 0.92]	16 (38.1)	8 (19.0)	7 (10.4)
Major depressive disorder, single episode (F32)	591 (74.5)	1.34 (0.51)	0.70 (0.59)	1.15 [1.05; 1.26]	332 (56.2)	108 (18.3)	44 (5.5)
Major depressive disorder, recurrent (F33)	1163 (77.8)	1.39 (0.52)	0.82 (0.59)	0.98 [0.91; 1.05]	529 (45.5)	230 (19.8)	83 (5.6)
Persistent mood (affective) disorders (F34)	160 (72.4)	1.30 (0.54)	0.85 (0.62)	0.81 [0.63; 0.99]	62 (38.8)	28 (17.5)	17 (7.7)
Agoraphobia and Panic disorder (F40.0. F41.0)	287 (64.1)	1.26 (0.47)	0.69 (0.57)	1.00 [0.86; 1.14]	149 (51.9)	47 (16.4)	23 (5.1)
Social phobias (F40.1)	334 (75.4)	1.32 (0.48)	0.76 (0.56)	0.96 [0.83; 1.09]	174 (52.1)	49 (14.7)	25 (5.6)
Specific (isolated) phobias (F40.2)	70 (39.8)	1.20 (0.40)	0.63 (0.46)	1.29 [0.97; 1.61]	44 (62.9)	8 (11.4)	6 (3.4)
Generalized anxiety disorder (F41.1)	133 (73.1)	1.31 (0.50)	0.79 (0.62)	0.84 [0.64; 1.04]	64 (48.1)	27 (20.3)	12 (6.6)
Obsessive-compulsive disorder (F42)	137 (62.8)	1.28 (0.47)	0.76 (0.60)	0.90 [0.70; 1.09]	65 (47.4)	26 (19.0)	13 (6.0)
Adjustment disorders (F43.2)	280 (48.0)	1.16 (0.44)	0.61 (0.49)	1.02 [0.87; 1.16]	154 (55.0)	34 (12.1)	28 (4.8)
Post-traumatic stress disorder (F43.1)	178 (85.6)	1.63 (0.64)	0.96 (0.77)	0.96 [0.78; 1.13]	79 (44.4)	39 (21.9)	13 (6.3)
Somatoform disorders (F45)	224 (62.2)	1.21 (0.46)	0.70 (0.51)	1.03 [0.87; 1.19]	105 (46.9)	47 (21.0)	15 (4.2)
Eating disorders (F50)	169 (68.7)	1.35 (0.52)	0.77 (0.55)	0.97 [0.78; 1.15]	77 (45.6)	31 (18.3)	15 (6.1)
Borderline personality disorder (F60.3)	138 (86.3)	1.74 (0.60)	1.16 (0.81)	0.89 [0.69; 1.08]	43 (30.7)	46 (33.6)	9 (5.6)
Other personality disorders (F60 except F 60.3)	39 (65.0)	1.34 (0.52)	0.82 (0.54)	0.85 [0.48; 1.22]	15 (38.5)	9 (23.1)	4 (6.7)
Disorders with onset in childhood and adolescence (F90-F98)	49 (64.5)	1.23 (0.47)	0.92 (0.62)	0.59 [0.28; 0.89]	17 (34.7)	7 (14.3)	5 (6.6)
Other	188 (53.1)	1.23 (0.47)	0.75 (0.59)	0.77 [0.60; 0.93]	90 (47.9)	20 (10.6)	21 (5.9)

Note. * Values and percentages based on complete sample.

Table 4
Clinically significant changes in depressive symptoms for patients scoring in the clinical range at baseline.

Index diagnoses or diagnostic categories	*N* (%)	Pre *M (SD)*	Post *M (SD)*	ES	Recovered, *N* (%)	Reliably improved, *N* (%)	Reliably deteriorated *N* (%)*
Disorders due to psychoactive substance use (F10 to F19)	29 (58.0)	26.28 (9.12)	13.28 (10.57)	1.29 [0.79; 1.78]	15 (51.7)	3 (10.3)	1 (2.2)
Schizophrenia and other psychotic disorders (F20 to F29)	70 (59.3)	24.33 (7.39)	14.16 (9.94)	1.14 [0.84; 1.44]	32 (45.7)	8 (11.4)	2 (1.7)
Bipolar disorder (F31)	48 (70.6)	26.58 (7.86)	15.56 (11.76)	0.97 [0.62; 1.31]	22 (45.8)	7 (14.6)	4 (6.0)
Major depressive disorder, single episode (F32)	626 (82.7)	27.43 (8.15)	11.72 (10.32)	1.45 [1.34; 1.56]	402 (64.2)	59 (9.4)	12 (1.5)
Major depressive disorder, recurrent (F33)	1195 (82.8)	28.42 (8.56)	14.16 (11.18)	1.30 [1.25; 1.41]	641 (53.6)	182 (15.2)	28 (1.9)
Persistent mood (affective) disorders (F34)	167 (75.6)	26.21 (8.36)	13.63 (10.51)	1.19 [0.99; 1.38]	81 (48.5)	15 (9.0)	8 (3.6)
Agoraphobia and Panic disorder (F40.0. F41.0)	214 (52.7)	23.67 (7.11)	11.77 (10.52)	1.19 [1.01; 1.36]	119 (55.6)	11 (5.1)	10 (2.2)
Social phobias (F40.1)	304 (69.1)	25.37 (8.11)	12.07 (10.13)	1.32 [1.16; 1.47]	171 (56.3)	31 (10.2)	3 (0.7)
Specific (isolated) phobias (F40.2)	56 (34.1)	23.21 (6.90)	9.36 (7.55)	1.38 [1.01; 1.75]	38 (67.9)	2 (3.6)	1 (0.6)
Generalized anxiety disorder (F41.1)	116 (67.4)	23.42 (7.44)	11.19 (10.40)	1.08 [0.85; 1.31]	63 (54.3)	6 (5.2)	7 (3.8)
Obsessive-compulsive disorder (F42)	120 (57.1)	24.52 (8.40)	12.53 (10.75)	1.18 [0.95; 1.41]	57 (47.5)	13 (10.8)	6 (2.8)
Adjustment disorders (F43.2)	300 (56.2)	22.40 (6.53)	9.48 (8.53)	1.41 [1.25; 1.57]	193 (64.3)	13 (4.3)	7 (1.2)
Post-traumatic stress disorder (F43.1)	169 (84.5)	30.38 (9.64)	15.76 (13.95)	1.17 [0.97,1.36]	86 (50.9)	27 (16.0)	5 (2.4)
Somatoform disorders (F45)	197 (59.0)	24.14 (7.12)	12.84 (9.47)	1.26 [1.07; 1.45]	92 (46.7)	20 (10.2)	6 (1.7)
Eating disorders (F50)	173 (72.7)	27.37 (8.95)	13.93 (11.15)	1.19 [1.00; 1.39]	95 (54.9)	19 (11.0)	5 (2.0)
Borderline personality disorder (F60.3)	130 (82.8)	32.90 (10.73)	20.68 (15.17)	0.94 [0.73; 1.15]	51 (39.2)	24 (18.5)	7 (4.3)
Other personality disorders (F60 except F 60.3)	38 (65.5)	27.97 (8.73)	14.13 (10.93)	1.31 [0.87; 1.74]	18 (47.4)	7 (18.4)	1 (1.7)
Disorders with onset in childhood and adolescence (F90-F98)	38 (48.7)	24.21 (8.00)	14.45 (11.21)	0.95 [0.56; 1.33]	17 (44.7)	3 (7.9)	4 (5.3)
Other	158 (46.7)	24.70 (7.88)	13.84 (11.04)	0.92 [0.74; 1.11]	77 (48.7)	11 (7.0)	1 (0.3)

Note. * Values and percentages based on complete sample.

4. Discussion

The aim of this study was to determine the effectiveness of CBT treatments as provided at university outpatient clinics in Germany. We found significant, medium to large improvements in psychological distress and large improvements in depressive symptoms from baseline to post-treatment across almost all diagnostic categories and levels of comorbidity. Direct improvement ratings by patients and therapists were high, with levels of improvement commonly exceeding 90%. This large-scale multi-site assessment documents that effective CBT interventions can be transferred from clinical trials to outpatient treatment centers. This adds to the growing evidence that CBT for specific disorders can be delivered effectively outside of RCTs (Öst et al., 2023; Schumm et al., 2022) and in unselected patient populations (Taubitz et al., 2022).

4.1. Psychological distress

In the total sample, improvement in psychological distress was slightly lower than that reported in previous practice-based studies. For example, a study on licenced psychotherapists providing services in private practices in Germany reported pre-post changes in the GSI of $d = 0.94$ (Strauss et al., 2015). Since our effectiveness analyses included patients independent of their baseline levels of distress, lower overall effects were expected as compared to studies limiting their analyses to patients with elevated baseline levels (Hans & Hiller, 2013a, 2013b). Accordingly, in the subgroup analysis of patients starting treatment in the clinical range of the GSI, the average baseline score as well as the observed pre-post changes were almost the same as in the aforementioned study (Strauss et al., 2015).

There is, however, evidence for a relative lack of sensitivity to change of the BSI as compared to more disorder-specific scales as lower rates of recovery have been found in studies on common mental disorders using this outcome measure (Conway et al., 2003; Von Brachel et al., 2019). This is in line with the repeated finding that symptom specific measures show larger effect sizes than instruments with lower specificity (Minami et al., 2007). As a significant proportion of patients with clinical diagnoses did not meet the clinical threshold of the GSI at baseline, lower rates of clinical change were expected. In this study, for example, less than half of the patients treated for specific phobias did not have clinical levels of distress at baseline, resulting in a high proportion of patients who were not expected to and did not show substantial improvements in distress. Accordingly, the GSI effect size especially for those subgroups with a high rate of patients starting in the non-clinical range (e.g., specific phobias or adjustment disorder), is substantially lower in the complete sample. Therefore, judging treatment effectiveness based on the GSI effect sizes of the complete sample is especially biased for these subgroups of patients. In the subsample starting in the clinical GSI range, the effect sizes are higher (e.g., specific phobias: $d = 1.29$; adjustment disorder: $d = 1.02$) indicating that the treatments provided are also highly effective for these diagnostic groups. Future research will need to identify predictors of successful outcome, possibly using modern analytic approaches such as machine learning (Taubitz et al., 2022).

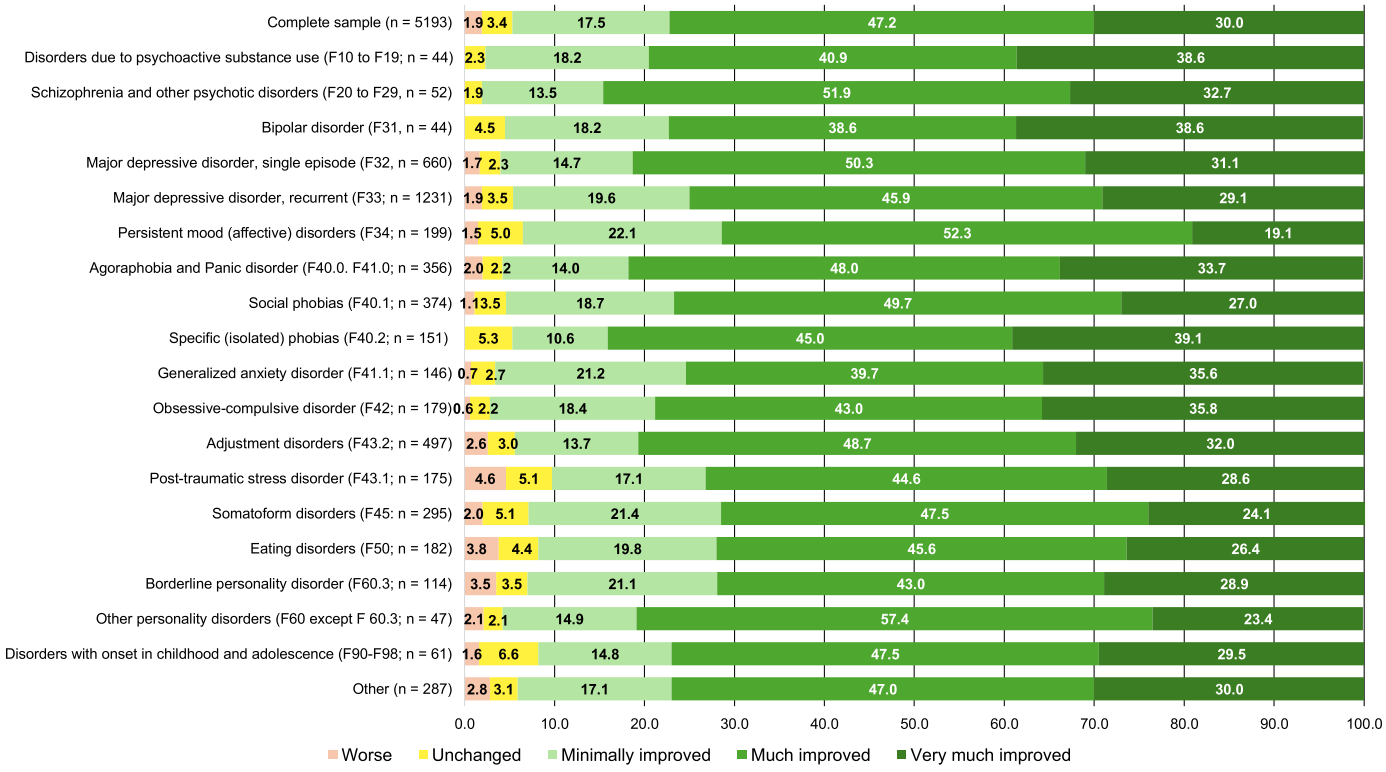

Fig. 1. Patients' clinical global impression Improvement (CGI-I) ratings across different diagnostic categories.

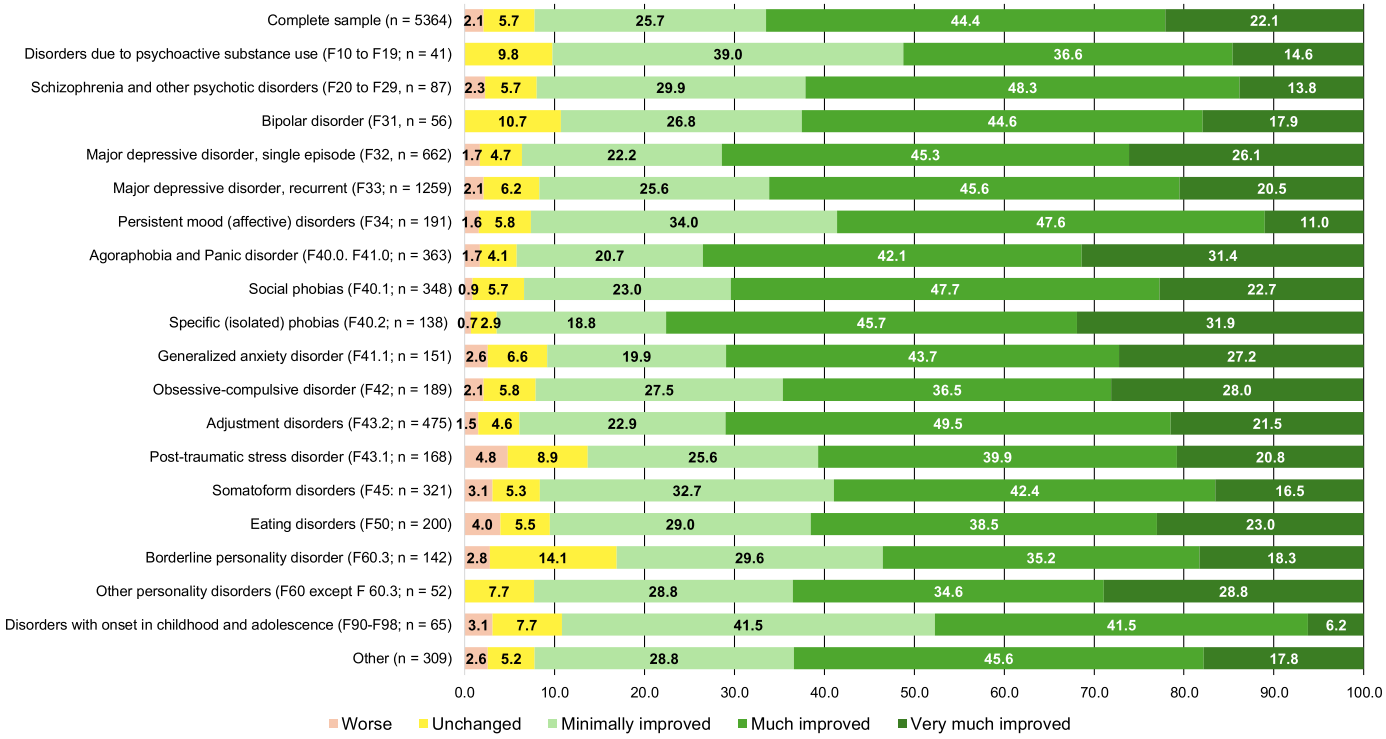

Fig. 2. Therapists' Clinical Global Impression Improvement (CGI-I) ratings across different diagnostic categories.

4.2. *Depressive symptoms*

Improvements in depressive symptoms were large in almost all diagnostic categories, with patients treated for Major depressive disorder, the largest subgroup of patients, showing the greatest reductions. The large reductions found in most patient groups point to the high comorbidity of depression with other mental disorders (e.g., anxiety disorders) as well as the broad relevance of symptoms measured with depression questionnaires as indicators of general distress (Böhnke et al., 2014). Because questionnaires such as the BDI cover a spectrum of

symptoms such as changes in sleep and appetite, lack of concentration, or agitation that are relevant to several mental disorders such as anxiety, posttraumatic stress disorder, or adjustment disorders, such scales may be useful for assessing therapeutic change beyond depression. Consequently, the BDI shows on average higher pre-post effect sizes than the ones observed in the GSI of the BSI. Concerning reliable and clinically significant change, there is a huge heterogeneity. In this study, about 54.7% of patients starting in the clinical range of the BDI at baseline recovered. This is in line with some studies that showed rates above 50% (Cahill et al., 2010), while others reported rates of 34% (Holmqvist et al., 2014). It is important to note that it is difficult to compare rates of recovered or reliably improved patients across studies since these are highly dependent on various methodological choices (e.g., reliability of the measure used, normative sample characteristics).

4.3. Direct assessment of change

High rates of improvement were reported by both patients and therapists with patients providing an even more positive evaluation of their experienced changes. Comparable rates of around 90% have been reported in other studies (Berk et al., 2008), and both patients and therapists report significant improvement in psychopathological symptoms after treatment. A major advantage of direct assessments is their lower dependency on baseline assessments (Flückiger et al., 2007). While pre-post effects (i.e., change scores) are highly correlated with baseline scores), direct assessments show only weak associations. Thus, direct measures provide a useful complementary perspective to pre-post effects especially in samples with heterogeneous baseline scores like the present one. As would be expected based on the high correlation between baseline and change scores, some of the diagnostic groups with low effect sizes started treatment with a relatively low symptom impairment (e.g., patients with specific phobias). While these groups show markedly lower effect sizes than the group with the highest pre-post effects (i.e., patients with a single episode of major depressive disorder), effectiveness based on the direct assessments from the patients' and therapists' perspective are comparable. Despite the advantages of direct assessment of change, the interpretability of the CGI-I may be limited by social desirability, as patients and therapists may be inclined to evaluate patients' progress more positively because CGI-I scores may be discussed during the final therapy session as part of the final evaluation.

4.4. Limitations and future directions

Although findings derived from university-based outpatient clinics may not be directly applicable to treatments provided by other outpatient care providers (e.g., private practitioners) that do not have close associations with research institutions, results resemble other naturalistic studies including routine treatments in the German health care system. This comparability is reflected in similar treatment lengths, reliance on clinical decision-making rather than treatment manuals, as well as similar comorbidity patterns (Strauss et al., 2015). The slightly lower pre-post changes in the complete sample may give rise to the hypothesis, that therapists' young age and relative inexperience could have a negative influence. However, studies so far showed only weak support for the notion that therapists show improved outcomes with increasing age and experience (Wampold & Owen, 2021; Tracey et al., 2024). Thus, we believe that it is more likely that these reduced effects were a result of slightly lower baseline scores than less effective treatments. A further enhancement of effect sizes might be achievable through a broader implementation of evidence-based methods (e.g., exposure treatments Barkham et al., 2023; Pittig et al., 2019).

A limitation of the present study is the lack of a control group. As a consequence, the observed uncontrolled pre-post effects cannot be attributed to the treatment alone. That is, it is unclear whether the observed changes would have occurred if a patient had received no treatment at all. However, since studies on the natural course of, e.g., depressive symptoms, yielded significantly smaller effects ($d = 0.37$) for waitlist-control groups, it is likely that larger changes in depressive symptomatology are associated with outpatient treatment.

In addition, the direct assessments of change did not explicitly ask patients or therapists whether they attributed changes experienced to treatment or to circumstantial factors. Subjective believes about the reasons for change may be a worthwhile addition to direct assessments of change. For future development of KODAP and similar practice-based research networks, it may be interesting to include waitlist data to create synthetic controls to adjust for treatment effects (Kaiser et al., 2023).

Another limitation relates to potential differences between sites. Because some clinics focus on specific disorders and/or treatments, there may be systematic differences between sites in outcomes and/or other characteristics that make it difficult to compare treatment effectiveness of different disorders. Thus, future investigations of the KODAP network should include the specific profile of the participating clinics and possible resulting differences, not only in terms of therapy outcome, but also in other parameters such as the duration of therapy.

5. Conclusion

Evidence from this collaborative study of 29 university-based outpatient clinics in Germany underscores the effectiveness of CBT in this setting. Improvements of psychological distress and depressive symptoms in patients with a wide range of mental disorders were observed. Research-practice networks not only generate data on the effects of treatments delivered outside of RCTs but also provide research infrastructures that allow for the integration of future practice-oriented research projects on a large scale.

CRediT authorship contribution statement

Julia Velten: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Hanna Christiansen:** Writing – review & editing, Supervision, Project administration. **Jürgen Hoyer:** Writing – review & editing, Supervision, Project administration. **Tina In-Albon:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Tania Lincoln:** Writing – review & editing, Supervision, Project administration. **Wolfgang Lutz:** Writing – review & editing, Supervision, Project administration. **Jürgen Margraf:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Henning Schöttke:** Writing – review & editing, Supervision, Project administration. **Rudolf Stark:** Writing – review & editing, Project administration, Funding acquisition. **Katja Werheid:** Writing – review & editing, Supervision, Project administration. **Ulrike Willutzki:** Writing – review & editing, Supervision, Project administration. **Georg W. Alpers:** Writing – review & editing, Project administration. **Stephan Bartholdy:** Writing – review & editing, Project administration. **Elisa-Maria Berger:** Writing – review & editing, Project administration. **Eva-Lotta Brakemeier:** Writing – review & editing, Project administration. **Anne-Kathrin Bräscher:** Writing – review & editing, Project administration. **Timo Brockmeyer:** Writing – review & editing, Project administration. **Isabel Dziobek:** Writing – review & editing, Project administration. **Lydia Fehm:** Writing – review & editing, Project administration. **Thomas Forkmann:** Writing – review & editing, Project administration. **Julia Glombiewski:** Writing – review & editing, Project administration. **Sylvia Helbig-Lang:** Writing – review & editing, Project administration. **Andrea Hermann:** Writing – review & editing, Project administration. **Anke Kirsch:** Writing – review & editing, Project administration. **Tim Klucken:** Writing – review & editing, Project administration. **Patrizia Odyniec:** Writing – review & editing, Project administration. **Anya Pedersen:** Writing – review & editing, Project administration. **Babette Renneberg:** Writing – review & editing, Project administration. **Almut Rudolph:** Writing – review & editing, Project

administration. **Brian Schwartz:** Writing – review & editing, Project administration. **Tobias Teismann:** Writing – review & editing, Project administration. **Gabriele Wilz:** Writing – review & editing, Project administration. **Julian A. Rubel:** Writing – review & editing, Project administration.

Statement of ethics

Study approval statement: This study was approved by the ethics committee of the Faculty of Psychology at Ruhr University Bochum (Reference No. 228).

Consent to participate statement

All patients gave their written informed consent.

Funding sources

Financial support was provided by unith.ev, Kinder-und Jugendlichenpsychotherapie Verhaltenstherapie e.V., the Mental Health Research and Treatment Center at Ruhr University Bochum, the Alexander von Humboldt Professorship awarded to JM, the departments of Clinical Child and Adolescent Psychology and Psychotherapy and Clinical Psychology and Psychotherapy at the University of Koblenz-Landau.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Amelie Scupin and Paul Gloyer for their help with data aggregation and preparation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2025.104691.

Data availability

Data will be made available on request.

References

Barkham, M., De Jong, K., Delgadillo, J., & Lutz, W. (2023). Routine outcome monitoring (ROM) and Feedback: Research review and recommendations. *Psychotherapy Research, 33*(7), 841–855. https://doi.org/10.1080/10503307.2023.2181114

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory–II. In *Psychological assessment.* APA PsycTests.

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

Bleichhardt, G., Timmer, B., & Rief, W. (2004). Cognitive-behavioural therapy for patients with multiple somatoform symptoms—a randomised controlled trial in tertiary care. *Journal of Psychosomatic Research, 56*(4), 449–454. https://doi.org/10.1016/S0022-3999(03)00630-5

Böhnke, J. R., Lutz, W., & Delgadillo, J. (2014). Negative affectivity as a transdiagnostic factor in patients with common mental disorders. *Journal of Affective Disorders, 166*, 270–278. https://doi.org/10.1016/j.jad.2014.05.023

Borkovec, T. D. (2002). Training clinic research and the possibility of a national training clinics practice research network. *The Behavior Therapist, 25*(98), 102.

Busner, J., & Targum, S. D. (2007). The clinical global impressions scale: Applying a research tool in clinical practice. *Psychiatry (Edgmont), 4*(7), 28.

Cahill, J., Barkham, M., & Stiles, W. B. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology, 49*(4), 421–453. https://doi.org/10.1348/014466509X470789

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry, 23*(4), 318–327.

Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology, 14*, 159–183.

Conway, S., Audin, K., Barkham, M., Mellor-Clark, J., & Russell, S. (2003). Practice-based evidence for a brief time-intensive multi-modal therapy guided by group-analytic principles and method. *Group Analysis, 36*(3), 413–435. https://doi.org/10.1177/05333164030363015

Derogatis, L. R. (1975). Brief symptom inventory. In *European journal of psychological assessment.* APA PsycTests.

Driessen, E., & Hollon, S. D. (2010). Cognitive behavioral therapy for mood disorders: Efficacy, moderators and mediators. *Psychiatria Clinica, 33*(3), 537–555. https://doi.org/10.1016/j.psc.2010.04.005

First, M. B., Williams, J. B., Karg, R. S., & Spitzer, R. L. (2016). *SCID-5-CV: Structured clinical interview for DSM-5 disorders: Clinician version (No Title)*.

Flückiger, C., Regli, D., Grawe, K., & Lutz, W. (2007). Differences and similarities between pre-post and retrospective measuring. *Psychotherapy Research, 17*(3), 359–364.

Guy, W. (1976). Clinical global impression scale. *The ECDEU Assessment Manual for Psychopharmacology-Revised Volume DHEW Publ No ADM, 76*(338), 218–222.

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy, 51*(9), 597–606. https://doi.org/10.1016/j.brat.2013.06.004

Hans, E., & Hiller, W. (2013a). A meta-analysis of nonrandomized effectiveness studies on outpatient cognitive behavioral therapy for adult anxiety disorders. *Clinical Psychology Review, 33*(8), 954–964. https://doi.org/10.1016/j.cpr.2013.07.003

Hans, E., & Hiller, W. (2013b). Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: A meta-analysis of nonrandomized effectiveness studies. *Journal of Consulting and Clinical Psychology, 81*(1), 75–88. https://doi.org/10.1037/a0031080

Hautzinger, M., Keller, F., & Kühner, C. (2023). *BDI-II. Beck-Depressions-Inventar. Revision. 3. Auflage.* Frankfurt: Pearson Assessment.

Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research, 36*(5), 427–440. https://doi.org/10.1007/s10608-012-9476-1

Holmqvist, R., Ström, T., & Foldemo, A. (2014). The effects of psychological treatment in primary care in Sweden—a practice-based study. *Nordic Journal of Psychiatry, 68*(3), 204–212. https://doi.org/10.3109/08039488.2013.797023

Hoyer, J., Velten, J., Benecke, C., Berking, M., Heinrichs, N., In-Albon, T., Lincoln, T., Lutz, W., Schlarb, A., & Schöttke, H. (2015). Koordination der Forschung an Hochschulambulanzen für Psychotherapie: Status quo und Agenda. *Zeitschrift für Klinische Psychologie und Psychotherapie, 44*(2), 80–87. https://doi.org/10.1026/1616-3443/a000308

In-Albon, T., Christiansen, H., Imort, S., Krause, K., Schlarb, A., Schneider, S., Schwarz, D., Weber, L., & Velten, J. (2019). KODAP research network: Pilot data of a project for coordinating research at university outpatient psychotherapy clinics for children and adolescents in Germany. *Zeitschrift für Klinische Psychologie und Psychotherapie, 48*(1). https://doi.org/10.1026/1616-3443/a000528

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

Kadouri, A., Corruble, E., & Falissard, B. (2007). The improved clinical global impression scale (iCGI): Development and validation in depression. *BMC Psychiatry, 7*(1), 7. https://doi.org/10.1186/1471-244X-7-7

Kaiser, T., Brakemeier, E.-L., & Herzog, P. (2023). What if we wait? Using synthetic waiting lists to estimate treatment effects in routine outcome data. *Psychotherapy Research*, 1–15. https://doi.org/10.1080/10503307.2023.2182241

Kathmann, N., Jacobi, T., Elsner, B., & Reuter, B. (2022). Effectiveness of individual cognitive-behavioral therapy and predictors of outcome in adult patients with obsessive-compulsive disorder. *Psychotherapy and Psychosomatics, 91*(2), 123–135. https://doi.org/10.1159/000520454

Keller, F., & Alexandrowicz, R. W. (2024). Assessing individual change: A comparison of reliable change indices based on classical test theory and various item response theory models. In M. Stemmler, W. Wiedermann, & F. L. Huang (Eds.), *Dependent data in social sciences research: Forms, issues, and methods of analysis* (2nd ed.). Heidelberg: Springer.

Kliem, S., & Brähler, E. (2017). *BSI Brief Symptom Inventory (Short-form of the Symptom-Checklist-90-R) von L. R. Derogatis, German version)*. Pearson Assessment.

Kroenke, K. (2007). Efficacy of treatment for somatoform disorders: A review of randomized controlled trials. *Psychosomatic Medicine, 69*(9), 881–888. https://doi.org/10.1097/PSY.0b013e31815b00c4

Lincoln, T. M., Ziegler, M., Mehl, S., Kesting, M.-L., Lüllmann, E., Westermann, S., & Rief, W. (2012). Moving from efficacy to effectiveness in cognitive behavioral therapy for psychosis: A randomized clinical practice trial. *Journal of Consulting and Clinical Psychology, 80*(4), 674. https://doi.org/10.1037/a0028665

Lutz, W., Neu, R., & Rubel, J. A. (2019). *Evaluation und Effekterfassung in der Psychotherapie* (Vol. 5). Göttingen: Hogrefe.

Lutz, W., Schiefele, A.-K., Wucherpfennig, F., Rubel, J., & Stulz, N. (2016). Clinical effectiveness of cognitive behavioral therapy for depression in routine care: A propensity score based comparison between randomized controlled trials and clinical practice. *Journal of Affective Disorders, 189*, 150–158. https://doi.org/10.1016/j.jad.2015.08.072

Margraf, J., Cwik, J. C., Pflug, V., & Schneider, S. (2017). Structured clinical interviews for mental disorders across the life span: Psychometric quality and further

developments of the DIPS open access interviews. *Zeitschrift für Klinische Psychologie und Psychotherapie, 46*(3), 176–186. https://doi.org/10.1026/1616-3443/a000430

Margraf, J., Hoyer, J., Fydrich, T., In-Albon, T., Lincoln, T., Lutz, W., Schlarb, A., Schöttke, H., Willutzki, U., & Velten, J. (2021). The cooperative revolution reaches clinical psychology and psychotherapy: An example from Germany. *Clinical Psychology in Europe, 3*(1), e4459. https://doi.org/10.32872/cpe.4459

Minami, T., Wampold, B. E., Serlin, R. C., Kircher, J. C., & Brown, G. S. J. (2007). Benchmarks for psychotherapy efficacy in adult major depression. *Journal of Consulting and Clinical Psychology, 75*(2), 232–243. https://doi.org/10.1037/0022-006X.75.2.232

Olatunji, B. O., Cisler, J. M., & Deacon, B. J. (2010). Efficacy of cognitive behavioural therapy for anxiety disorders: A review of meta-analytic findings. *Psychiatria Clinica, 33*, 557–577. https://doi.org/10.1016/j.psc.2010.04.002

Öst, L.-G., Enebrink, P., Finnes, A., Ghaderi, A., Havnen, A., Kvale, G., Salomonsson, S., & Wergeland, G. J. (2023). Cognitive behavior therapy for adult anxiety disorders in routine clinical care: A systematic review and meta-analysis. *Clinical Psychology: Science and Practice, 30*(3), 272–290. https://doi.org/10.1037/cps0000144

Pittig, A., Kotter, R., & Hoyer, J. (2019). The struggle of behavioral therapists with exposure: Self-reported practicability, negative beliefs, and therapist distress about exposure-based interventions. *Behavior Therapy, 50*(2), 353–366. https://doi.org/10.1016/j.beth.2018.07.003

Radhakrishnan, M., Hammond, G., Jones, P. B., Watson, A., McMillan-Shields, F., & Lafortune, L. (2013). Cost of improving access to psychological therapies (IAPT) programme: An analysis of cost of session, treatment and recovery in selected primary care trusts in the east of england region. *Behaviour Research and Therapy, 51*(1), 37–45. https://doi.org/10.1016/j.brat.2012.10.001

Schumm, H., Krüger-Gottschalk, A., Dyer, A., Pittig, A., Cludius, B., Takano, K., Alpers, G. W., & Ehring, T. (2022). Mechanisms of change in trauma-focused treatment for PTSD: The role of rumination. *Behaviour Research and Therapy, 148*, Article 104009. https://doi.org/10.1016/j.brat.2021.104009

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist, 50*(12), 965.

Shadish, W. R., Matt, G. E., Navarro, A. M., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin, 126*(4), 512–529. https://doi.org/10.1037//0033-2909.126.4.512

Strauss, B. M., Lutz, W., Steffanowski, A., Wittmann, W. W., Boehnke, J. R., Rubel, J., Scheidt, C. E., Caspar, F., Vogel, H., Altmann, U., Steyer, R., Zimmermann, A., Bruckmayer, E., von Heymann, F., Kramer, D., & Kirchmann, H. (2015). Benefits and challenges in practice-oriented psychotherapy research in Germany: The TK and the QS-PSY-BAY projects of quality assurance in outpatient psychotherapy. *Psychotherapy Research, 25*(1), 32–51. https://doi.org/10.1080/10503307.2013.856046

Taubitz, F.-S., Büdenbender, B., & Alpers, G. W. (2022). What the future holds: Machine learning to predict success in psychotherapy. *Behaviour Research and Therapy, 156*, Article 104116. https://doi.org/10.1016/j.brat.2022.104116

Tolin, D. F. (2010). Clinical psychology review is cognitive – behavioral therapy more effective than other therapies? A meta-analytic review. *Clinical Psychology Review, 30*(6), 710–720. https://doi.org/10.1016/j.cpr.2010.05.003

Tracey, T. J. G., Lichtenberg, J. W., Goodyear, R. K., & Wampold, B. E. (2024). Do therapists get better with experience?. In F. T. L. Leong, J. L. Callahan, J. Zimmerman, M. J. Constantino, & C. F. Eubanks (Eds.), *APA handbook of psychotherapy: Evidence-based practice, practice-based evidence, and contextual participant-driven practice* (Vol. 2, pp. 255–269) American Psychological Association. https://doi.org/10.1037/0000354-017.

Turner, D. T., Van Der Gaag, M., Karyotaki, E., & Cuijpers, P. (2014). Psychological interventions for psychosis: A meta-analysis of comparative outcome studies. *American Journal of Psychiatry, 171*(5), 523–538. https://doi.org/10.1176/appi.ajp.2013.13081159

Velten, J., Bräscher, A.-K., Fehm, L., Fladung, A.-K., Fydrich, T., Heider, J., Hentschel, S., Limberg-Thiesen, A., Lutz, W., Margraf, J., Schöttke, H., Witthöft, M., & Hoyer, J. (2018). Behandlungsdiagnosen in universitären Ambulanzen für psychologische Psychotherapie im Jahr 2016. *Zeitschrift für Klinische Psychologie und Psychotherapie, 47*, 175–185. https://doi.org/10.1026/1616-3443/a000490

Velten, J., Margraf, J., Benecke, C., Berking, M., In-albon, T., Lincoln, T., Lutz, W., Schlarb, A., Schöttke, H., Willutzki, U., & Hoyer, J. (2017). Methodenpapier zur Koordination der Datenerhebung und -auswertung an Hochschul- und Ausbildungsambulanzen für Psychotherapie (KODAP). *Zeitschrift für Klinische Psychologie und Psychotherapie, 46*(3), 169–175. https://doi.org/10.1026/1616-3443/a000431

Velten, J., Pantazidis, P., Benecke, A., Bräscher, A.-K., Fehm, L., Fladung, A.-K., … Hoyer, J. (2021). Wie häufig werden Diagnosen aus dem Bereich der sexuellen Funktionsstörungen an deutschen Hochschulambulanzen für Psychotherapie an psychologischen Instituten vergeben? *Zeitschrift für Sexualforschung, 34*(01), 5–14. https://doi.org/10.1055/a-1362-2243.

Von Brachel, R., Hirschfeld, G., Berner, A., Willutzki, U., Teismann, T., Cwik, J. C., Velten, J., Schulte, D., & Margraf, J. (2019). Long-term effectiveness of cognitive behavioral therapy in routine outpatient care: A 5- to 20-year follow-up study. *Psychotherapy and Psychosomatics, 88*(4), 225–235. https://doi.org/10.1159/000500188

von Glischinski, M., von Brachel, R., & Hirschfeld, G. (2019). How depressed is "depressed"? A systematic review and diagnostic meta-analysis of optimal cut points for the Beck depression inventory revised (BDI-II). *Quality of Life Research, 28*(5), 1111–1118. https://doi.org/10.1007/s11136-018-2050-x

Wampold, B. E., & Owen, J. (2021). Therapist effects: History, methods, magnitude. In *Bergin and garfield's handbook of psychotherapy* (pp. 297–326). John Wiley & Sons.

Westbrook, D., & Kirk, J. (2005). The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy, 43*(10), 1243–1261. https://doi.org/10.1016/j.brat.2004.09.006

Wittchen, H.-U., Wunderlich, U., Gruschwitz, S., & Zaudig, M. (1997). SKID-I. In *Strukturiertes klinisches interview für DSM-IV*.

Zala, D., Brabban, A., Stirzaker, A., Radhakrishnan, M., & Paul, K. (2019). The cost-effectiveness of the improving access to psychological therapies (IAPT) programme in severe mental illness: A decision analytical model using routine data. *Community Mental Health Journal, 55*(5), 873–883. https://doi.org/10.1007/s10597-019-00390-z