Behavioural and Cognitive Psychotherapy, 2023, **51**, 320–334
doi:10.1017/S1352465823000073

BABCP
www.babcp.com

MAIN

Effectiveness and long-term stability of outpatient cognitive behavioural therapy (CBT) for children and adolescents with anxiety and depressive disorders under routine care conditions

Daniel Walter[1,2]*, Ulrike Behrendt[2], Eva Katharina Matthias[2], Martin Hellmich[3], Lydia Dachs[2], Hildegard Goletz[2], Anja Goertz-Dorten[1,2], Daniela Perri[2], Christiane Rademacher[1,2], Stephanie Schuermann[1,2], Katrin Woitecki[2], Tanja Wolff Metternich-Kaizman[1,2] and Manfred Doepfner[1,2]

[1]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Robert-Koch-Str. 10, 50931 Cologne, Germany, [2]School of Child and Adolescent Cognitive Behavior Therapy (AKiP), Faculty of Medicine and University Hospital Cologne, University of Cologne, Pohligstr. 9, 50969 Cologne, Germany and [3]Institute for Medical Statistics and Computational Biology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany
*Corresponding author. Email: daniel.walter@uk-koeln.de

(Received 21 July 2022; revised 30 December 2022; accepted 15 February 2023; first published online 13 March 2023)

Abstract

Background: Randomised controlled trials (RCTs) have provided considerable evidence for the short-term efficacy of cognitive behavioural therapy (CBT) in children and adolescents with depressive and anxiety disorders. However, the effectiveness and long-term stability of treatment effects under routine care conditions remain unproven.

Aims: This observational study investigates the effectiveness and stability of CBT under routine care conditions within a large sample of clinically referred youth with depressive and anxiety disorders.

Method: Two hundred and twenty former patients (age 6–18 years at start of treatment) underwent a follow-up assessment (follow-up interval: $M = 5.3$ years, $SD = 2.47$). Parent and self-ratings of behavioural and emotional problems were obtained at the beginning and end of treatment and at follow-up. Additionally, at follow-up, a telephone interview and questionnaires exploring other mental symptoms and life satisfaction were administered.

Results: A repeated measures ANOVA yielded statistically significant, medium to large pre– post symptom reductions ($\eta_p^2 = .15$ to $\eta_p^2 = .47$) and small to medium post-follow-up symptom reductions ($\eta_p^2 = .03$ to $\eta_p^2 = .19$). At follow-up, between 57 and 70% of the sample reported a decrease in different emotional symptoms since the end of treatment, and 80% reported improved life satisfaction.

Conclusions: These findings provide evidence for the effectiveness and stability of treatment effects of CBT in youth with depressive and anxiety disorders under routine care conditions. Due to the lack of a direct control condition and a substantial proportion of missing data, the results must be interpreted with caution.

Keywords: child and adolescent psychotherapy; cognitive behavioural therapy; internalising disorders; routine treatment

Introduction

Depressive and anxiety disorders are among the most common mental disorders in children and adolescents between the ages of 5 and 17 years (worldwide prevalence: depressive disorder 6.2%; anxiety disorder 3.2%) (Erskine *et al.*, 2017). Affected individuals also show a considerable risk for

adverse development, such as co-morbid mental disorders, an increased risk of suicide, and substantial psychosocial impairments like school abstinence and drop-out, unemployment, and peer problems (Beesdo *et al.*, 2009; de Lijster *et al.*, 2018; Naicker *et al.*, 2013). Unsurprisingly, treatment of these young people incurs high costs for the health care system (Pella *et al.*, 2020; Ssegonja *et al.*, 2019).

Accordingly, adequate treatment for affected individuals is crucial (Lepine and Briley, 2011). Over the last decades, intensive research has been conducted to develop and examine effective treatments (Higa-McMillan *et al.*, 2016; Weersing *et al.*, 2017; Weisz *et al.*, 2017), and treatment guidelines based on empirical evidence for depression and anxiety in youth have been formulated (National Institute for Health and Care Excellence, 2014; National Institute for Health and Care Excellence, 2019). These guidelines consistently recommend cognitive behavioural therapy (CBT) as a first-line treatment for depression and anxiety. Additionally, depending on the severity of the disorder, the patients' impairment, and the effects of CBT, additional pharmacotherapy should be considered, although the empirical evidence for pharmacotherapy remains much poorer (National Institute for Health and Care Excellence, 2014; National Institute for Health and Care Excellence, 2019).

This strong evidence in favour of CBT was derived from efficacy trials using randomised controlled study designs, which are considered as the gold standard within psychotherapy research. However, their external validity has been criticised because treatment settings often differ considerably from routine care conditions (Weisz *et al.*, 2005). Accordingly, it has been argued that randomised controlled trial (RCT) results cannot be generalised to routine care situations (Bear *et al.*, 2019; Carr, 2009; Weisz *et al.*, 2013). By contrast, effectiveness studies provide a high external validity, as they can provide evidence that a particular intervention works under real-life conditions. Thus, both efficacy and effectiveness studies are needed in psychotherapy research. Many researchers call for effectiveness studies to determine the replicability of effects found in RCTs under routine care conditions (Bear *et al.*, 2019; Carr, 2009; Roest *et al.*, 2021; Weersing and Weisz, 2002; Weisz *et al.*, 2005).

There is extensive evidence to support the short-term efficacy of CBT in children and adolescents with anxiety and depressive disorders under controlled conditions (RCTs). The study findings have been aggregated in various meta-analyses, yielding moderate to high between-group effect sizes (ES) ranging from $d = 0.54$ to $d = 0.88$ (Eckshtain *et al.*, 2020; Higa-McMillan *et al.*, 2016; Kreuze *et al.*, 2018; Reynolds *et al.*, 2012; Weersing *et al.*, 2017; Weisz *et al.*, 2017; Weitz *et al.*, 2018; Wergeland *et al.*, 2021).

However, far fewer studies have examined whether the changes found in RCTs can be replicated in routine care conditions (Lee *et al.*, 2013; Mahdi *et al.*, 2018). In a large multicentre study encompassing $n = 9895$ patients (age $M = 22.7$ years, $SD = 5.3$), McAleavey *et al.* (2019) assessed the effectiveness of all forms of routine psychotherapy. The authors reported a medium pre–post reduction for depression ($d = 0.64$) and a small reduction for generalised anxiety disorder ($d = 0.47$). A recent meta-analysis encompassed 58 studies, including eight examining CBT for depression and 22 for anxiety (Wergeland *et al.*, 2021). All studies were conducted in a non-university setting such as school, community mental health centres, or other clinical routine care facilities (Wergeland *et al.*, 2021). The medium treatment duration in the studies investigating anxiety or depression was $M = 12.5$ weeks ($M = 11.7$ treatment sessions). The authors reported statistically significant, large pre–post symptom reductions in the treatment groups for depression ($d = 1.24$) and anxiety ($d = 1.32$). Overall, these results provide first evidence for the short-term effectiveness of CBT under routine care conditions in terms of reducing symptoms of internalising disorders. Nevertheless, several important limitations must be kept in mind (McAleavey *et al.*, 2019; Wergeland *et al.*, 2021). First, the treatment ingredients were poorly described or differed within studies, leaving it unclear which type of intervention was examined. Second, a substantial number of trials lacked detailed study information (e.g. information on setting or

https://doi.org/10.1017/S1352465823000073 Published online by Cambridge University Press

sample characteristics), impeding the classification of these studies as effectiveness trials. Third, most of the studies examined very small samples. Fourth, the majority of studies applied a limited treatment intensity of 12 sessions or less. Finally, the role of pharmacotherapy was frequently not reported (Wergeland *et al.*, 2021).

Beside short-term effects, the question of the stability of the obtained treatment effects is crucial. Several highly controlled efficacy studies have provided results regarding shorter stability periods, which were aggregated by Weersing *et al.* (2017) and Wergeland *et al.* (2021). The authors included studies investigating a follow-up period of 12–18 months and found remission rates to be higher at follow-up (depression 66.7%; anxiety 70.0%) than at post-treatment (depression 55.5%; anxiety 60.3%), indicating symptom improvements during the follow-up period. Less controlled effectiveness studies, which investigated all forms of psychotherapy including CBT and examined a comparable follow-up period of up to 1.5 years, mostly replicated these findings (Wergeland *et al.*, 2021).

While such results seem promising, in addition to the above-mentioned shortcomings, it is important to note some further important aspects that particularly concern the follow-up period. First, a substantial proportion of the study samples was lost to follow-up. Most of the studies dealt with this methodological problem by using the last observation carried forward method (Lachin, 2015), but this procedure leads to a bias in favour of the assumption of stability (Newgard and Lewis, 2015). Second, the amount of professional help during follow-up (i.e. further therapy, medication) mostly remains unclear. Lastly, only shorter follow-up periods with a maximum of 1.5 years were considered, while evidence on long-term effects remains unknown.

To contribute further knowledge to the research field, the present study investigated the effectiveness and long-term stability of CBT using a large, clinically referred sample, employing a broader follow-up range of 1–10 years, and applying a more conservative and elaborated method for handling missing data. Besides symptoms of depression and anxiety, further psychological symptoms and information concerning life satisfaction were rated by patients and their parents. Based on the current state of research, we hypothesised (1) statistically and clinically significant reductions in self- and parent-rated symptoms of depression, anxiety, and other behavioural and emotional symptoms during treatment; (2) no statistically significant deteriorations at follow-up assessment; and (3) statistically larger symptom reductions at post-treatment based on self- and parent ratings in patients receiving additional anti-depressant medication compared with those without pharmacotherapy.

Method

Participants

The following inclusion criteria were applied: (1) meeting the ICD-10 criteria for an anxiety and/or depressive disorder based on clinical judgement, (2) a minimum of ten treatment sessions, and (3) outpatient CBT between 2006 and 2015. Children and adolescents either self-referred to outpatient treatment or were referred by their parents, other inpatient or outpatient departments at the University of Cologne, or other clinics and private psychotherapy or psychiatry practices in the greater Cologne area. The study was approved by the ethics committee of the University of Cologne and written informed consent was obtained from all patients and parents.

Procedure

Study eligibility was assessed 1–10 weeks before the start of treatment and participants were consecutively included in the study. The first assessment took place within the first treatment

https://doi.org/10.1017/S1352465823000073 Published online by Cambridge University Press

sessions (pre-assessment) and consisted of standardised questionnaires. The second assessment took place at the end of treatment (post-assessment). Depending on the patient's age and the specialist outpatient clinic in which they were treated, an individualised set of standardised self- and parent-rated questionnaires was selected. For the follow-up assessment, patients who had completed treatment between 1 and 10 years earlier were contacted by one of the authors (U.B.), who provided information on the follow-up assessment. Standardised questionnaires were again completed, and a semi-structured telephone interview was conducted.

Measures

Basic Documentation Form

The standardised 'Basic Documentation Form' (Doepfner and Steinhausen, 2012) captures sociodemographic data such as gender and age, as well as treatment characteristics such as treatment duration and number of sessions. In addition, the form encompasses the following clinical ratings: (1) intelligence level [ranging from 1 (very high) to 8 (severely impaired)]; (2) global functioning [ranging from 0 (very good functioning in all areas) to 8 (needs persistent support 24/7)] based on Axis Six of the Multiaxial Classification of Child and Adolescent Psychiatric Disorders (World Health Organization, 1996); (3) overall clinical improvement [ranging from 1 (very much improved/remitted) to 5 (worsened)] (short version of the Clinical Global Impressions Scale-Improvement; Busner and Targum, 2007); and (4) the cooperation of children/adolescents and parents [ranging from 1 (no cooperation) to 5 (very good cooperation)].

Diagnostic interviews

All clinical diagnoses at the start of treatment were based on clinical examinations applying the clinical rating scales of the 'Diagnostiksystem für psychische Störungen nach ICD-10 und DSM-5 für Kinder und Jugendliche' (DISYPS) (Doepfner and Goertz-Dorten, 2017; Doepfner *et al.*, 2009). Good internal consistency (range from $\alpha r = .69–.95$) was found in both the clinical and the field sample. The correlations between clinical ratings according to the adolescent and the parent interviews were in the moderate range (Doepfner and Goertz-Dorten, 2017).

Telephone interview

The semi-structured interview conducted at follow-up with either the patients or their parents (if patients were minors) was designed to obtain the following information: current anxiety and depressive symptoms [clinical rating on a 5-point Likert scale ranging from −2 (large deterioration) to +2 (large improvement)] [i.e. 'Compared to treatment end, how do you currently perceive depressive symptoms of your child (such as depressive mood, reduction of energy or decrease in activity)?', 'Compared to treatment end, how do you currently perceive symptoms of anxiety of your child (such as fear of scrutiny by other people of fear of open spaces)?'], educational level (e.g. school diploma, professional training), occupational functioning (e.g. current employment, previous employment, changes of employment), current living situation, partnership, treatments due to mental health problems within the follow-up period, and delinquent behaviour (police contacts, convictions).

Parent and self-rating scales

At the pre-, post- and follow-up assessments, the German versions of the parent-rated Child Behavior Checklist (CBCL) and the self-rated Youth Self Report (YSR) (Doepfner *et al.*, 2014) were administered to assess emotional and behavioural problems. The CBCL consists of 118 items while the YSR consists of 112 items; for both forms, items are aggregated into eight

https://doi.org/10.1017/S1352465823000073 Published online by Cambridge University Press

narrowband syndrome scales and three broadband scales (Internalizing problems, Externalizing problems, Total problems). Representative German norms have been published for both parent rating and self-rating, and reliability and validity of the German versions have been demonstrated (Doepfner *et al.*, 2014). At all three assessment time points, the occurrence of specific depressive and anxiety symptoms was assessed using the following scales from the DISYPS: the FBB-DES (parent rating) and SBB-DES (self-rating) for depressive symptoms, and the FBB-ANZ (parent rating) and SBB-ANZ (self-rating) for anxiety symptoms (Doepfner and Goertz-Dorten, 2017; Doepfner, 2000).

Life satisfaction

The German-language Questionnaire on Life Satisfaction (FLZ; Fahrenberg *et al.*, 2000) is a self-rating scale used in the present study to assess different aspects of life satisfaction. The following subscales were assessed: Health, Job/Career, Financial Situation, Leisure/Recreational Activities, One's Own Person, Sexuality, Relationships with Others, Living Conditions and General Life Satisfaction. Each subscale consists of seven items that are rated on a 7-point Likert scale [ranging from 1 (very dissatisfied) to 7 (very satisfied)]. Reliability and validity of the questionnaire have been demonstrated (Fahrenberg *et al.*, 2000). The FLZ was assessed at the follow-up assessment for patients aged 16 years or over.

Brief Symptom Checklist (BSCL)

The BSCL by Franke (2017) is a self-rating questionnaire assessing mental health problems within the last 7 days. It consists of 53 items that are grouped into the following nine scales: depression, anxiety, hostility, somatisation, obsessive-compulsive, interpersonal sensitivity, anxiety, paranoid ideation and psychoticism. All items are rated on a 5-point Likert scale ranging from 0 (not at all) to 4 (very much). The BSCL was normalised on a representative sample of the population and separate norm values are available. In the present study, the BSCL was assessed at follow-up from the age of 18 years.

Outpatient treatment

Treatment took place at the university outpatient clinic of a school for child and adolescent cognitive behavioural therapy in Germany. The German health insurance system covers treatment costs up to a maximum of 100 treatment sessions of CBT based on an individualised treatment plan that has to be examined by a certified reviewer. Therapies were conducted by postgraduate students with a Master's degree (psychology or education) that were in the second half of their 5-year training in CBT. During the second half, 600 sessions of psychotherapy have to be delivered and these sessions are guided by an accredited CBT supervisor (one supervision session every four therapy sessions). Therapies were based on currently recommended cognitive behavioural methods. Information on the individual treatment modules, as rated by the therapists in the Basic Documentation Form at post-treatment, is provided in Table 1. All treatments included patient- and parent-focused interventions. Interventions with the patient included psychoeducation, graduated exposure or cognitive methods such as cognitive restructuring, whereas parent-focused interventions focused on psychoeducation, help to implement token systems or interventions enhancing the relationship of parents and patients. Fifteen per cent of patients additionally received psychopharmacological treatment (usually SSRIs). The average treatment length was $M = 15.88$ months ($SD = 7.75$, range 3–43) and the average number of sessions was $M = 41.59$ ($SD = 19.96$). The German health insurance system covered all treatment costs.

https://doi.org/10.1017/S1352465823000073 Published online by Cambridge University Press

Table 1. Most frequent interventions of the main sample ($n = 220$)

Intervention	Percentage of the total sample
Patient-focused interventions in total	100.0
Psychoeducation and cognitive methods	99.5
Token economy	79.5
Social skills training	72.7
Parent-/family-focused interventions in total	95.0
Psychoeducation and cognitive methods	94.5
Guidance to implement token economy at home	69.5
Methods to enhance the relationship between parents and children/adolescents	56.4
School-focused interventions in total	29.5
Psychoeducation and cognitive methods	25.0
Guidance to implement token economy at school	12.7
Methods to enhance the relationship between teacher and children/adolescents	3.7
Socio-therapeutic interventions in total	24.5
Counselling from social workers	11.8
Social-educational family support	7.7
Medication in total	15.5
SSRIs	5.9
Other anti-depressants	2.7

Statistical analysis

First, we checked the representativeness of the sample by comparing the data of the main sample ($n = 220$) with youth who could not be contacted and therefore did not participate ($n = 277$). The main sample of $n = 220$ was used for all further analyses. *t*-tests for dependent samples (continuous variables) or chi-squared tests (dichotomous variables) were used to compare sociodemographic and pre-assessment data in parent and self-rating and clinical ratings of treatment characteristics and effects. ES for dependent samples ($(M_{incomplete}–M_{complete})/SD_{pooled}$) (Cohen, 1988) or odds ratios were calculated to determine the magnitude of differences.

To examine overall changes from pre- to post-, pre- to follow-up, and post- to follow-up assessment, a repeated measures ANOVA was conducted. Partial eta squared was calculated to determine the magnitude of changes (Keppel, 1991). For all analyses, the significance level was set at $\alpha<5\%$.

Following the clinical interview at follow-up, a range of standardised questionnaires were administered. Due to a substantial proportion of missing data from these questionnaires at follow-up (up to 14.2 % at post-assessment and up to 88.2% at follow-up), missing data at pre-, post- and follow-up assessments were replaced using multiple imputation for the main scales of the CBCL, YSR, FBB-DES, SBB-DES, FBB-ANZ and SBB-ANZ (Rubin, 1987). All variables were imputed together and $n = 20$ datasets were created. To examine the robustness of symptom reductions in patients with complete data, the imputation data sets were subsequently post-processed by factor multiplication and addition of offsets (tipping point analysis; the symptom severity of missing data was decreased at pre-assessment and increased at post- and follow-up assessment) (Van Buuren, 2018). Imputed data were modified by up to 50% for the symptom scales and 80% for the total scales. Again, repeated measures ANOVAs were conducted to examine significance following imputation and modification.

To assess clinical relevance, we combined two criteria (Jacobson and Truax, 1991). First, we analysed whether participants had changed to normal functioning ($t<60$, Stanine<7), and second, we calculated the reliable change index (RCI; Jacobson and Truax, 1991) to analyse the statistical reliability of the changes. We conducted these analyses for the total scores of the CBCL, YSR, FBB-DES/ANZ and SBB-DES/ANZ. Patients were divided into five groups:

https://doi.org/10.1017/S1352465823000073 Published online by Cambridge University Press

Table 2. Axis I diagnoses of the main sample ($n = 220$)

Diagnoses	*n* (%)
Anxiety disorders	122 (55.5)
Depressive episodes	38 (17.3)
Mixed anxiety and depressive disorders	4 (1.8)
Depressive conduct disorders	12 (5.5)
Dysthymia	6 (2.7)
Other emotional disorders	10 (4.5)
Other mixed disorder of conduct and emotions	1 (0.5)
Obsessive-compulsive disorders	7 (3.2)
Somatoform disorders	5 (2.3)
Attention deficit hyperactivity disorders	7 (3.2)
Conduct disorders	2 (0.9)
Tic disorders	4 (1.8)
Non-organic enuresis	1 (0.5)
Eating disorders	1 (0.5)

(1) improved and clinically normalised; (2) improved but still in a clinical range; (3) unchanged and in a normal range; (4) unchanged and still in a clinical range; (5) worsened.

Finally, we examined whether patients who received monotherapy (CBT) differed from patients who additionally received pharmacotherapy regarding the total scales of the CBCL, YSR and FBB/SBB-DES/ANZ (ANCOVAs with post-assessment scores as dependent variables and pre-assessment scores of the analysed scales as covariate).

Results

Sample characteristics

In total, 497 children and adolescents fulfilled the inclusion criteria between January 2006 and December 2015 and received treatment at the outpatient unit, University Hospital Cologne. Of these, 432 (86.2%) had complete pre- and post-test data for self-ratings and parent ratings; $n = 262$ (57.6%) could not be contacted at follow-up, $n = 15$ (3.0%) actively refused to participate, and one former patient had died.

Two hundred and twenty of the 497 (44.3%) former patients (124 female; 56.4%) agreed to participate in the present study. This sample was used for the main analyses. The mean age was $M = 14.1$ years ($SD = 2.86$, range 6–18 years) at the start of treatment and $M = 20.89$ ($SD = 3.80$, range 11–28 years) at follow-up; $n = 168$ (76.4%) had an average intelligence level, $n = 37$ above-average (16.8%), and $n = 15$ below-average (6.8%). The most common ICD-10-based clinical diagnoses were (see Table 2): anxiety disorders ($n = 122$; 56.0%), depressive episodes/recurrent depressive disorders ($n = 38$; 17.0%), depressive conduct disorders ($n = 12$; 6.0%), and other emotional disorders ($n = 10$, 5.0%). Over half of the patients had more than one mental disorder: $n = 103$ (46.8%) had two and $n = 29$ (13.8%) had three or more; $n = 96$ (43.6%) patients had one or more family members with a mental disorder and $n = 68$ (31.0%) had separated parents. Global functioning at the start of treatment was: 1 (superior functioning and satisfactory; $n = 12$, 5.5%); 2 (mild impairment, $n = 28$, 12.7%); 3 (moderate impairment, $n = 79$, 35.9%); 4 (serious impairment in at least one area, $n = 79$, 35.9%); 5 (serious impairment in most of the areas, $n = 20$, 9.1%); or 6 (severe and profound impairment in most of the areas, $n = 2$, 0.9%).

Representativeness of complete data

Table S1 (see Supplementary material) shows the comparison of patients with complete data and participants with missing interview data who were excluded. Most variables showed no

https://doi.org/10.1017/S1352465823000073 Published online by Cambridge University Press

statistically significant differences between the two groups. The following statistically significant, small differences emerged: the main sample showed a lower YSR Externalizing problems score at pre-assessment ($d = -0.20$), better global functioning at post-assessment ($d = -0.23$), higher therapist-rated treatment success regarding psychological symptoms ($d = -0.22$) and overall life functioning ($d = -0.32$), and better cooperation of parents and patients ($d = 0.24$).

Treatment effectiveness

Table 3 presents changes of the main sample after imputation for all assessment points including effect sizes and modifications of the imputed data.

The results revealed highly significant parent-rated symptom reductions from pre- to post-assessment, with large ES ranging from $\eta_p^2 = .25$ to $\eta_p^2 = .47$ for parent rating and medium to large ES for adolescent rating ($\eta_p^2 = .15$ to $\eta_p^2 = .39$). After modification of the imputed data (tipping point analyses), the results remained stable [parent rating: medium to large symptom reductions ($\eta_p^2 = .18$ to $\eta_p^2 = .37$); self-rating: medium to large reductions ($\eta_p^2 = .09$ to $\eta_p^2 = .33$)].

Treatment stability and state of health at follow-up

Post- to follow-up assessment revealed statistically highly significant, small to medium symptom reductions in parent rating ($\eta_p^2 = .04$ to $\eta_p^2 = .19$) and self-rating ($\eta_p^2 = .03$ to $\eta_p^2 = .12$; YSR total scale: no significant change). Again, these results remained stable when using tipping point analyses (except for the FBB-ANZ-Total and the YSR Externalizing problems scale, which became non-significant at a 50% deterioration of the imputed data).

From pre- to follow-up assessment, statistically highly significant, large symptom reductions in parent rating ($\eta_p^2 = .13$ to $\eta_p^2 = .58$) and self-rating ($\eta_p^2 = .13$ to $\eta_p^2 = .29$), which remained stable when using tipping point analyses (except for the CBCL-Total problems scale and the FBB-DES-Total scale, which became non-significant at 50% deterioration).

The analysis of the telephone interviews revealed that 70.5% of the sample reported an improvement or large improvement in anxiety symptoms since discharge, whereas 5.4% reported a deterioration or a large deterioration (depressive symptoms: 57.3% improvement or large improvement, 7.7% deterioration or large deterioration, respectively). The overall life functioning was most often reported as largely improved (50.0%) or improved (29.1%); 43.2% of the participants reported some form of treatment due to mental health problems. Only 9.5% had no current occupation at all. Further details are presented in Table 4.

BSCL and FLZ

Results of the BSCL and FLZ at follow-up assessment are presented in Table S2 of the Supplementary material. Between 2.7 and 5.5% of the sample reported mental symptoms in the clinical range; 19.7% of the former patients reported below-average general life satisfaction.

Clinical significance

Results on the clinical significance of the changes on the total scales of the CBCL, YSR and FBB/SBB-DES/ANZ for the main sample are presented in Table 5. In parent rating, between 23.6 and 45.0% of the sample who were in the clinical range at the start of treatment were normalised at the end of treatment (self-rating: 28.2 to 66.4%). From post- to follow-up, between 10.9 and 38.6% were improved and clinically normalised according to parent rating, and between 9.0 and 25.2% according to self-rating. At the end of treatment, 18.6% remained in the clinical range on the

https://doi.org/10.1017/S1352465823000073 Published online by Cambridge University Press

Table 3. Changes in behavioural and emotional problems from pre-, post- and follow-up assessment on the broadband scales of the CBCL, YSR, and the total scales of the FBB-DES/ANZ and SBB-DES/ANZ for the main sample using imputed data ($n = 220$)

	Pre-assessment		Post-assessment		Follow-up assessment			Pre to post	Post to follow-up	Pre to follow-up	Pre to post	Post to follow-up	Pre to follow-up
	M	*SD*	*M*	*SD*	*M*	*SD*	*F*	η_p^2	η_p^2	η_p^2	50%/80%	50%/80%	50%/80%
Parent rating													
CBCL													
Internalizing problems	17.78	8.93	10.20	7.58	8.51	5.09	2664.44	.44*	.04*	.48*	.37*	.00*	.26*
Externalizing problems	10.13	8.25	6.71	5.96	3.86	3.71	1849.90	.25*	.17*	.37*	.18*	.08*	.25*
Total problems	42.48	21.21	25.38	16.90	16.62	9.54	4037.72	.47*	.19*	.58*	.34*	.01*	.32
FBB-DES	17.27	11.49	9.70	8.88	12.29	6.24	1020.13	.32*	.06*	.13*	.24*	.27*	.00
FBB-ANZ	21.14	13.46	12.12	10.24	9.71	6.472	1279.69	.38*	.04*	.38*	.30*	.00	.21*
Adolescent rating													
YSR													
Internalizing problems	19.14	10.64	11.86	9.22	13.64	8.23	1047.09	.39*	.03*	.20*	.33*	.17*	.02*
Externalizing problems	11.33	7.33	8.91	6.19	7.73	4.48	593.39	.15*	.04*	.19*	.09*	.00	.04*
Total problems	47.57	23.99	32.10	20.86	31.87	17.04	1226.97	.39*	.00	.29*	.26*	.14*	.01*
SBB-DES	21.91	14.49	11.96	10.67	15.97	6.71	990.85	.35*	.12*	.13*	.28*	.39*	.01*
SBB-ANZ	21.45	13.31	12.54	11.78	14.63	10.32	507.58	.31*	.04*	.17*	.24*	.15*	.02*

*$p < 0.001$.

Table 4. Results of the clinical interview at follow-up for the main sample ($n = 220$)

	n (%)
Further therapy within the follow-up interval	
Outpatient	72 (33.0)
Inpatient	24 (11.0)
Day clinic	13 (6.0)
Pharmacotherapy	46 (21.0)
Total	95 (43.2)
Educational attainment	
Lower-track secondary school certificate ('Hauptschulabschluss')	21 (9.5)
Medium-track secondary school certificate ('Realschulabschluss')	77 (35.0)
Entrance qualification for university of applied sciences ('Fachhochschulreife')	20 (9.1)
Higher-track secondary school certificate ('Abitur'/A-level equivalent)	59 (26.8)
University or university of applied sciences degree	10 (4.5)
No school-leaving certificate	33 (15.0)
Current occupational status	
School	67 (30.5)
Internship	5 (2.3)
Training	29 (13.2)
Employment	28 (12.7)
Studies	46 (20.9)
None	21 (9.5)
Unknown	24 (10.9)

CBCL total score (YSR: 23.9%). At follow-up, between 50.0 and 86.8% were clinically normalised according to parent rating and between 39.2 and 66.0% according to self-rating.

Effects of additional anti-depressant pharmacotherapy

A comparison between the group of patients receiving monotherapy and the group additionally receiving anti-depressant pharmacotherapy ($n = 34$; 15.5%) revealed the following statistically significant differences: patients receiving psychopharmacotherapy were older at the start of treatment ($t = -2.18$; $p < .05$; $d = -.41$) and reported higher scores at pre-assessment on the following scales: CBCL: Internalizing problems ($t = -2.239$; $p < .05$; $d = -.42$), Total problems ($t = -3.037$; $p < .01$; $d = -.57$); YSR: Internalizing problems ($t = -4.153$; $p < .001$; $d = -.83$); FBB-DES ($t = -4.426$; $p < .001$; $d = -.83$), SBB-DES ($t = -3.305$; $p < .01$; $d = -.66$), FBB-ANZ ($t = -2.00$; $p < .05$; $d = -.53$).

When pre-assessment scores were entered as covariates, almost all scales differed significantly between the groups at post-assessment, insofar as symptom scores were higher at the end of treatment in the group receiving combined treatment (CBCL Internalizing problems $F_{1,189} = 8.44$, $p < .01$, $\eta_p^2 = .23$; CBCL Total problems $F_{1,189} = 8.13$, $p < .05$, $\eta_p^2 = .04$; YSR Externalizing problems $F_{1,161} = 12.90$, $p < .001$, $\eta_p^2 = .07$; YSR Total problems $F_{1,161} = 8.80$, $p < .01$, $\eta_p^2 = .05$; FBB-DES $F_{1,182} = 7.49$, $p < .01$, $\eta_p^2 = .04$; SBB-DES $F_{1,159} = 16.36$, $p < .001$, $\eta_p^2 = 0.9$; FBB-ANZ, $F_{1,121} = 4.8$, $p < .05$, $\eta_p^2 = .04$ and SBB-ANZ $F_{1,94} = 7.58$, $p < .01$, partial $\eta_p^2 = .08$).

Discussion

The aim of the present study was to extend existing knowledge about evidence-based CBT under routine care conditions by investigating the changes in mental symptoms, and their long-term

https://doi.org/10.1017/S1352465823000073 Published online by Cambridge University Press

Table 5. Clinical significance of changes of the imputed parent and adolescent ratings of the CBCL, YSR, FBB-DES/ANZ and SBB-DES/ANZ total scales

			Worsened	Unchanged and still in a clinical range	Unchanged and in a normal range	Improved and still in a clinical range	Improved and clinically normalised
		*N**	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Parent rating							
Pre to post	CBCL	220	8 (3.6)	24 (10.9)	72 (32.7)	17 (7.7)	99 (45.0)
	FBB-DES	220	11 (5.0)	26 (11.8)	105 (47.7)	19 (8.6)	59 (26.8)
	FBB-ANZ	140	2 (1.4)	33 (23.6)	57 (40.7)	15 (10.7)	33 (23.6)
Post to follow-up	CBCL	220	26 (11.8)	3 (1.4)	106 (48.2)	0 (0.0)	85 (38.6)
	FBB-DES	220	66 (30.0)	35 (15.9)	86 (39.1)	9 (4.1)	24 (10.9)
	FBB-ANZ	140	8 (5.7)	21 (15.0)	85 (60.7)	6 (4.3)	20 (14.3)
Adolescent rating							
Pre to post	YSR	188	5 (2.7)	32 (17.0)	61 (32.5)	13 (6.9)	72 (38.3)
	SBB-DES	188	4 (2.1)	24 (12.8)	100 (53.2)	7 (3.7)	53 (28.2)
	SBB-ANZ	107	18 (16.8)	1 (0.9)	10 (9.3)	7 (6.5)	71 (66.4)
Post to follow-up	YSR	188	51 (27.1)	4 (2.1)	81 (43.1)	4 (2.1)	43 (22.9)
	SBB-DES	188	58 (30.9)	24 (12.8)	84 (44.7)	5 (2.7)	17 (9.0)
	SBB-ANZ	107	58 (54.2)	1 (0.9)	15 (14.0)	6 (5.6)	27 (25.2)

*Dependent on age at start of treatment and the specialist outpatient clinic.

stability, in a large sample of clinically referred adolescents with anxiety/depressive disorders who received CBT at a university outpatient clinic. We assessed pre- to post-, post- to follow-up and pre- to follow-up changes in parent- and adolescent-rated behavioural and emotional symptoms. Moreover, we investigated whether patients who received monotherapy (CBT) differed from those who received additional pharmacotherapy. Lastly, a standardised telephone interview was conducted at follow-up.

Overall, we found large, statistically significant reductions in depressive and anxiety symptoms as well as other emotional and behavioural symptoms during treatment. Moreover, the analysis of clinical significance revealed that a large percentage of patients were clinically normalised at the end of the treatment. These results are in line with previous efficacy trials (Eckshtain *et al.*, 2020; Higa-McMillan *et al.*, 2016; Kreuze *et al.*, 2018; Reynolds *et al.*, 2012; Weersing *et al.*, 2017; Weisz *et al.*, 2017; Weitz *et al.*, 2018; Wergeland *et al.*, 2021) and effectiveness trials (McAleavey *et al.*, 2019; Wergeland *et al.*, 2021), which reported similar symptom reductions during treatment. Therefore, the findings of our study, which investigated a large clinical sample and describes the different treatment modules, add important knowledge to the research field and support the assumption that CBT interventions may be delivered effectively in a routine care setting. This confirms our first hypothesis.

The analyses of post- to follow-up changes revealed significant, small to medium symptom reductions, therefore indicating possible evidence for the long-term stability of changes achieved during therapy as well as for additional improvements after the end of therapy. The results of the clinical interview at follow-up showed a further considerable reduction in anxiety and depressive symptoms and an improvement of daily-life functioning since the end of treatment. Moreover, according to our results on clinical significance at follow-up, a further substantial proportion of the sample that had been in the clinical range at post-assessment had transitioned into a clinically normalised range during the follow-up period. These findings support comparable effectiveness studies that examined much shorter follow-up periods, mostly of less than 12 months (James *et al.*, 2015; Wergeland *et al.*, 2021). As such, our findings add evidence for the long-term stability of changes during treatment and for further symptom reductions during follow-up. Taken together, our second hypothesis can therefore be confirmed.

https://doi.org/10.1017/S1352465823000073 Published online by Cambridge University Press

The additional information on mental symptoms and life satisfaction provided at follow-up (BSCL and FLZ) point in the same positive direction: only 6–12% of the sample reported symptoms in the clinical range, which is comparable to population-based epidemiological data (11% in the clinical range) (Franke, 2017). Moreover, most of the former patients reported average or above-average life satisfaction. Interestingly, several researchers have reported a negative relationship between internalising disorders and life satisfaction (Mahmoud *et al*., 2012; Meule and Voderholzer, 2020; Rissanen *et al*., 2011), insofar as life satisfaction improves with decreasing depressive and/or anxiety symptoms. Similarly, our follow-up assessment revealed a decrease in depressive and anxiety symptoms during treatment, a stability or further symptom decrease after the end of treatment, and high life satisfaction at the follow-up assessment. When interpreting these results concerning the follow-up period, it is important to keep in mind that about 43% of the sample reported having received further professional help due to mental symptoms in the interim, in the form of counselling, psychosocial treatments, or pharmacotherapy. Future studies therefore need to disentangle the effects of this kind of professional support during follow-up from the treatment stability of the initial therapy.

Our comparison between patients who solely received CBT and those who additionally received anti-depressant medication revealed differences on most of the examined variables at post-treatment. Adolescents additionally receiving anti-depressants showed higher self- and parent-rated mental problems at post-treatment, even after controlling for pre-treatment differences between these two groups. Our third hypothesis was therefore not confirmed. Based on our data, it seems that patients with more severe symptoms may require a combination of CBT and pharmacotherapy to achieve a significant symptom reduction, but nevertheless remain more impaired at the end of treatment compared with those receiving CBT alone (Vitiello and Ordóñez, 2016). However, it should be noted that we did not assess the exact medication dosage and were therefore unable to investigate the role of dosage effects, which may have affected our findings.

Although the present study provides important evidence on the effectiveness and stability of CBT, some important limitations should be mentioned. First, a substantial proportion of the sample did not participate in the follow-up assessment, and others participated in the interview but did not complete the questionnaires, leading to the possibility of a substantial selection bias. This large proportion of missing data is a substantial weakness of the present study. Nevertheless, our analyses of representativeness and missing data revealed small (though significant) or no differences between completers and non-completers. Moreover, our results using imputed data demonstrated a considerable robustness of the data: even after deteriorating the imputed data up to 80%, most of the findings remained statistically significant. However, we cannot definitively rule out the possibility of unsystematic missing data. Second, due to the lack of a direct control condition, we cannot rule out that the changes observed were attributable to confounding factors such as natural developmental trends. However, the 1- to 3-year stability of mental disorders in adolescents has been demonstrated in several studies, including a representative cross-sectional study in Germany consisting of almost 3000 4- to 18-year-olds, which reported no significant decreases in behavioural and emotional problems (assessed using the CBCL and YSR) with increasing age over a 2- to 3-year period (Doepfner *et al*., 1997). Third, while the therapists received guidance from supervisors and had regular discussions regarding the implementation of CBT, we did not assess treatment integrity formally. The therapies were conducted in a university outpatient clinic by therapists with advanced CBT training; therefore, future studies should examine whether the results differ for therapy delivered in outpatient units or private practice under routine care conditions. Similar analyses should examine the generalisability of the present results obtained in Germany (including a large number of treatment sessions that goes beyond treatment intensities in most other international studies) to other countries. Studies

https://doi.org/10.1017/S1352465823000073 Published online by Cambridge University Press

on differential effects should be conducted to determine the specific effect of each treatment ingredient (e.g. additional effects of parent-focused interventions). Fourth, although we assessed professional support due to mental problems during the follow-up period, we did not investigate the intensity and frequency of this support in detail. Therefore, we cannot quantify the potential effect of this professional support on the stability of the changes. Future research should prospectively assess both the type and intensity of professional support during a follow-up period.

In conclusion, the present study demonstrates the potential benefits and stability of routine CBT for youth within a natural treatment setting and in a large sample of clinically referred youth over a period of up to ten years after the end of treatment. As such, our findings support the results of RCTs that demonstrated the efficacy and shorter-term stability of CBT for children and adolescents under more controlled but less representative conditions.

Supplementary material. To view supplementary material for this article, please visit: https://doi.org/10.1017/S1352465823000073

Data availability statement. The data that support the findings of this study are available on request from the corresponding author, D.W.

Acknowledgements. None.

Author contributions. **Daniel Walter:** Conceptualization (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Project administration (equal), Supervision (lead), Validation (equal), Writing – original draft (equal), Writing – review & editing (equal); **Ulrike Behrendt:** Data curation (equal), Formal analysis (equal), Investigation (equal)Eva MatthiasFormal analysis (equal), Methodology (equal), Writing – original draft (equal), Writing – review & editing (equal); **Martin Hellmich:** Formal analysis (equal), Methodology (equal); **Lydia Dachs:** Data curation (equal); **Hildegard Goletz:** Data curation (equal); **Anja Goertz-Dorten:** Data curation (equal); **Daniela Perri:** Data curation (equal); **Christiane Rademacher:** Data curation (equal); **Stephanie Schuermann:** Data curation (equal); **Katrin Woitecki:** Data curation (equal); **Tanja Wolff Metternich-Kaizman:** Data curation (equal); **Manfred Doepfner:** Conceptualization (equal), Supervision (equal), Writing – review & editing (equal).

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflicts of interest. All authors work or worked at the School of Child and Adolescent Cognitive Behavior Therapy (AKiP).

Ethical standards. This study was conducted according to the guidelines of the Declaration of Helsinki and subsequent revisions, and was approved by the Ethics Committee of the Medical Faculty of the University of Cologne (protocol code 16-022).

References

Bear, H., Edbrooke-Childs, J., Norton, S., Krause, K., & Wolpert, M. (2019). Systematic review and meta-analysis: outcomes of routine specialist mental health care for young people with depression and/or anxiety. *Journal of the American Academy of Child and Adolescent Psychiatry*, *59*, 810–841.

Beesdo, K., Knappe, S., & Pine, D. S. (2009). Anxiety and anxiety disorders in children and adolescents: developmental issues and implications for DSM-V [Review]. *Psychiatric Clinics of North America*, *32*, 483–524. https://doi.org/10.1016/j.psc.2009.06.002

Busner, J., & Targum, S. D. (2007). The Clinical Global Impressions Scale: applying a research tool in clinical practice. *Psychiatry*, *4*, 28–37.

Carr, A. (2009). *What Works with Children, Adolescents and Adults? A Review of Research on the Effectiveness of Psychotherapy*. Routledge.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edn). Erlbaum.

de Lijster, J. M., Dieleman, G. C., Utens, E., Dierckx, B., Wierenga, M., Verhulst, F. C., & Legerstee, J. S. (2018). Social and academic functioning in adolescents with anxiety disorders: a systematic review. *Journal of Affective Disorders*, *230*, 108–117. https://doi.org/10.1016/j.jad.2018.01.008

Doepfner, M., & Goertz-Dorten, A. (2017). *DISYPS-III. Diagnostik-System für psychische Störungen nach ICD-10 und DSM-V für Kinder und Jugendliche - III* [DISYPS-III: Diagnostic system for psychiatric disorders in children and adolescents]. Hogrefe.

Doepfner, M., Goertz-Dorten, A., & Lehmkuhl, G. (2009). *Diagnostik-System für psychische Störungen nach ICD-10 und DSM-IV für Kinder und Jugendliche-II: DISYPS-II; Manual.* Huber, Hogrefe.

Doepfner, M., Plueck, J., Berner, W., Fegert, J., Huss, M., Lenz, K., . . . , & Lehmkuhl, G. (1997). Psychische Auffälligkeiten von Kindern und Jugendlichen in Deutschland - Ergebnisse einer repräsentativen Studie: Methodik, Alters-, Geschlechts- und Beurteilereffekte [Mental disturbances in children and adolescents in Germany. Results of a representative study: age, gender and rater effects]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, *25*, 218–233.

Doepfner, M., Plueck, J., Kinnen, C., & Arbeitsgruppe Deutsche Child Behavior Checklist (2014). Manual deutsche Schulalter-Formen der Child Behavior Checklist von Thomas M. Achenbach. Elternfragebogen über das Verhalten von Kindern und Jugendlichen (CBCL/6-18R), Lehrerfragebogen über das Verhalten von Kindern und Jugendlichen (TRF/6-18R), Fragebogen für Jugendliche (YSR/11-18R) [Manual of the German versions of the Child Behavior Checklist for school-aged children and adolescents by Thomas M. Achenbach. Child Behavior Checklist (CBCL/6-18R), Teacher Report Form (TRF/6-18R), Youth Self Report (YSR/11-18R)]. Hogrefe.

Doepfner, M., & Steinhausen, H. C. (2012). *Störungsübergreifende Verfahren zur Diagnostik psychischer Störungen* [Transdiagnostic instruments to assess mental health disorders]. Hogrefe.

Eckshtain, D., Kuppens, S., Ugueto, A., Ng, M. Y., Vaughn-Coaxum, R., Corteselli, K., & Weisz, J. R. (2020). Meta-analysis: 13-year follow-up of psychotherapy effects on youth depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, *59*, 45–63. https://doi.org/10.1016/j.jaac.2019.04.002

Erskine, H. E., Baxter, A. J., Patton, G., Moffitt, T. E., Patel, V., Whiteford, H. A., & Scott, J. G. (2017). The global coverage of prevalence data for mental disorders in children and adolescents. *Epidemiology and Psychiatric Sciences*, *26*, 395–402. https://doi.org/10.1017/S2045796015001158

Fahrenberg, J., Myrtek, M., Schumacher, J., & Braehler, E. (2000). *Fragebogen zur Lebenszufriedenheit (FLZ). Handanweisung*. Hogrefe.

Franke, G. H. (2017). *BSCL-53®-S. Brief Symptom-Checklist–Standard–Deutsches Manual.* Hogrefe.

Higa-McMillan, C. K., Francis, S. E., Rith-Najarian, L., & Chorpita, B. F. (2016). Evidence base update: 50 years of research on treatment for child and adolescent anxiety. *Journal of Clinical Child and Adolescent Psychology*, *45*, 91–113. https://doi.org/10.1080/15374416.2015.1046177

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*, 12–19.

James, A. C., James, G., Cowdrey, F. A., Soler, A., & Choke, A. (2015). Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane Database of Systematic Reviews* (2), CD004690. https://doi.org/10.1002/14651858.CD004690.pub4

Keppel, G. (1991). *Design and Analysis: A Researcher's Handbook*. Prentice-Hall, Inc.

Kreuze, L. J., Pijnenborg, G. H. M., de Jonge, Y. B., & Nauta, M. H. (2018). Cognitive-behavior therapy for children and adolescents with anxiety disorders: a meta-analysis of secondary outcomes. *Journal of Anxiety Disorders*, *60*, 43–57. https://doi.org/10.1016/j.janxdis.2018.10.005

Lachin, J. (2015). Fallacies of last observation carried forward analyses. *Clinical Trials*. https://doi.org/https://doi.org/10.1177/1740774515602688

Lee, C. M., Horvath, C., & Hunsley, J. (2013). Does it work in the real world? The effectiveness of treatments for psychological problems in children and adolescents. *Professional Psychology: Research and Practice*, *44*, 81.

Lepine, J. P., & Briley, M. (2011). The increasing burden of depression. *Neuropsychiatric Disease and Treatment*, *7* (suppl 1), 3–7. https://doi.org/10.2147/NDT.S19617

Mahdi, M., Jhawar, S., Benett, S.D., & Shafran, R. (2018). Cognitive behavioral therapy for childhood anxiety disorders: what happens to comorbid mood and behavioral disorders? A systematic review. *Journal of Affective Disorders*, *251*, 141–148.

Mahmoud, J. S., Staten, R. T., Hall, L. A., & Lennie, T. A. (2012). The relationship among young adlts college students' depression, anxiety, stress, demographics, life satisfaction, and coping styles. *Issues in Mental Health Nursing*, *33*, 149–156.

McAleavey, A. A., Youn, S. J., Xiao, H., Castonguay, L. G., Hayes, J. A., & Locke, B. D. (2019). Effectiveness of routine psychotherapy: method matters. *Psychotherapy Research*, *29*, 139–156. https://doi.org/10.1080/10503307.2017.1395921

Meule, A., & Voderholzer, U. (2020). Life satisfaction in persons with mental disorders. *Quality of Life Research*, *29*, 3043–3052. https://doi.org/10.1007/s11136-020-02556-9

Naicker, K., Galambos, N. L., Zeng, Y., Senthilselvan, A., & Colman, I. (2013). Social, demographic, and health outcomes in the 10 years following adolescent depression. *Journal of Adolescent Health*, *52*, 533–538. https://doi.org/10.1016/j.jadohealth.2012.12.016

National Institute for Health and Care Excellence (NICE) (2014). *Anxiety Disorders*. https://www.nice.org.uk/guidance/qs53

National Institute for Health and Care Excellence (NICE) (2019). *Depression in Children and Young People: Identification and Management*. https://www.nice.org.uk/guidance/NG134

Newgard, C., & Lewis, R. (2015). MIssing data – how to best account for what is not known. *Clinical Review & Education*, *314*, 940–941.

Pella, J. E., Slade, E. P., Pikulski, P. J., & Ginsburg, G. S. (2020). Pediatric anxiety disorders: a cost of illness analysis. *Journal of Abnormal Child Psychology*, *48*, 551–559. https://doi.org/10.1007/s10802-020-00626-7

Reynolds, S., Wilson, C., Austin, J., & Hooper, L. (2012). Effects of psychotherapy for anxiety in children and adolescents: a meta-analytic review. *Clinical Psychology Reviews*, *32*, 251–262. https://doi.org/10.1016/j.cpr.2012.01.005

Rissanen, T., Viinamäki, H., Honkalampi, K., Lehto, S. M., MHintikka, J., Saharinen, T., & Koivumaa-Honkanen, H. (2011). Long term life dissatisfaction and subsequent major depressive disorder and poor mental health. *BMC Psychiatry*, *11*, 1–6.

Roest, S. L., Siebelink, B. M., van Ewijk, H., Vermeiren, R., Middeldorp, C. M., & van der Lans, R. M. (2021). Sociodemographic and clinical characteristics in child and youth mental health; comparison of routine outcome measurements of an Australian and Dutch outpatient cohort. *Epidemiology and Psychiatric Sciences*, *30*, e74. https://doi.org/10.1017/S2045796021000652

Rubin, D. B. (1987). *Multiple Imputation for Survey Nonresponse*. New York: Wiley.

Ssegonja, R., Alaie, I., Philipson, A., Hagberg, L., Sampaio, F., Moller, M., . . . **, & Feldman, I.** (2019). Depressive disorders in adolescence, recurrence in early adulthood, and healthcare usage in mid-adulthood: a longitudinal cost-of-illness study. *Journal of Affective Disorders*, *258*, 33–41. https://doi.org/10.1016/j.jad.2019.07.077

Van Buuren, S. (2018). *Flexible Imputation of Missing Data*. CRC press.

Vitiello, B., & Ordóñez, A. E. (2016). Pharmacological treatment of children and adolescents with depression. *Expert Opinion on Pharmacotherapy*, *17*, 2273–2279. https://doi.org/10.1080/14656566.2016.1244530

Weersing, V., & Weisz, J. (2002). Community clinic treatment of depressed youth: benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology*, *70*, 299–310.

Weersing, V. R., Jeffreys, M., Do, M. T., Schwartz, K. T., & Bolano, C. (2017). Evidence base update of psychosocial treatments for child and adolescent depression. *Journal of Clinical Child and Adolescent Psychology*, *46*, 11–43. https://doi.org/10.1080/15374416.2016.1220310

Weisz, J. R., Doss, A. J., & Hawley, K. M. (2005). Youth psychotherapy outcome research: a review and critique of the evidence base. *Annual Review of Psychology*, *56*, 337–363.

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., . . . **, Fordwood, S. R.** (2017). What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. *American Psychologist*, *72*, 79–117. https://doi.org/10.1037/a0040360

Weisz, J. R., Ugueto, A. M., Cheron, D. M., & Herren, J. (2013). Evidence-based youth psychotherapy in the mental health ecosystem. *Journal of Clinical Child and Adolescent Psychology*, *42*, 274–286.

Weitz, E., Kleiboer, A., van Straten, A., & Cuijpers, P. (2018). The effects of psychotherapy for depression on anxiety symptoms: a meta-analysis. *Psychological Medicine*, *48*, 2140–2152. https://doi.org/10.1017/S0033291717003622

Wergeland, G. J. H., Riise, E. N., & Ost, L. G. (2021). Cognitive behavior therapy for internalizing disorders in children and adolescents in routine clinical care: a systematic review and meta-analysis. *Clinical Psychology Reviews*, *83*, 101918. https://doi.org/10.1016/j.cpr.2020.101918

World Health Organization (1996). *Multiaxial Classification of Child and Adolescent Psychiatric Disorders: The ICD-10 Classification of Mental and Behavioural Disorders in Children and Adolescents*. Cambridge: Cambridge University Press.

Cite this article: Walter D, Behrendt U, Matthias EK, Hellmich M, Dachs L, Goletz H, Goertz-Dorten A, Perri D, Rademacher C, Schuermann S, Woitecki K, Wolff Metternich-Kaizman T, and Doepfner M (2023). Effectiveness and long-term stability of outpatient cognitive behavioural therapy (CBT) for children and adolescents with anxiety and depressive disorders under routine care conditions. *Behavioural and Cognitive Psychotherapy* **51**, 320–334. https://doi.org/10.1017/S1352465823000073