Received: 14 June 2022 | Revised: 14 September 2022 | Accepted: 8 November 2022

DOI: 10.1002/cpp.2799

RESEARCH ARTICLE

WILEY

The effectiveness of cognitive behavioural therapy for social anxiety disorder in routine clinical practice

Nexhmedin Morina[1] | Julienne Seidemann[1] | Tanja Andor[1] | Lisa Sondern[1] | Paul-Christian Bürkner[2] | Isabelle Drenckhan[1] | Ulrike Buhlmann[1]

[1]Institute of Psychology, University of Münster, Münster, Germany

[2]Cluster of Excellence SimTech, University of Stuttgart, Stuttgart, Germany

Correspondence
Nexhmedin Morina and Ulrike Buhlmann, Clinical Psychology and Psychotherapy, University of Münster, Fliednerstraße 21, 48149 Münster, Germany.
Email: morina@uni-muenster.de and ulrike.buhlmann@wwu.de

Funding information
The authors received no external funding for this study.

Abstract

Numerous randomized controlled trials have shown cognitive behaviour therapy (CBT) to be effective in treating social anxiety disorder (SAD). Yet, less is known about the effectiveness of CBT for SAD conducted by psychotherapists in training in routine clinical practice. In this study, 231 patients with SAD were treated with CBT under routine conditions and were examined at pre- and post-treatment as well as at 6 and 12 months follow-up. We applied self-reports to assess symptoms of SAD (defined as primary outcome), depression and psychological distress (defined as secondary outcome). We conducted both completer and intent-to-treat analyses and also assessed the reliability of change with the reliable change index. Results revealed significant reductions in symptoms of SAD between pre- and post-assessments, with effect sizes ranging from $d = 0.9$ to 1.2. Depending on the SAD specific questionnaire applied, 47.8% to 73.5% of the sample showed a reliable positive change, whereas 1.9% to 3.8% showed a reliable negative change. Depressive symptoms and psychological distress also decreased significantly from pre- to post-assessment, with large effect sizes. Significant treatment gains regarding both primary and secondary outcomes were further observed at 6 and 12 months follow-up. The current findings based on a large sample of patients suggest that psychotherapists in CBT training working under routine conditions can effectively treat symptoms of SAD, depression and psychological distress.

KEYWORDS
CBT, effectiveness, psychotherapists in training, routine treatment, social anxiety disorder

1 | INTRODUCTION

Social anxiety disorder (SAD) is defined as excessive fear of negative evaluation and rejection by other people (American Psychiatric Association, 2013). This condition is one of the most prevalent mental disorders and is associated with significant impairments in various areas of functioning, reduced quality of life and high socioeconomic costs (Fehm et al., 2005; Konnopka & König, 2020; Ruscio et al., 2008; Stein et al., 2017). This condition is further associated with co-occurring depressive symptoms and high levels of psychological distress (Adams et al., 2016; Kashdan et al., 2009).

The most researched treatment for SAD is cognitive behaviour therapy (CBT), which aims at modifying maladaptive cognitions and behaviour using both cognitive (e.g. cognitive restructuring) and behavioural (e.g. exposure) strategies (Hofmann & Otto, 2017; Mayo-Wilson et al., 2014). Meta-analytic reviews have revealed large treatment effects of CBT for SAD (Carpenter et al., 2018; Mayo-Wilson et al., 2014). Several effectiveness trials have examined the extent to which treatment effects of individual CBT generalize to naturalistic clinical settings. Five effectiveness trials on individual CBT were included in the meta-analysis conducted by Hans and Hiller (2013) (Heinrichs et al., 2009; Joormann et al., 2005; Leveni et al., 2002; Lincoln et al., 2003; van Velzen et al., 1997). The authors reported combined Hedges' g effect sizes for both individual and group CBT for SAD, which ranged from 0.67 for trials reporting intention to treat (ITT) analyses, to 0.90 for those reporting completer effect size. The authors further reported that treatment gains were maintained over 12 months after treatment termination, with a tendency towards further improvement during the follow-up period (Hans & Hiller, 2013). In a trial published later, Crecelius and Hiller (2014) reported large effect sizes for treatment completers ($d = 1.0$) and medium ITT effect sizes ($d = 0.5–0.6$) at post-treatment, respectively. A comparison of post-treatment to follow-up effects resulted in nonsignificant differences. Furthermore, Hoyer et al. (2017) also reported large treatment effects among patients with SAD treated with CBT in routine care. The large effect sizes were found at both post-treatment and follow-up. Recently, Butler et al. (2021) examined the effectiveness of individual CBT for SAD in a quasi-naturalistic clinical setting and reported effect sizes ranging from $d = 0.63$ to 0.98, depending on the instrument assessing SAD symptoms. Butler et al. did not report any follow-up data. However, procedures in the trial by Butler et al. were similar to those of randomized controlled trials (e.g. treatment followed the structure of a treatment manual), so that their results may not depict the reality in routine care. Noteworthy, four of the nine trials described above were conducted with therapists in training (Crecelius & Hiller, 2014; Heinrichs et al., 2009; Lincoln et al., 2003; van Velzen et al., 1997). All of these trials reported large effect sizes following treatment, suggesting that CBT conducted by therapists in training also leads to significant treatment effects.

Relative to trials examining treatment efficacy in academic research units, effectiveness studies are characterized by a larger heterogeneity regarding both treatment as well as patient characteristics. For example, therapists treating patients under routine conditions may feel less restricted to implement procedures in a manner consistent with any prescribed protocol. Furthermore, patients in routine care might present with characteristics that may differ significantly from one effectiveness trial to another. Thus, the existing level of heterogeneity regarding treatment characteristics, patient characteristics, as well as the level of clinical experience makes it critical that we increase the number of effectiveness trials to increase the validity of data resulting from these trials. To this end, we aimed at assessing the effectiveness of CBT as provided by therapists in training under routine clinical practice conditions. As reported above, several of the effectiveness trials were conducted with therapists in training and still accomplished large treatment effects. Accordingly, we hypothesized first that patients with SAD will report significantly lower symptoms of SAD following treatment relative to pre-assessment. We further hypothesized that treatment gains will remain stable over the period of 6 and 12 months follow-up. Finally, we hypothesized that treatment will also result in significant reductions of symptoms of depression and psychological distress.

Key Practitioner Message

- Psychotherapists in cognitive behaviour therapy (CBT) training treated 231 patients with social anxiety disorder (SAD) under routine conditions
- Large reductions in symptoms of SAD, depression and psychological distress were reported
- Treatment gains were maintained at 6 and 12 months follow-up

2 | METHOD

2.1 | Participants, treatment and procedure

Eligible participants were all patients seeking treatment at the academic outpatient clinic for psychotherapy at the University of Münster, Germany, who fulfilled criteria for current SAD and were willing to receive psychological treatment for their symptoms. Noteworthy, patients with SAD were treated by psychotherapists in training regardless of their SAD symptom severity. For a detailed description of the sample selection, see the flow chart in Figure 1. A total of 313 potential participants with SAD provided their written informed consent. Following the exclusion of 82 participants (26.2%; see Figure 1), 231 patients with SAD were included in the study and were treated between 2008 and 2016. Of all patients, 143 (61.9%) presented with SAD as a primary diagnosis and the remaining 88 patients presented with SAD as a secondary diagnosis (38.1%). Altogether, 39% of patients reported involvement in previous psychological treatments. During treatment for this study, patients did not receive any other psychological treatment. Post-assessment was conducted with 83.2% of patients. In the following, we report both completer and ITT analyses. Sample characteristics are described in Table 1. In terms of severity of SAD symptoms, our sample's level of symptomatology was comparable with that of other clinical samples from effectiveness research. For example, our sample reported a mean score of on the *Social Interaction Anxiety Scale* (SIAS; see below) of 41.02 (SD = 15.71), which is similar to the SIAS mean reported in other trials using this scale (Butler et al., 2021; Heinrichs et al., 2009; Lincoln et al., 2003).

Altogether, 205 therapists with a master's degree in clinical psychology and currently in CBT training delivered treatment under close supervision by licensed psychotherapists. The training teaches delivery of preventive, diagnostic and therapeutic intervention services for

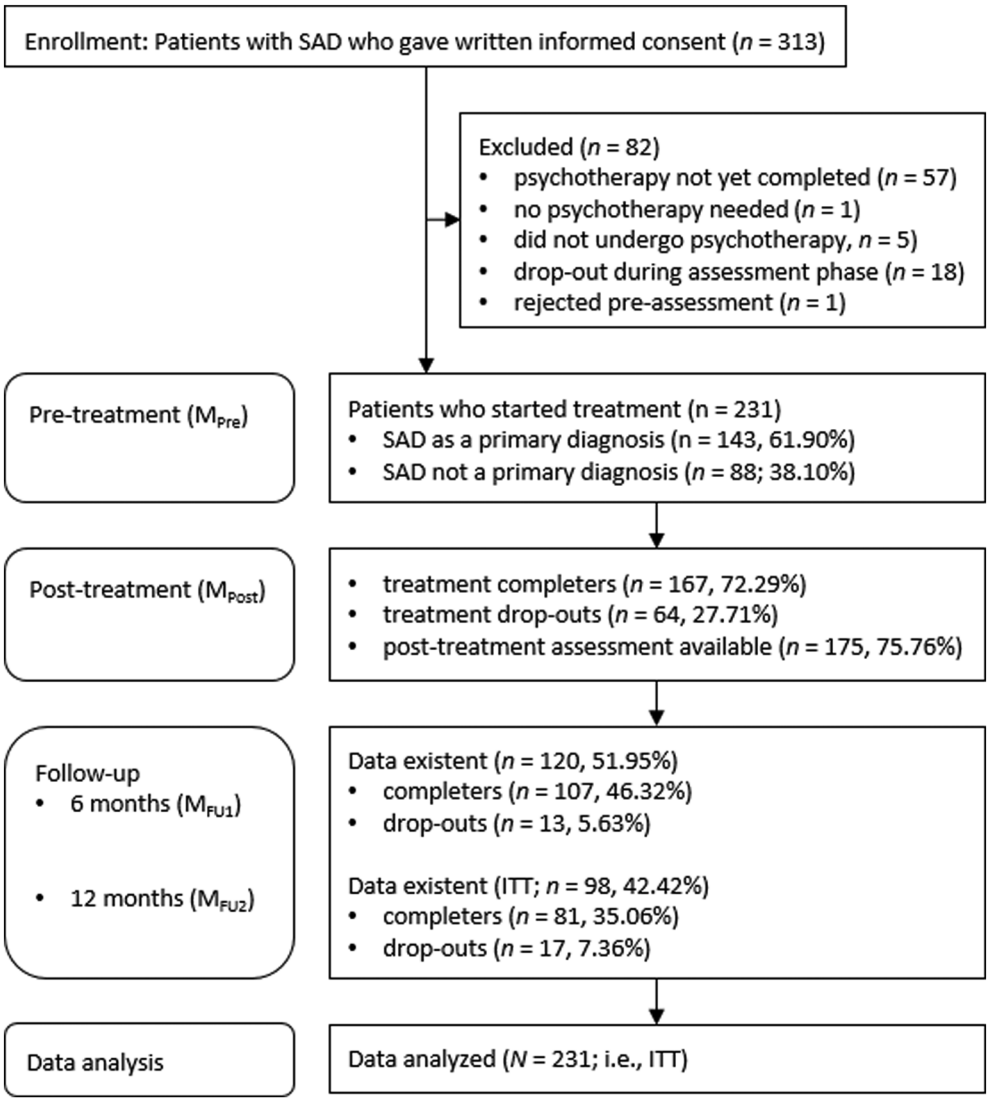

FIGURE 1 Flow chart. *Note*: ITT, intent-to-treat; SAD, social anxiety disorder

psychological and physical health and prepares the therapists to become licensed CBT psychotherapists. The therapists in this study had previously undergone an internship training programme in a psychiatric ward defined as a 1200-h experience. Furthermore, they had undergone a 2-day workshop on CBT for SAD, which was a regular part of the training. During this workshop, the cognitive therapy model of Stangier et al. (2006) was discussed as the primary example of how CBT can be applied to treat SAD. The therapists were not expected to follow a specific treatment manual, yet they were expected to use CBT techniques. To this end, supervision was provided by licensed CBT supervisors every other week. To become a licensed CBT supervisor in Germany, one has to prove a 5-year track of clinical work in CBT as well as a teaching record.

Treatment started with a detailed assessment consisting of five to seven sessions. In Germany, all patients are offered these trial sessions to enable clinical judgement about the need for treatment. Between the first and third sessions, the patients completed the pre-assessment. Diagnoses were made using the Structured Clinical Interview for DSM-IV (First, Gibbon, et al., 1997; First, Spitzer, et al., 1997). Prior to conducting the SCID for this study, the assessors had been trained on the administration of the SCID for a total of 11 h. At post-assessment and at follow-up, the same battery of self-reports used at pre-treatment was again completed by the patient. For the follow-up assessments, the patients were sent the questionnaires by mail. Similar to the approach by Heinrichs et al. (2009) and following CBT guidelines for SAD recommended by Stangier et al. (2006), we retrospectively used the patients' records to rate treatment components applied by the therapists. Initially and following a training for the purpose of this study, the first nine randomly chosen patient records were independently rated by a master's student in clinical psychology and one advanced psychotherapist in training. Based on the resulting excellent inter-rater reliability of $\kappa = .84$, the remaining patient records were rated only by the master's student. Uncertain cases ($n = 10$) were again reviewed by both raters. Results revealed that the most frequently applied treatment components were psychoeducation (90.5%), cognitive techniques (other than behavioural experiments; 87.4%), behavioural experiments (71.4%), individual social skills training (32.0%), exposure (23.4%) and relapse prevention (73.6%).

TABLE 1 Sample characteristics ($n = 231$)

Sample characteristics	*M* ± *SD* (range) or %	Missing data (%)
Age (years)	29.12 ± 9.37 (17.28–68.47)	0.00
Female	56.71%	0.00
Number of treatment sessions	39.08 ± 17.10 (3–80)	0.00
Number of assessment sessions	6.19 ± 0.60 (4–9)	5.63
Education: ≥12 years	77.49%	0.43
Psychopharmacological treatment, pre-assessment	30.30%	1.30
Comorbid diagnoses:	71.00%	0.00
• Number of comorbid diagnoses	1.16 ± 1.01 (0–4)	0.00
• Affective disorders	43.72%	0.00
• Substance use disorder (abuse or dependence)	11.25%	0.00
• Anxiety disorders (other than SAD)	22.08%	0.00
• Eating disorder	4.33%	0.00
• Somatoform disorder	3.90%	0.00
• Personality disorder	10.82%	0.00
• Obsessive–compulsive disorder	2.16%	0.00
• PTSD	3.46%	0.00

Abbreviations: PTSD, post-traumatic stress disorder; SAD, social anxiety disorder.

2.2 | Measures

2.2.1 | Primary outcome measures

We used the following self-reports as the primary outcome measures: the *Social Interaction Anxiety Scale* (German version: Stangier & Heidenreich, 1999; SIAS; Mattick & Clarke, 1998), the *Social Phobia Scale* (SPS; Mattick & Clarke, 1998; German version: Stangier & Heidenreich, 1999) and the *Self-assessment in Social Situations Questionnaire* ('*Fragebogen zur Selbstbeschreibung in sozialen Situationen*' FSSS; Kolbeck, 2008). The SPS was designed to measure specific scrutiny fears while being observed by others (e.g. drinking, eating or using public toilets). The SIAS on the other hand provides a measure of the more generalized social interaction anxieties (e.g. talking to others or attending social gatherings). Both the SPS and the SIAS include 20 items each and are rated on a 5-point rating scale. Total scores range from 0 to 80, where a higher score indicates greater severity of social anxiety. Both scales have good psychometric properties and are sensitive to treatment change (Mattick & Clarke, 1998). In our sample, the internal consistencies of the SPS and the SIAS were 0.90 and 0.91, respectively.

The FSSS measures anxiety related to social interactions and perceived competence deficiencies during the last 6 months. The scale comprises 59 items that are rated on a 4-point Likert-scale. The author of the RSSS has reported good psychometric properties (Kolbeck, 2008). The internal consistency of the FSSS in our sample was 0.96.

2.2.2 | Secondary outcome measures

To assess secondary outcomes, we used the *Symptom Checklist-90-R* (SCL-90-R; Derogatis & Cleary, 1977; German version by Franke, 2002) and the *Beck Depression Inventory* (BDI, Beck et al., 1996; German version: Hautzinger et al., 2006). The SCL-90-R comprises 90 items, and participants rate on a 5-point rating scale the degree to which they have experienced each of the respective symptoms during the past week. The total score is used as an indicator for general psychological distress. The SCL-90-R has provided good psychometric properties and sensitivity to psychotherapeutic change (Derogatis & Savitz, 1999). The internal consistency of the SCL-90-R in our sample was 0.97.

The BDI is a 21-item self-reporting questionnaire for evaluating the existence and severity of symptoms of depression, and the items are rated on a 4-point scale. The BDI has revealed very good psychometric properties, and the scale has been used widely in clinical trials (Beck et al., 1996). Initially, the original version of the BDI (Beck et al., 1987) was used. Yet, the outpatient clinic decided at a later time to replace the BDI-I with the BDI-II (Beck et al., 1996). The majority of participants (90.5%) were assessed with the first version of the BDI (which applied to all assessment points). The items of both versions are very similar, with the major difference that patients filling out the BDI-I were asked to consider each statement for the past week, whereas the BDI-II asked for potential symptoms during the past 2 weeks. When considering the reliability of change assessed with the Reliable Change Index (RCI; Jacobson & Truax, 1992; see below), we relied with respect to the BDI-I on the meta-analysis by Seggar et al. (2002) that reported an RCI of 8.46 (rounded to 9). For the BDI-II, we applied an RCI of 7.7 (rounded to 8) based on results reported by Hautzinger et al. (2006). The internal consistency of the BDI in our sample was 0.87.

2.3 | Statistical analysis

To assess potential differences from pre- to post-assessment or follow-up, we performed dependent t-tests. We further assessed the course of symptomatology over all four measurement points by using multilevel models. We calculated the magnitude of treatment effect (Cohen's d) by subtracting the post-test mean score of (*Mpost*) from the pre-test mean score (*Mpre*) and dividing the result by the pooled standard deviation (*SDpool*). Effect sizes may be conservatively interpreted with Cohen's convention of small (0.2), medium (0.5) and large (0.8) effects (Cohen, 2013).

We calculated the reliability of change using the *RCI* (Jacobson & Truax, 1992). To calculate the RCI, the post-treatment score is

TABLE 2 Descriptive values, *t*-tests of dependent data, effect sizes, RCI, response, deterioration and relapse rates

		Pre–post											Post-FU2[a]	
Scale	Sample	*n*	M_{pre}	SD_{pre}	M_{post}	SD_{post}	*t*	*df*	*P*	*ES*	RCI[b]	Response %	Deterioration %	Relapse %	Response %
SPS	Compl.	134	30.11	14.45	16.95	12.90	11.10	133	<.001	0.96	13.09	47.01	0.75	1.52	12.12
	ITT	231	31.27	14.93	17.61	12.55	13.89	230	<.001	0.99		47.84	2.62	10.30	18.68
SIAS	Compl.	134	41.02	15.71	27.38	13.67	12.23	133	<.001	0.93	11.47	52.24	0.75	6.06	15.15
	ITT	231	42.05	15.63	27.10	13.64	15.79	230	<.001	0.96		55.28	1.86	17.12	19.68
FSSS	Compl.	132	96.95	29.60	61.46	29.26	15.00	131	<.001	1.21	17.50	75.00	2.27	17.19	20.31
	ITT	231	100.33	31.56	64.62	30.09	17.74	230	<.001	1.22		73.53	3.81	27.25	29.39
SCL-90-R	Compl.	141	0.92	0.59	0.41	0.33	10.51	140	<.001	1.06	0.27	60.28	3.55	20.90	5.97
	ITT	231	0.93	0.56	0.44	0.35	12.91	230	<.001	1.04		60.65	4.72	24.65	17.34
BDI	Compl.	143	15.01	9.25	5.57	5.63	12.80	142	<.001	1.23	9/8[c]	44.76	0	11.69	1.30
	ITT	231	15.71	8.85	6.27	5.83	16.17	230	<.001	1.26		47.25	0.58	18.07	7.75

Abbreviations: BDI, Beck Depression Inventory; Compl., completer sample; Deterioration, deterioration rate; ES, effect size; FSSS, 'self-assessment in social situations'; ITT, intent-to-treat sample; RCI, Reliable Change Index, calculated by SE_{diff} (the standard error of differences) of the imputed ITT-sample x 1.96; Relapse, relapse rate; Response, response rate; SIAS, Social Interaction Anxiety Scale; SCL-90-R, Symptom Checklist-90-R; SPS, Social Phobia Scale.

[a]For all patients with post-treatment measurements but without FU2 measurements, the data were multiple imputed. The number of FU2 data existent for the ITT-samples: $n = 83$ SPS and SIAS and $n = 79$ FSSS, $n = 81$ SCL-90-R, $n = 90$ BDI.

[b]The calculated RCI score is used as a standard to inform how many patients reported a positive treatment response or deterioration following treatment.

[c]RCI was 9 for BDI-I and 8 for BDI-II.

subtracted from the pre-treatment score, and the result is divided by the standard error of the differences. If the patient's RCI is larger than the *z*-score desired level of significance (set at 1.96, $p < 0.05$), the change in pre-to-post-treatment scores is defined as a reliable positive change (i.e. treatment response). A patient with an RCI smaller than −1.96 is considered to have experienced a negative change (i.e., deterioration), and patients with a RCI between these two values are considered to have experienced no reliable change. Note that the RCIs for the BDI scores were not calculated based on the present data but rather on representative large clinical samples were given (see above). We further applied the RCI to examine the number of patients reporting relapse or response from post-treatment to follow-up. In this regard, the follow-up score was subtracted from the post-treatment score, and the result was divided by the standard error of the differences. Response was defined as reporting an RCI larger than 1.96, and relapse was defined as reporting an RCI smaller than −1.96 (see also Table 2).

Missing values of post-assessment nonresponders were handled by conducting multiple imputation using the R package 'mice' (van Buuren & Groothuis-Oudshoorn, 2011). Missing values were imputed 20 times. In case single items were missing in a questionnaire, we used the method of *response function imputation* (Sijtsma & van der Ark, 2003).

3 | RESULTS

3.1 | Comparison of completers versus drop-outs

Of 231 patients who started treatment, 64 patients (27.71%) terminated treatment earlier than clinically indicated. Of these 64 patients, 29.2% reported motivational reasons for their choice to terminate treatment early, 16.7% reported dissatisfaction with treatment and 10.4% of patients failed to keep therapeutic agreements. An additional 25.0% of all drop-outs were categorized as not related to treatment quality as they occurred for reasons such as change of residence (68.8%) or because the therapist recommended an alternative type of treatment (12.5%).

Two logistic regression analyses of the drop-out-status and socio-demographic and clinical or treatment characteristics showed in sum eight relevant predictors. Younger age ($b = -0.08$, $p < .01$; $OR = 0.92$), part-time employment (compared to full-time employment; $b = 2.77$, $p < .01$; $OR = 15.95$), a comorbid affective disorder ($b = 1.13$, $p < .05$; $OR = 3.11$), lower motivation at the beginning of treatment ($b = -1.25$, $p < .01$; $OR = 0.29$), fewer treatment sessions ($b = -0.07$, $p < .001$; $OR = 0.93$) and number of treatment elements ($b = -0.62$, $p < .01$; $OR = 0.54$) were associated with higher risk of drop-out. Furthermore, patients with no behavioural experiments ($b = -0.99$, $p < .05$; $OR = 0.373$) or no relapse prevention ($b = -3.28$, $p < .001$; $OR = 0.04$) belonged more often to the drop-out group rather than the completer group. However, there were no significant differences regarding symptoms of SAD between completers and drop-outs neither at post-treatment nor at follow-up (independent *t*-tests, all *p*s > .50).

3.2 | Treatment effectiveness and reliable change

All calculated effect sizes, response and deterioration rates are presented in Tables 2. For all measured symptoms, paired *t*-tests of both completer and ITT samples showed statistically significant symptom reduction (all *p*s < .01). Regarding the self-reported primary outcome measures (SPS, SIAS and FSSS), the effect sizes ranged from $d = 0.9$ to 1.2 for the completer and the imputed ITT samples, respectively. Depending on the disorder specific questionnaire applied, 47.8% to 73.5% of the imputed ITT-sample showed a reliable positive change, whereas 1.9% to 3.8% showed a reliable negative change. Similarly, 47% to 75% of the completer sample showed a reliable positive change, whereas 0.8% to 2.3% showed a reliable negative change.

Regarding the secondary outcome measures, the effect size on psychological distress was $d = 1.0$ for the completer and the imputed ITT samples, respectively. A reliable positive change was reported by 60.7% of the imputed ITT-sample, whereas no patient showed a negative change. The amount of depression changed with an effect size of $d = 1.2$ (completers) or $d = 1.3$ (imputed ITT-sample), respectively. Of these, 47.3% of the imputed ITT-sample showed a reliable positive change and 0.6% a negative change.

3.3 | Maintenance of treatment gains

With respect to follow-up, SAD symptoms as assessed with all three primary outcome measures tended to fall from post-treatment to the second follow-up (FU2). For the SPS, the symptom reduction from post-treatment to FU2 was significant ($t[39.126] = 2.195$, $p < .05$). However, the changes from post-treatment to FU1 and FU2 on the other self-reports were nonsignificant (all *p*s > .05), suggesting overall stable treatment outcomes.

The relapse and response rates are also reported in Tables 2. On the primary outcome measures, 18.7% to 29.4% of the imputed ITT-sample showed a reliable positive change between post-treatment and the FU2 measurement point. On the other hand, 10.3% to 27.3% of the imputed ITT-sample showed a reliable negative change. Following the development of all patients with a regularly ended therapy (completers), only 1.5% to 17.2% showed a reliable negative change and 12.1% to 20.3% showed a reliable positive one. Based on these values, 82.8% up to 98.5% of all completers were able to maintain or even improve their post-treatment outcome during the following 1-year time period.

Regarding psychological distress, 17.3% of the ITT sample and 6.0% of the completer sample showed a reliable positive change, whereas 24.7% of the ITT sample and 20.9% of the completer sample showed a reliable negative change, respectively. Regarding depression, 7.8% of the ITT sample and 1.3% of the completer sample showed a reliable positive change, whereas 18.1% of the ITT sample and 11.7% of the completer sample showed a reliable negative change, respectively. Altogether, 79.1% and 88.3% of all completers were able to maintain or even improve their general psychopathology or depression, respectively.

4 | DISCUSSION

We aimed at examining the effectiveness of CBT for SAD. Our results suggest that CBT utilized under routine clinical settings significantly reduces SAD symptoms, with treatment gains reported at both post-treatment and at follow-up. Patients also reported significant reduction of depressive symptoms and overall psychological distress.

The large treatment effects found in our study replicate and extend those reported by several effectiveness trials (Crecelius & Hiller, 2014; Gaston et al., 2006; Heinrichs et al., 2009; Lincoln et al., 2003). The overall efficacy in our trial needs to be interpreted in light of the fact that 71% of patients had at least one comorbid disorder, with 38% of participants even reporting another mental disorder as a primary diagnosis. This shows that CBT significantly reduces SAD symptoms even in the presence of other psychological complaints. Perhaps even more relevant is the finding that treatment CBT also resulted in significant reduction of symptoms of depression and overall psychological distress. Furthermore, these results were achieved by CBT therapists in training. Note that other effectiveness studies have also made use of therapists in training and reported large effect sizes (Crecelius & Hiller, 2014; Heinrichs et al., 2009; Lincoln et al., 2003). Obviously, a large number of patients is being treated every year by therapists in training. Accordingly, the findings that therapists in training can achieve large treatment effects is very reassuring to patients entering psychological treatment by therapists in training.

The positive treatment effects were further supported by the number of patients showing a reliable positive change according to the RCI. Depending on the SAD instrument, up to 73.5% of the patients reported a reliable positive change, whereas up to 3.8% showed a negative change. With respect to symptoms of psychological distress, 60.7% of patients showed a reliable positive change, and none showed a reliable negative change. Finally, 47.3% of patients reported a reliable positive change on their depressive symptoms and only 0.6% a reliable negative change. Altogether, our findings lend credit to the reported CBT efficacy as assessed in randomized controlled trials (Carpenter et al., 2018; Mayo-Wilson et al., 2014). As such, they clearly suggest that the results achieved using an evidence-based treatment developed in RCT designs generalize to a more naturalistic setting.

A significant difference between our effectiveness trial and the RCTs conducted in academic settings relates, however, to treatment duration. While patients in our trial received an average of 39 sessions, Mayo-Wilson et al. (2014) reported in their meta-analyses on the efficacy of RCTs for SAD an average of 12 sessions. It is noteworthy that effectiveness trials have generally reported a high number of sessions than that reported in RCTs. Boettcher et al. (2020) reported an average of 63 sessions, whereas other trials have reported an average number of sessions ranging from 25 to 33 (Crecelius & Hiller, 2014; Gaston et al., 2006; Hoyer et al., 2017; McEvoy, 2007). The longer duration in the effectiveness trials might be required if patients present with a multitude of problems that need a longer time to be addressed. The multitude of problems can also explain the wide variation in number of treatment sessions, which in our trial ranged from 3 to 80. Overall, 71% of patients in our trial reported comorbid disorders. Note that other German effectiveness trials have also reported a wide variation in number of sessions. For example, Crecelius and Hiller (2014) reported a rage of 1 to 80. The variation in number of sessions may also arise from the fact that in our trial therapists and patients did not commit to a specific number of sessions at the beginning of treatment. This is in contrast to efficacy trials where an upper limit of sessions is regularly communicated to patients at the beginning of treatment. For example, several efficacy trials with patients with SAD in Germany included a maximum of 25 sessions and achieved large effect sizes (e.g. Hoyer et al., 2017; Leichsenring et al., 2013; Stangier et al., 2003). The lack of commitment to a limited number of treatment sessions might negatively influence behaviour change motivation with some patients, which may offer an explanation as to why some patients needed up to 80 sessions. Yet, we did not examine this hypothesis in our trial. Accordingly, it remains for future studies to examine reasons behind longer duration of SAD treatments in routine care. A recent systematic review on the dose–response effect in routinely delivered psychological interventions for different mental disorders suggests an optimal dose of treatment ranging between 4 and 26 sessions (Robinson et al., 2020). However, none of the 26 included studies had been conducted with SAD patients. In fact, there is lack of research on the dose–response effect with clinical samples with severe mental disorders altogether (Robinson et al., 2020). Relatedly, we need to increase our knowledge about the group of patients in routine care that might need a substantially longer intervention relative to other patients. Accordingly, it remains for future research to examine the optimal dose–effect relationship in patients with SAD treated in routine outpatient settings as well as means to adjust treatment by addressing potential obstacles to improvement. Furthermore, future research needs to examine potential mechanisms of change in routine settings, a topic that we did not investigate in our trial.

The results of our study may be limited by the lack of a control group, thus precluding us from drawing definite conclusions on the relative treatment efficacy. Yet, the current effect sizes are comparable with those of previous effectiveness trials, and more importantly, they exceed those of waitlist comparison groups. For example, a meta-analysis by Steinert et al. (2017) that investigated the effects occurring in waitlist control groups in randomized controlled trials for patients with SAD reported a pooled effect of $d = 0.13$, suggesting no significant symptom reduction. In comparison, the reduction of symptoms of SAD in our trial resulted in large effect sizes, ranging from $d = 0.9$ to 1.2. Accordingly, we have reason to claim that CBT conducted in routine settings is significantly more effective in reducing SAD than no intervention. Another limitation is that we used the patient's chart as source of information for rating what therapy elements were applied in the respective treatment. Some therapists might have used some treatment technique that they failed to report in the patient's chart. Future research should use a more detailed recording device, such as a standardized and detailed questionnaire following each session. Furthermore, the rate of missing data was relatively high. We dealt with this limitation by reconstructing missing

measurements using multiple imputation. Relatedly, 28% of patients terminated treatment early. Yet, similar rates have also been reported by previous trials (Crecelius & Hiller, 2014; Hoyer et al., 2017). Another limitation is that we did not control for the effects of unspecific treatment factors (e.g. therapeutic alliance). Finally, no clinician-administered measures were conducted by independent clinical raters.

In conclusion, our findings provide evidence that CBT can effectively reduce SAD and comorbid symptoms in naturalistic settings. The findings further demonstrate that therapists in training can effectively treat patients with SAD. The validity of the current findings is supported by calculating and reporting reliability of change scores as well as by assessing treatment gains up to 1 year after treatment. Future research needs to investigate which therapeutic elements are mostly associated with behaviour change as well as the needed amount of sessions to achieve a reliable change of SAD and related symptoms.

ACKNOWLEDGEMENT

Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no competing interests.

ETHICS STATEMENT

The study protocol was approved by the local research ethics committee at the University of Munster. All participants provided written informed consent prior to participating in the study.

DATA AVAILABILITY STATEMENT

Upon publication of the results, the datasets generated and/or analysed during the current study will be available from the corresponding authors on reasonable request.

ORCID

Nexhmedin Morina https://orcid.org/0000-0002-2331-9140

REFERENCES

Adams, G. C., Balbuena, L., Meng, X., & Asmundson, G. J. G. (2016). When social anxiety and depression go together: A population study of comorbidity and associated consequences. *Journal of Affective Disorders*, *206*, 48–54. https://doi.org/10.1016/j.jad.2016.07.031

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Pub.

Beck, A. T., Steer, R. A., & Brown, G. K. (1987). *Beck depression inventory*. Harcourt Brace Jovanovich.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory (BDI-II)* (Vol. 10). Pearson.

Boettcher, J., Weinbrecht, A., Heinrich, M., & Renneberg, B. (2020). Die Behandlung der sozialen Angststörung und ängstlich-vermeidenden Persönlichkeitsstörung in der Versorgung: eine naturalistische Studie zu einer kombinierten Einzel-und Gruppentherapie. *Verhaltenstherapie*, *30*(3), 189–199. https://doi.org/10.1159/000497620

Butler, R. M., O'Day, E. B., Swee, M. B., Horenstein, A., & Heimberg, R. G. (2021). Cognitive behavioral therapy for social anxiety disorder: Predictors of treatment outcome in a quasi-naturalistic setting. *Behavior Therapy*, *52*(2), 465–477. https://doi.org/10.1016/j.beth.2020.06.002

Carpenter, J. K., Andrews, L. A., Witcraft, S. M., Powers, M. B., Smits, J. A. J., & Hofmann, S. G. (2018). Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. *Depression and Anxiety*, *35*(6), 502–514. https://doi.org/10.1002/da.22728

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Taylor and Francis. http://gbv.eblib.com/patron/FullRecord.aspx?p=1192162

Crecelius, V., & Hiller, W. (2014). Die Evaluation von Therapieeffekten bei sozialen Angststörungen in einer Hochschulambulanz mit verhaltenstherapeutischem Schwerpunkt. *Zeitschrift für Klinische Psychologie Und Psychotherapie*, *43*, 112–120. https://doi.org/10.1026/1616-3443/a000256

Derogatis, L. R., & Cleary, P. A. (1977). Confirmation of the dimensional structure of the SCL-90: A study in construct validation. *Journal of Clinical Psychology*, *33*(4), 981–989. https://doi.org/10.1002/1097-4679(197710)33:4<981::AID-JCLP2270330412>3.0.CO;2-0

Derogatis, L. R., & Savitz, K. L. (1999). *The SCL-90-R, brief symptom inventory, and matching clinical rating scales*.

Fehm, L., Pelissolo, A., Furmark, T., & Wittchen, H.-U. (2005). Size and burden of social phobia in Europe. *European Neuropsychopharmacology*, *15*(4), 453–462. https://doi.org/10.1016/j.euroneuro.2005.04.002

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

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *User's guide for the structured clinical interview for DSM-IV axis I disorders: SCID-I*. American Psychiatric Press.

Franke, G. H. (2002). *Symptom-checkliste von LR Derogatis: SCL-90-R; manual*. Beltz Test.

Gaston, J. E., Abbott, M. J., Rapee, R. M., & Neary, S. A. (2006). Do empirically supported treatments generalize to private practice? A benchmark study of a cognitive-behavioural group treatment programme for social phobia. *British Journal of Clinical Psychology*, *45*(1), 33–48. https://doi.org/10.1348/014466505X35146

Hans, E., & Hiller, W. (2013). 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

Hautzinger, M., Keller, F., & Kühner, C. (2006). *Beck depressions inventar (BDI-II)*. Harcourt Test Services.

Heinrichs, N., Nowak, C., Oertel, J., & Kröger, C. (2009). Die Wirksamkeit ambulanter Psychotherapie der Sozialen Angststörung in einer universitären Ambulanz: Wird die Forschung in die Praxis transportiert? *Zeitschrift für Klinische Psychologie Und Psychotherapie*, *38*(3), 181–193. https://doi.org/10.1026/1616-3443.38.3.181

Hofmann, S. G., & Otto, M. W. (2017). *Cognitive behavioral therapy for social anxiety disorder: Evidence-based and disorder specific treatment techniques*. Routledge.

Hoyer, J., Čolić, J., Pittig, A., Crawcour, S., Moeser, M., Ginzburg, D., Lin, J., Wiltink, J., Leibing, E., & Stangier, U. (2017). Manualized cognitive therapy versus cognitive-behavioral treatment-as-usual for social anxiety disorder in routine practice: A cluster-randomized controlled trial. *Behaviour Research and Therapy*, *95*, 87–98. https://doi.org/10.1016/j.brat.2017.05.012

Jacobson, N. S., & Truax, P. (1992). *Clinical significance: A statistical approach to defining meaningful change in psychotherapy research*.

Joormann, J., Kosfelder, J., & Schulte, D. (2005). The impact of comorbidity of depression on the course of anxiety treatments. *Cognitive Therapy and Research*, *29*(5), 569–591. https://doi.org/10.1007/s10608-005-3340-5

Kashdan, T. B., Morina, N., & Priebe, S. (2009). Post-traumatic stress disorder, social anxiety disorder, and depression in survivors of the Kosovo war: Experiential avoidance as a contributor to distress and quality of

life. *Journal of Anxiety Disorders*, *23*(2), 185–196. https://doi.org/10.1016/j.janxdis.2008.06.006

Kolbeck, S. (2008). *Zur psychometrischen Differenzierbarkeit von sozialen Ängsten und sozialen Defiziten. Eine empirische Studie an nichtklinischen und klinischen Stichproben*. Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky. BibTeX.

Konnopka, A., & König, H. (2020). Economic burden of anxiety disorders: A systematic review and meta-analysis. *PharmacoEconomics*, *38*(1), 25–37. https://doi.org/10.1007/s40273-019-00849-7

Leichsenring, F., Salzer, S., Beutel, M. E., Herpertz, S., Hiller, W., Hoyer, J., Huesing, J., Joraschky, P., Nolting, B., Poehlmann, K., Ritter, V., Stangier, U., Strauss, B., Stuhldreher, N., Tefikow, S., Teismann, T., Willutzki, U., Wiltink, J., & Leibing, E. (2013). Psychodynamic therapy and cognitive-behavioral therapy in social anxiety disorder: A multicenter randomized controlled trial. *American Journal of Psychiatry*, *170*(7), 759–767. https://doi.org/10.1176/appi.ajp.2013.12081125

Leveni, D., Piacentini, D., & Campana, A. (2002). Effectiveness of cognitive-behavioural treatment in social phobia: A description of the results obtained in a public mental health service. *Epidemiology and Psychiatric Sciences*, *11*(2), 127–133. https://doi.org/10.1017/S1121189X00005583

Lincoln, T. M., Rief, W., Hahlweg, K., Frank, M., von Witzleben, I., Schroeder, B., & Fiegenbaum, W. (2003). Effectiveness of an empirically supported treatment for social phobia in the field. *Behaviour Research and Therapy*, *41*(11), 1251–1269. https://doi.org/10.1016/S0005-7967(03)00038-X

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, *36*(4), 455–470. https://doi.org/10.1016/S0005-7967(97)10031-6

Mayo-Wilson, E., Dias, S., Mavranezouli, I., Kew, K., Clark, D. M., Ades, A. E., & Pilling, S. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: A systematic review and network meta-analysis. *The Lancet Psychiatry*, *1*(5), 368–376. https://doi.org/10.1016/S2215-0366(14)70329-3

McEvoy, P. M. (2007). Effectiveness of cognitive behavioural group therapy for social phobia in a community clinic: A benchmarking study. *Behaviour Research and Therapy*, *45*(12), 3030–3040. https://doi.org/10.1016/j.brat.2007.08.002

Robinson, L., Delgadillo, J., & Kellett, S. (2020). The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research*, *30*(1), 79–96. https://doi.org/10.1080/10503307.2019.1566676

Ruscio, A. M., Brown, T. A., Chiu, W. T., Sareen, J., Stein, M. B., & Kessler, R. C. (2008). Social fears and social phobia in the USA: Results from the National Comorbidity Survey Replication. *Psychological Medicine*, *38*(1), 15–28. https://doi.org/10.1017/S0033291707001699

Seggar, L. B., Lambert, M. J., & Hansen, N. B. (2002). Assessing clinical significance: Application to the Beck depression inventory. *Behavior Therapy*, *33*(2), 253–269. https://doi.org/10.1016/S0005-7894(02)80028-4

Sijtsma, K., & van der Ark, L. A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research*, *38*(4), 505–528. https://doi.org/10.1207/s15327906mbr3804_4

Stangier, U., Clark, D. M., & Ehlers, A. (2006). *Soziale phobie*. Hogrefe Verlag.

Stangier, U., Heidenreich, T., Berardi, A., Golbs, U., & Hoyer, J. (1999). Die erfassung sozialer phobie durch social interaction anxiety scale (SIAS) und die social phobia scale (SPS). *Zeitschrift für Klinische Psychologie*, *28*, 28–36. https://doi.org/10.1026//0084-5345.28.1.28

Stangier, U., Heidenreich, T., Peitz, M., Lauterbach, W., & Clark, D. M. (2003). Cognitive therapy for social phobia: Individual versus group treatment. *Behaviour Research and Therapy*, *41*(9), 991–1007. https://doi.org/10.1016/S0005-7967(02)00176-6

Stein, D. J., Lim, C. C. W., Roest, A. M., de Jonge, P., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Benjet, C., Bromet, E. J., Bruffaerts, R., de Girolamo, G., Florescu, S., Gureje, O., Haro, J. M., Harris, M. G., He, Y., Hinkov, H., Horiguchi, I., Hu, C., … WHO World Mental Health Survey Collaborators. (2017). The cross-national epidemiology of social anxiety disorder: Data from the world mental health survey initiative. *BMC Medicine*, *15*(1), 143. https://doi.org/10.1186/s12916-017-0889-2

Steinert, C., Stadter, K., Stark, R., & Leichsenring, F. (2017). The effects of waiting for treatment: A meta-analysis of waitlist control groups in randomized controlled trials for social anxiety disorder. *Clinical Psychology & Psychotherapy*, *24*(3), 649–660. https://doi.org/10.1002/cpp.2032

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

van Velzen, C. J. M., Emmelkamp, P. M. G., & Scholing, A. (1997). The impact of personality disorders on behavioral treatment outcome for social phobia. *Behaviour Research and Therapy*, *35*(10), 889–900. https://doi.org/10.1016/S0005-7967(97)00052-1

How to cite this article: Morina, N., Seidemann, J., Andor, T., Sondern, L., Bürkner, P.-C., Drenckhan, I., & Buhlmann, U. (2022). The effectiveness of cognitive behavioural therapy for social anxiety disorder in routine clinical practice. *Clinical Psychology & Psychotherapy*, 1–9. https://doi.org/10.1002/cpp.2799