Clinical supervision in cognitive behavior therapy improves therapists' competence: a single-case experimental pilot study

Sven Alfonsson , Tobias Lundgren & Gerhard Andersson

Routledge
Taylor & Francis Group

Clinical supervision in cognitive behavior therapy improves therapists' competence: a single-case experimental pilot study

Sven Alfonsson [a,b], Tobias Lundgren [a] and Gerhard Andersson [a,c]

[a]Department of Clinical Neuroscience Karolinska Institutet & Stockholm Health Care Services, Centre for Psychiatry Research, Stockholm, Sweden; [b]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden; [c]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

ABSTRACT

Clinical supervision is a cornerstone in psychotherapists' training but there are few empirical evaluations on the effects of supervision on therapists' competencies. The aim of this study was therefore to evaluate the effects of standardized supervision on rater-assessed competency in Cognitive Behavior Therapy (CBT). Six therapists with basic training in CBT were provided with protocol-based clinical supervision in CBT in a single-case experimental multiple baseline design. The supervision focused on specific CBT competencies and used experiential learning methods such as role-play. Each therapist recorded weekly treatment sessions during phases without and with supervision. The therapists' CBT competence was assessed by third-party raters using the Revised Cognitive Therapy Scale (CTS-R). Statistical analyses showed that the therapists' CTS-R scores increased significantly during the phase with supervision with a mean item increase of $M = 0.71$ (range = 0.50–1.0) on the supervision focus areas. This is one of the first empirical studies that can confirm that supervision affect CBT competencies. The results also suggest that supervision can be manualized and that supervisees have a positive perception of more active training methods. Further studies are needed to replicate the results and to find ways to improve the impact of supervision.

Background

Clinical supervision is an integral part of therapist training and continued professional development that aims at improving therapists' competencies for the benefits of their patients (B. E. Wampold & Holloway, 1997). The format and content of clinical supervision in psychotherapy vary between settings but typically takes the form of individual tuition with a focus on specific therapeutic competencies or patient cases (Bernard & Goodyear, 2014; Scaife & Inskipp, 2001). Clinical supervision is commonly based on an assumed causal chain between supervision, therapist practice and patient well-being but surprisingly little empirical

research has been conducted regarding the effects of clinical supervision in psychotherapy and the evidence base for any effects on psychotherapy competencies or patient outcomes is limited (Branson et al., 2015; Freitas, 2002; D. Milne et al., 2008). This stands in sharp contrast to the vast empirical research literature regarding the effects of psychotherapy, especially Cognitive Behavior Therapy (CBT) (Cuijpers et al., 2016; Hofmann et al., 2012).

Several research overviews have concluded that methodological weaknesses are common among studies of clinical supervision and these weaknesses include underdeveloped theoretical frameworks, not using objective and validated measurements and poor study designs or procedures (D. Milne et al., 2008; Olds & Hawkins, 2014). To support evidence-based practices in clinical supervision, research in this area should arguably follow scientific standards more rigorously using adequate empirical designs and methods. This may include standardizing the intervention (i.e. supervision) as well as the outcome measures (i.e. supervisee- or patient outcomes). Regarding outcome measures, competence in CBT can be measured with standardized coding instruments such as the Cognitive Therapy Adherence and Competence Scale (CTACS) (Barber et al., 2003), the Cognitive Therapy Scale—Revised (CTS-R) (Blackburn et al., 2001) and the Assessment of Core CBT Skills (ACCS) (Muse et al., 2017). None of these instruments claim to encompass all therapeutic competencies but they are all designed to capture key features of CBT therapist competence and assess therapeutic skills and CBT competence measured with such instruments may be associated with patient outcomes, though results are currently mixed (Kazantzis et al., 2018; Liness et al., 2019).

Models of CBT competencies have been developed to guide therapist training and to assess therapist competence in clinical practice. They are for example, used by the Improving Access to Psychological Therapies initiative in the UK to accredit therapists. While clinical supervision can be provided for various reasons, such as providing emotional support or training in case formulation, supervision aiming at improving therapists' in-session skills may be focused on competencies specified by the CTS-R or the ACCS. While there are many models for clinical supervision in psychotherapy, there is a need for more standardized manuals or protocols that targets specified competencies that can be measured (Corrie & Lane, 2015; Falender & Shafranske, 2014; D. L. Milne & Reiser, 2017; Scott, 2013; C. E. Watkins, 2012). Supervision could be standardized and designed to first map therapists' current competencies and then help therapists improve identified weaknesses in specific areas. For example, a therapist that according to the CTS-R shows sub-par skills for eliciting key cognitions (i.e., receives a low score on this item of the CTS-R) can be identified and the supervision may be focused on training this specific skill.

Besides standardizing the content of clinical supervision and designing it to correspond to models of CBT competence, the learning methods used by clinical supervisors probably have an impact on supervisees' skills development (Bennett-Levy et al., 2009). Building on learning theory as well as the declarative-procedural-reflective model, clinical supervision may be more effective if it focuses on direct observation, modeling, role-play, self-experimental work and self-reflection (Johnston & Milne, 2012; Pretorius, 2006; C. E. Watkins & Scaturo, 2013). Current models for quality assessment tools of clinical supervision, such as the Supervisor Competency Scale (Naeem et al., 2010) or the Supervision: Adherence and Guidance Evaluation scale (D. L. Milne et al., 2011), include and highlight the importance of these types of experiential learning. But while these

principles have long been proposed, few studies have shown the efficacy of using these learning principles in clinical supervision (James et al., 2007).

In a recent systematic review that included only experimental research (Alfonsson et al., 2018), we found only five empirical studies evaluating the effects of clinical supervision on CBT competence. Of these, three were of adequate scientific quality and only the study by Rakovshik et al. (2016) showed an unambiguous effect of clinical supervision on CBT competence, as measured with the Cognitive Therapy Scale. The study was conducted in a clinical context and compared online supervision with control conditions without supervision. The supervision seems to have been semi-structured and focused on improving specific case conceptualizations. The supervisor was described to have used methods for experiential learning such as modeling and role-play but the supervision was not manualized. The study showed a significant improvement in CTS scores for participants in the supervision condition with a mean item increase of 1.23 points (from 2.33 to 3.56). It is therefore clearly possible to conduct empirical studies on clinical supervision but there are still methodological shortcomings that need further development. One such important area is the standardization of the supervision intervention based on models of CBT competence.

To summarize, the value of clinical supervision in psychotherapy is expressed by several educational, governing and practitioners' bodies but there is very limited empirical support for this standpoint. Taken together, there is a need for high quality, empirical studies on the effects on clinical supervision on therapists' competencies to promote evidence based practice throughout psychotherapy training and practice (Reiser & Milne, 2012). To assess the effects of clinical supervision, supervision should be standardized, manualized and evaluated in experimental studies. The primary aim of this pilot study was therefore to assess whether standardized clinical supervision affects in-session CBT competence. More specifically, we wanted to investigate whether weekly protocol-based clinical supervision would affect supervised therapists' CBT competence as measured with the CTS-R. The research hypothesis was that therapists would show an increase in CTS-R scores during a period of clinical supervision compared to a baseline period without clinical supervision. A secondary aim was to assess the feasibility of protocol-based supervision, to collect feedback on the standardized supervision format from therapists and to assess the study procedures before a full-scale study.

Methods

Study design

This study used a single-case experimental design (SCED) with parallel replication in six participants using multiple baselines. The use of the experimental multiple-baseline design makes it possible to draw causal inferences of treatment effects even though generalizability is lower than for experimental group studies (T. R. Kratochwill & Levin, 2015). Randomizing the baseline length for each participant makes it possible to conduct statistical analyses with increased statistical power (B. E. Wampold & Worsham, 1986; Edgington & Onghena, 2007). Statistical power is generally difficult to assess for single-case designs but based on recommendations from the literature, including six participants was judged to be enough to detect moderate treatment effects using this design (Ferron & Onghena, 1996).

Outcome measurements and data collection

The primary outcome variable, CBT competence, was measured with the CTS-R (Blackburn et al., 2001). The CTS-R can be used to assess CBT competence by coding CBT sessions on 12 items; Agenda Setting & Adherence, Feedback, Collaboration, Pacing and Efficient Use of Time, Interpersonal Effectiveness, Eliciting Appropriate Emotional Expression, Eliciting Key Cognitions, Eliciting Behaviors, Guided Discovery, Conceptual Integration, Application of Change Methods, and Homework Setting. In the CTS-R manual, each item is described and provided with examples of how it should be scored in accordance with different therapist behaviors and skills. For example, when scoring the item Guided Discovery the coder should consider whether the style of the therapist is open and inquisitive and whether the therapist efficiently uses Socratic methods to help the patient better understand her problems. Each of these items is scored on a scale from 0 to 6 with an item score of 3 indicating an acceptable level of competence in that area. Consequently, commonly used thresholds for CBT competence is to receive at least 3 points on each item on the CTS-R as well as a total score of 35 or 40 points according to the CTS-R manual (Mark Latham, personal communication, 2016) The CTS-R is widely used and have undergone psychometric evaluation which shows that inter-rater reliability is moderate to high for both total scores and individual items scores (Blackburn et al., 2001). Unfortunately, the test-retest reliability of the CTS-R has not been investigated.

Secondary outcome variables were measured with self-report instruments collected through a secure webpage. The perceived quality of the supervision was measured with the short-form version of the Supervision Working Alliance Inventory (SWAI) (Smith et al., 2002), which is an instrument used to measure supervisees' perception of the clinical supervision and their collaboration with the supervisor. The SWAI has 12 items that are scored on a scale from 0 to 6 and result in a total score of 0 to 72. Overall satisfaction with supervision was measured with the Supervisee Satisfaction Questionnaire-8 (SSQ-8). The SSQ-8 is analog to the commonly used Client Satisfaction Questionnaire (Attkisson & Zwick, 1982) but is directed to therapists in training rather than clients. The SSQ-8 has 8 items that are scored on a scale from 0 to 3 and provides a total score between 0 and 24. Therapists' symptoms of stress and fatigue during the study period were measured with the Copenhagen Burnout Inventory (CBI) (Kristensen et al., 2005). The CBI comprises 19 items scored on a scale from 0 to 4 and provides a total score of 0 to 76.

All participants were also asked to provide qualitative feedback on their participation in the study and on the standardized supervision format. Before the study, all participants were asked to complete a questionnaire regarding background variables. The CBI was used before and after the study while the other self-report instruments were used at the study end.

Procedure and participants

Participating therapists were recruited by advertising for therapists interested in evaluating competence-focused supervision in professional online forums and e-mailing lists for psychotherapists. Inclusion criteria were having had basic training as a CBT therapist, currently providing CBT treatments and being able to record treatment sessions. In this context, basic training corresponds to a one-year course in psychotherapy which does not lead to accreditation but allows for providing treatment under formal supervision of a senior

psychotherapist. All participants in this study had the same level of formal therapist training despite different educational backgrounds. Exclusion criteria were having had advanced training in CBT or currently having competence-focused clinical supervision. A total of eight potential participants underwent initial screening. Of these, one potential participant was excluded for having started advanced CBT training and one potential participant withdrew from the study for unknown reasons before being included.

A total of six therapists were included in the study. Four were women, two were men and the mean age of the sample was 33.1 years (range 30–38 years). The participants had diverse backgrounds and training and had provided CBT treatments for a mean of 2.8 years (range 1–7 years). One participant reported markedly more work experience than the other but during the inclusion interview it became clear that this person not worked extensively with CBT and that the level of experience in providing treatment was similar to that of the other therapists. Three worked in primary psychiatric care, two worked in specialized psychiatric care and one worked in primary care. One participant had additional basic therapeutic training in a therapeutic method besides CBT. All identified their clinical practice as CBT and all had had at least some previous experience of clinical supervision in CBT. See Table 1 for participant details.

After inclusion, participating therapists were randomized to a baseline (phase A) length of 3 to 6 weeks without contact with a supervisor, after which they started weekly individual supervision (phase B). The total study length was 12 weeks for each participant so the supervision phase ranged from five to eight weeks depending on the length of the baseline phase. While the different lengths of the intervention phases may result in different effects on the therapists, this design was chosen to enable statistical analyses, see below. The therapists informed patients in their caseload about the study and recruited two patients who agreed to be recorded but no additional data was collected from these patients. To be included, patients were required to fulfill the criteria of at least one DSM-5 diagnosis except for severe disorders such as schizophrenia, development disorder or personality disorder and being able to attend to regular treatment with CBT. During each week of the study, the therapist provided audio recorded treatment sessions

Table 1. Participant training, CBT experience and clinical context.

Participant	Education	Years of CBT experience	Clinical setting	CTS-R focus areas for supervision
A	Counselor	7	Primary psychiatric care	Agenda Setting and Adherence Collaboration Guided Discovery
B	Psychologist	3	Primary care	Eliciting Emotional Expression Conceptual Integration Homework Setting
C	Psychologist	3	Specialist psychiatric care	Pacing and Efficient Use of Time Eliciting Emotional Expression Eliciting Key Cognitions
D	Psychologist	1	Primary psychiatric care	Agenda Setting and Adherence Eliciting Key Cognitions Guided Discovery
E	Psychologist	2	Specialist psychiatric care	Feedback Eliciting Emotional Expression Guided Discovery
F	Counselor	1	Primary psychiatric care	Collaboration Guided Discovery Application Of Change Methods

CTS-R = Cognitive Therapy Rating Scale Revised.

from these patients to their supervisor who scrutinized the recordings in order to identify important issues for supervision. The recordings did not have to come from the same patient each week since the supervision was not case-focused but competence-focused (see below). In total, the therapists recorded 126 sessions from 15 different patients. All participating therapists as well as the patients who were recorded provided written informed consent. The study was approved by the Regional Ethics Committee Board (no. 2017/1643-31/5).

All session recordings were gathered at study end and each was given a random code number before being randomly sent for third-party coding. The recordings were coded according to the CTS-R by two trained psychotherapists with long experience of using the CTS-R and who were blinded to participating therapists, study hypotheses, study phases and the order of recorded sessions. To avoid inter-rater reliability issues with the coding instrument, all sessions from a participating therapist was coded by the same coder. Since the present study used a within-subject design, this procedure was deemed to produce the most reliable results.

Intervention

The intervention was delivered by one supervisor with advanced training in CBT, more than ten years of clinical work experience and several years of experience with clinical supervision and therapist training. The study intervention consisted of standardized clinical supervision that was conducted in weekly sessions during phase B for each participant, with each session lasting 50–60 minutes. Before the study, a supervision protocol was produced with detailed instructions for each session. The supervision protocol empathized three major themes for supervision; (1) the supervision should be structured and focus on the CTS-R competencies described in the protocol, (2) the supervisor should make use of collaborative exploration, e.g., through Socratic questioning, to support reflective learning and help identify ways to improve supervisees' competencies, and (3) the supervisor should rely on modeling and role-play to facilitate learning through experience-based training rather than through verbal instructions. The protocol had 13 modules, corresponding to the 12 items in the CTS-R and an introduction module. The introduction module provided overall instructions and a supervision framework for supervisors. It described the overall structure of each supervision session, except for the first session, please see below. The introductory module also provided instructions on how to make a baseline assessment of the supervisee's CBT competencies and map potential areas for development. In the 12 main supervision modules, the corresponding item from the CTS-R were briefly described and explained with typical examples and suggestions. For example, module 1 addressed item 1 of the CTS-R, Agenda Setting & Adherence, and included a brief rationale and description of how agenda setting and structure is commonly used in CBT and provided several examples of how to set and manage an agenda. The module also included several examples on how supervisees may struggle with agenda-setting and examples on how supervisors may discuss this topic and suggestions for modeling and roleplaying agenda setting and adhering to the session structure. The instructions emphasized the importance of exercising these behaviors during the supervision session and rehearsing with the supervisee

1. Check-in
2. Agenda setting
3. Follow-up from last session and homework
4. Supervisee report and supervisor feedback on recorded session
5. Problem solving or repetition of previous material
6. Modelling and/or role-play of the chosen focus competences
7. Collaborative homework assignment
8. Summary and supervisee feedback

Figure 1. Supervision session structure.

rather than discussing them. The module ended with suggestions of suitable assignments for the supervisee but underscored that the tasks should be developed collaboratively.

Before the first supervision session at the start of phase B, the supervisor listened to the therapist's baseline session recordings and evaluated their CBT competence profile with strengths and weaknesses based on the CTS-R. In parallel, each therapist was provided with a CTS-R checklist before the first supervision session and asked to make a self-evaluation of their competence profile. During the first supervision session, the therapist and the supervisor discussed their evaluations of the therapist's competence profile and identified and ranked the three competencies that should be the focus for supervision sessions. These three focus areas of the CTS-R remained throughout the supervision period but the therapist and supervisor were free to allocate time to each item as they deemed necessary.

The structure for the subsequent supervision sessions is described in Figure 1. Before each supervision session, the supervisor listened to the latest treatment sessions recordings to be able to provide feedback to the supervisee regarding the chosen focus items. During the supervision sessions, the supervisor used the three selected modules from the supervisor protocol and explored ways for the supervisee to train these competencies with patients. During each supervision session, the focus was on modeling and roleplaying based on the chosen focus competencies for each supervisee. The sessions ended with collaboratively deciding on the supervisees' assignments and how to train the focus competencies in clinical work until the next supervision session.

All supervision sessions were recorded and a random sample of 30 recordings was selected by the first author to assess intervention fidelity. Of these recordings, 26 (87%) included skills training with role-playing and 30 (100%) included modeling. Overall, 25 (83%) of the scrutinized supervision sessions were assessed as adhering very closely to the prescribed structure and content.

Analyses

In single-case research, results are easier to interpret if baseline phases are stable (i.e., showing low degrees of trend and variability). The CTS-R scores from phase A of each participant therefore underwent a visual inspection before analyses to assess adequate stability. The CTS-R scores were then analyzed for each phase for each participant and the difference in phase means (Δ; M_B—M_A) was used as the test statistic. Both the total

score of the CTS-R and the scores of the three chosen focus competencies for each participant was analyzed with randomization tests.

Randomization test is a type of permutation test that does not require parametric data and is robust to non-normal data distributions and autocorrelation and can therefore be suitable for single-case studies (T. R. Kratochwill & Levin, 2010). Similar to Fisher's exact test, randomization tests are based on comparing the ranking of data in two sets (T. R. Kratochwill & Levin, 2010). When used in single-case research, the data points in both phases can be ranked and compared to all possible permutations of data in the sets. This provides a *p*-value that indicates how unlikely the given result is provided there are no systematic differences between phases, please see B. E. Wampold and Furlong (1981) for examples of this procedure. The null hypothesis in this analysis was that the obtained mean difference between phases would not be different and the exact *p*-value could be calculated by comparing the proportion of all possible mean differences that are smaller than the obtained mean difference between phases. The analytic procedures suggested by Dugard et al. (2011) were followed for the randomization test and all analyses were conducted in Microsoft Excel. The study design resulted in 4096 possible permutations which was deemed adequate for statistical testing. To increase precision, Monte Carlo random sampling procedures are often used and in this study, the CTS-R scores for each recorded session, each participant and each phase was calculated and the aggregated difference obtained was compared with the result of a random sample from the data distribution ($n = 1000$).

Effect sizes for phase differences were assessed using Non-overlap of all Pairs (NAP) calculations (Manolov et al., 2016; R. I. Parker et al., 2011). NAP was calculated by comparing each data point in phase A with each data point in phase B and the resulting NAP was the proportion of all comparisons with no overlap between data points. R. I. Parker and Vannest (2009) suggest that NAP values are adjusted for chance level agreement and that resulting NAP <.65 equals a small effect size, NAP = .66—.92 equals a medium effect size and NAP = .93–1.0 equals a strong effect size.

The secondary self-reported measures were summarized descriptively.

Results

Primary outcomes

The CTS-R total scores and focus items scores for each participant's phases are shown in Table 2. The mean increase was m = 3.53 for the total scores and m = 0.71 for the individual focus items. Visual inspection revealed that the total scores and the scores of the three chosen focus competencies for each participant all increased during the supervision phases compared to the baseline phases and that variance was limited, see Supplement A. The randomization tests showed that the aggregated differences between phases were statistically significant for both the total scores ($\Delta = 21.2, p = .005$) and the focus competencies ($\Delta = 4.21, p = .002$) of the CTS-R. In a post-hoc analysis, differences between phase scores in the CTS-R items that were not in focus, i.e. the remaining nine items that had not been the focus of the supervision, were also investigated and found to be non-significant ($p > .05$). NAP-values ranged from .18 to .94 for the CTS-R total scores and from .37 to .94 for the focus items. The combined NAP was .48 for the total scores and .68 for the focus items which correspond to small and medium effects sizes respectively.

Table 2. Observer-rated CTS-R scores for each participant in each phase and phase differences.

Participant	CTS-R total score Mean (SD)				CTS-R focus items Mean (SD)			
	Phase A	Phase B	Change score Δ	NAP	Phase A	Phase B	Change score Δ	NAP
A	23.3 (4.9)	23.9 (6.9)	0.6	.18	1.8 (0.51)	2.3 (0.75)	0.50	.44
B	44.8 (3.3)	49.5 (2.9)	4.7	.81	3.7 (0.51)	4.6 (0.62)	0.85	.89
C	33.5 (3.0)	36.9 (3.9)	3.4	.47	2.3 (0.50)	3.0 (0.78)	0.75	.94
D	44.3 (5.3)	45.1 (4.1)	0.8	.20	3.4 (1.1)	3.9 (0.70)	0.50	.37
E	22.7 (6.0)	32.3 (3.6)	9.6	.94	1.9 (0.6)	2.9 (0.3)	1.0	.91
F	39.5 (4.3)	41.7 (4.8)	2.2	.25	3.1 (0.50)	3.8 (0.73)	0.64	.56

CTS-R = Cognitive Therapy Rating Scale Revised, NAP = Non-overlap of All Pairs.

Secondary outcomes

The mean and range of the SWAI (M = 62.4, range 54–72) and SSQ-8 (M = 19.6, range 18–21), indicated a high degree of satisfaction with the supervision. The participants reported medium levels of work-related fatigue and stress on the CBI both before (M = 53.0) and after (M = 54.2) the study period. In the free text feedback after the study, all six participating therapists reported being satisfied with the structure of the standardized and manualized supervision format. Participants were especially satisfied with the concrete feedback and practical training during supervision sessions. The most common suggestion for future development was to include even more role-play during supervision. The most common negative feedback was increased self-criticism and workload.

Discussion

The main aim of the present study was to investigate whether structured and manualized clinical supervision could affect CBT competence and the results showed that rater-assessed CBT competence was significantly higher in the supervision phase compared to the baseline phase without supervision. The aggregated numerical increase in CTS-R total scores was modest (Δ = 21.2) but consistent across all six participants. The same aggregated pattern could be seen for the three focus items (Δ = 4.21) which also increased for all participants. The difference between the weak effect size seen in the total scores and the moderate effect size seen for the focus areas was expected given that more skills improvement should be seen in the areas targeted by supervision. Interestingly, the scores of the non-focus items did not statistically change between baseline and supervision phases which indicates that the supervision had a specific effect on the focused items that did not generalize to other competencies.

Compared to the study by Rakovshik et al. (2016), the effects of supervision were more modest in the current study. The effect of supervision in the Rakovshik study was a mean item improvement of 1.23 points on the CTS compared to 0.71 on the focus items of the CTS-R in this study. There may be several reasons for this difference, the simplest being that supervision was of higher quality in the study by Rakovshik et al. However, quality of supervision is difficult to assess and is seldom reported in supervision research which suggests another area for further methodological improvement (D. L. Milne & Reiser, 2011). Further, the participants in the current study had a higher mean baseline CBT competence than the participants in the Rakovshik et al study so there may have been less room for improvement. It is also possible that standardized supervision is less effective

than more individualized supervision, though this has to be confirmed in comparative studies. CBT competence in the Rakovshik et al. study was assessed by one of the authors and though measures were taken to blind the rater it is still possible that the ratings could have been biased. In the current study all therapy sessions were rated by third-part raters who were blinded to study conditions. The fidelity check showed that supervision was overall compliant to the supervision protocol. It is therefore somewhat surprising that the intervention developed for the current study with the specific aim to target the competencies modeled in the CTS-R did not have a larger impact on CTS-R scores. More explorative studies may be needed to identify optimal methods for skills training as well as methods for implementing skills in clinical practice. Live supervision may for example, provide a context in which the distance from learning and practice is much shorter than in delayed supervision (F. Weck et al., 2016; Jakob et al., 2013). In the current study, the most commonly chosen item for development items from the CTS-R was Guided Discovery which was chosen by four participants. Though it is not possible to draw any conclusions from this limited sample, it is possible that some competence areas are more suitable for traditional delayed video-based supervision while other areas may be more suitable for live supervision or other forms of skills training. It is also possible that some competencies require more training than others and that supervision may need to be tailored to fit each therapist's needs better.

The secondary aims of this study were to evaluate the standardized supervision protocol, study procedures and overall feasibility of conducting experimental studies in supervision. The participants overall reported moderate levels of work-related stress and the feedback after the study end was generally very positive regarding the supervision. The most common feedback was a wish for even more focus on skills training through modeling and role-play. There was also very positive feedback on the overall structure and focus on specific skills of the supervision. This suggests that experiential learning methods may be perceived positively by supervisees and not be as aversive as some supervisors may fear. There was very little negative feedback and all participants reported they would recommend the intervention to a colleague. The participants only suggested minor improvements in the supervision protocol and the overall study procedure and technological solutions for recording and sessions and reporting data were deemed adequate. Overall, this study shows that experimental studies in supervision are feasible and that full size randomized controlled studies can be planned and designed upon the methods used in this pilot study.

There are still very few empirical studies on the effects of clinical supervision on psychotherapy competence but the present study provides some evidence that supervision can be an effective way to improve therapists' skills (Alfonsson et al., 2018). Results from single-case studies should be interpreted with caution as generalizability is lower than for large randomized controlled studies. Also, it is not possible to explore possible moderating variables in single-case designs so future large scale studies may analyze potential predictor variables, such as therapists' training. However, the results of this study can be seen as a proof-of-concept since the supervision unanimously improved participants' CTS-R scores. The participants in this study had only basic training in psychotherapy but three of them still showed a high level of baseline CBT competence. This could be a result of the recruitment procedure for this study since advertisements may attract therapists who are already interested in developing their skills and actively

seeking to improve their competence. Whether the effects of supervision also hold in other training settings and other populations needs to be investigated. To our knowledge, supervision format, structure and content vary greatly in the clinical context and that supervision as it is commonly set up may have very different effects than the ones seen in this study. For example, the supervision experiential learning methods emphasized in this study, such as roleplaying, seems to be scarcely used case-focused clinical supervision and in this regard the supervision provided in this study may differ from standard practice. Also, therapists were instructed to reflect upon their competencies based on self-evaluation on the CTS-R before supervision and this may have affected their behavior. The impact of such specific features of the current intervention cannot be assessed in this study but may be the target of future research. In this study, supervisors were also able to listen to session recordings and this may seldom be the case in clinical contexts. Finally, it is important to remember that CBT competence was narrowly defined by the content of the CTS-R in this study and that clinical supervision may target other aspects of psychotherapy competence such as ethical and legal issues, that are not covered by the CTS-R.

There is a continued need for large and rigorous empirical studies on the effects of clinical supervision, especially on patient outcomes. Previous efforts have been hampered by small but important design flaws and unexpected confounding effects that make it impossible to draw firm conclusions (e.g., Bambling et al., 2006; F. Weck et al., 2016). There is therefore a need to confirm that CBT competence as modeled in instruments such as the CTS-R is beneficial for patients since studies in this area have shown mixed results (Webb et al., 2010). To provide some context, in the studies by Kazantzis et al. (2018) and by Liness et al. (2019), a 1-point increase on the CTS-R was associated with a 0.58-point improvement on the BDI and a 0.76-point improvement on the PHQ-9 respectively. Patient outcome data was not collected in the present study but the 3.53-point mean increase on the CTS-R could have had a marginal impact on patients' symptomology, though this is of course speculative. It is also largely unknown whether other treatment factors, such as following an evidence-based treatment protocol, are more important than therapist competence. There is a need to compare different supervision formats, such as individual versus group supervision and delayed video-based supervision versus live bug-in-the-ear supervision. There is a need to identify the optimal intensity and length of supervision for example, by comparing short intensive supervision with prolonged but less frequent supervision. Further, there is a need to investigate whether there is a causal link from clinical supervision to therapist competence and patient outcomes. The training structure of psychotherapists relies on this assumption but there are at present very little empirical support for this standpoint. Overall, there is a need for more high-quality empirical research in this field to promote evidence-based training in psychotherapy. In conclusion, the present study provides some evidence that structured manualized supervision can improve CBT competence but the results need to be replicated and extended in future larger studies.

Disclosure statement

The authors report no conflict of interest.

ORCID

Sven Alfonsson http://orcid.org/0000-0003-4570-5891
Tobias Lundgren http://orcid.org/0000-0002-3872-9660
Gerhard Andersson http://orcid.org/0000-0003-4753-6745

References

Alfonsson, S., Parling, T., Spännargård, Å., Andersson, G., & Lundgren, T. (2018). The effects of clinical supervision on supervisees and patients in cognitive behavioral therapy: A systematic review. *Cognitive Behaviour Therapy*, *47*(3), 206–228. https://doi.org/10.1080/16506073.2017.1369559

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

Bambling, M., King, R., Raue, P., Schweitzer, R., & Lambert, W. (2006). Clinical supervision: Its influence on client-rated working alliance and client symptom reduction in the brief treatment of major depression. *Psychotherapy Research*, *16*(3), 317–331. https://doi.org/10.1080/10503300500268524

Barber, J. P., Liese, B. S., & Abrams, M. J. (2003). Development of the cognitive therapy adherence and competence scale. *Psychotherapy Research*, *13*(2), 205–221. https://doi.org/10.1093/ptr/kpg019

Bennett-Levy, J., McManus, F., Westling, B. E., & Fennell, M. (2009). Acquiring and refining CBT skills and competencies: Which training methods are perceived to be most effective? *Behavioural and Cognitive Psychotherapy*, *37*(5), 571–583. https://doi.org/10.1017/S1352465809990270

Bernard, J., & Goodyear, R. (2014). *Fundamentals of clinical supervision*. Pearson Education.

Blackburn, I.-M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F. K. (2001). The revised cognitive therapy scale (CTS-R): Psychometric properties. *Behavioural and Cognitive Psychotherapy*, *29*(4), 431–446. https://doi.org/10.1017/S1352465801004040

Branson, A., Shafran, R., & Myles, P. (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour Research and Therapy*, *68*(5), 19–26. https://doi.org/10.1016/j.brat.2015.03.002

Corrie, S., & Lane, D. A. (2015). *CBT supervision*. Sage.

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry*, *15*(3), 245–258. https://doi.org/10.1002/wps.v15.3

Dugard, P., File, P., & Todman, J. (2011). *Single-case and small-n experimental designs: A practical guide to randomization tests*. Routledge.

Edgington, E.S. & Onghena, P. (2007). *Randomization tests*. (4. ed.) Boca Raton, Fla.: Chapman & Hall/CRC.

Falender, C. A., & Shafranske, E. P. (2014). Clinical supervision: The state of the art. *Journal of Clinical Psychology*, *70*(11), 1030–1041. https://doi.org/10.1002/jclp.2014.70.issue-11

Ferron, J., & Onghena, P. (1996). The power of randomization tests for single-case phase designs. *The Journal of Experimental Education*, *64*(3), 231–239. https://doi.org/10.1080/00220973.1996.9943805

Freitas, G. J. (2002). The impact of psychotherapy supervision on client outcome: A critical examination of 2 decades of research. *Psychotherapy: Theory, Research, Practice, Training*, *39*(4), 354. https://doi.org/10.1037/0033-3204.39.4.354

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

Jakob, M., Weck, F., & Bohus, M. (2013). Live supervision: From the one-way mirror to video-based online supervision. *Verhaltenstherapie*, 23(3), 170–180. https://doi.org/10.1159/000354234

James, I. A., Milne, D., Marie-Blackburn, I., & Armstrong, P. (2007). Conducting successful supervision: Novel elements towards an integrative approach. *Behavioural and Cognitive Psychotherapy*, 35(2), 191–200. https://doi.org/10.1017/S1352465806003407

Johnston, L. H., & Milne, D. L. (2012). How do supervisee's learn during supervision? A grounded theory study of the perceived developmental process. *The Cognitive Behaviour Therapist*, 5(1), 1–23. https://doi.org/10.1017/S1754470X12000013

Kazantzis, N., Clayton, X., Cronin, T. J., Farchione, D., Limburg, K., & Dobson, K. S. (2018). The cognitive therapy scale and cognitive therapy scale-revised as measures of therapist competence in cognitive behavior therapy for depression: Relations with short and long term outcome. *Cognitive Therapy and Research*, 42(4), 385–397. https://doi.org/10.1007/s10608-018-9919-4

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*, 15(2), 124. https://doi.org/10.1037/a0017736

Kratochwill, T. R., & Levin, J. R. (2015). *Single-case research design and analysis (Psychology revivals): New directions for psychology and education*. Routledge.

Kristensen, T. S., Borritz, M., Villadsen, E., & Christensen, K. B. (2005). The copenhagen burnout inventory: A new tool for the assessment of burnout. *Work & Stress*, 19(3), 192–207. https://doi.org/10.1080/02678370500297720

Liness, S., Beale, S., Lea, S., Byrne, S., Hirsch, C. R., & Clark, D. M. (2019). The sustained effects of cbt training on therapist competence and patient outcomes. *Cognitive Therapy and Research*, 43(3), 631–641. https://doi.org/10.1007/s10608-018-9987-5

Manolov, R., Losada, J. L., Chacón-Moscoso, S., & Sanduvete-Chaves, S. (2016). Analyzing two-phase single-case data with non-overlap and mean difference indices: Illustration, software tools, and alternatives. *Frontiers in Psychology*, 7, 32. https://doi.org/10.3389/fpsyg.2016.00032

Milne, D., Aylott, H., Fitzpatrick, H., & Ellis, M. V. (2008). How does clinical supervision work? Using a "best evidence synthesis" approach to construct a basic model of supervision. *The Clinical Supervisor*, 27(2), 170–190. https://doi.org/10.1080/07325220802487915

Milne, D. L., & Reiser, R. P. (2011). Observing competence in CBT supervision: A systematic review of the available instruments. *The Cognitive Behaviour Therapist*, 4(3), 89–100. https://doi.org/10.1017/S1754470X11000067

Milne, D. L., & Reiser, R. P. (2017). *A manual for evidence-based CBT supervision*. John Wiley & Sons.

Milne, D. L., Reiser, R. P., Cliffe, T., & Raine, R. (2011). SAGE: Preliminary evaluation of an instrument for observing competence in CBT supervision. *The Cognitive Behaviour Therapist*, 4(4), 123–138. https://doi.org/10.1017/S1754470X11000079

Muse, K., McManus, F., Rakovshik, S., & Thwaites, R. (2017). Development and psychometric evaluation of the Assessment of Core CBT Skills (ACCS): An observation-based tool for assessing cognitive behavioral therapy competence. *Psychological Assessment*, 29(5), 542. https://doi.org/10.1037/pas0000372

Naeem, F., Phiri, P., Rathod, S., Kingdon, D., Mueller, M., Kennerley, H., McManus F., & Westbrook, D. (2010). *The Oxford guide to surviving as a CBT therapist*. Oxford University Press.

Olds, K., & Hawkins, R. (2014). Precursors to measuring outcomes in clinical supervision: A thematic analysis. *Training and Education in Professional Psychology*, 8(3), 158. https://doi.org/10.1037/tep0000034

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, 40(4), 357–367. https://doi.org/10.1016/j.beth.2008.10.006

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, 35(4), 303–322. https://doi.org/10.1177/0145445511399147

Pretorius, W. M. (2006). Cognitive behavioural therapy supervision: Recommended practice. *Behavioural and Cognitive Psychotherapy*, *34*(4), 413–420. https://doi.org/10.1017/S1352465806002876

Rakovshik, S. G., McManus, F., Vazquez-Montes, M., Muse, K., & Ougrin, D. (2016). Is supervision necessary? Examining the effects of internet-based CBT training with and without supervision. *Journal of Consulting and Clinical Psychology*, *84*(3), 191. https://doi.org/10.1037/ccp0000079

Reiser, R. P., & Milne, D. (2012). Supervising cognitive-behavioral psychotherapy: Pressing needs, impressing possibilities. *Journal of Contemporary Psychotherapy*, *42*(3), 161–171. https://doi.org/10.1007/s10879-011-9200-6

Scaife, J., & Inskipp, F. (2001). *Supervision in the mental health professions: A practitioner's guide*. Psychology Press.

Scott, M. J. (2013). *Simply effective CBT supervision*. Routledge.

Smith, T. R., Younes, L. K., & Lichtenberg, J. W. (2002, August 22–25). *Examining the working alliance in supervisory relationships: The development of the working alliance inventory of supervisory relationships* [Paper presentation]. Annual Conference of the American Psychological Association, Chicago, IL, USA.

Wampold, B. E., & Furlong, M. J. (1981). Randomization tests in single-subject designs: Illustrative examples. *Journal of Behavioral Assessment*, *3*(4), 329–341. https://doi.org/10.1007/BF01350836

Wampold, B. E., & Holloway, E. L. (1997). Methodology, design, and evaluation in psychotherapy supervision research. In C. E. Watkins, Jr. (Ed.), *Handbook of psychotherapy supervision* (pp. 11–27). John Wiley & Sons, Inc.

Wampold, B. E., & Worsham, N. L. (1986). Randomization tests for multiple-baseline designs. *Behavioral Assessment*, *8*(2), 135–143.

Watkins, C. E. (2012). Psychotherapy supervision in the new millennium: Competency-based, evidence-based, particularized, and energized. *Journal of Contemporary Psychotherapy*, *42*(3), 193–203. https://doi.org/10.1007/s10879-011-9202-4

Watkins, C. E., Jr, & Scaturo, D. J. (2013). Toward an integrative, learning-based model of psychotherapy supervision: Supervisory alliance, educational interventions, and supervisee learning/relearning. *Journal of Psychotherapy Integration*, *23*(1), 75. https://doi.org/10.1037/a0031330

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *78*(2), 200. https://doi.org/10.1037/a0018912

Weck, F., Jakob, M., Neng, J. M., Höfling, V., Grikscheit, F., & Bohus, M. (2016). The effects of bug-in-the-eye supervision on therapeutic alliance and therapist competence in cognitive-behavioural therapy: A randomized controlled trial. *Clinical Psychology & Psychotherapy*, *23*(5), 386–396. https://doi.org/10.1002/cpp.v23.5