Cohort outcome study of cognitive analytic therapy in a private practice

IRENE ELIA

Abstract:

Background: The evidence base for effectiveness of cognitive analytic therapy (CAT) in private practice is limited. **Aims:** To assess effectiveness of the author's private practice in reducing global distress, measured by Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM), and to hypothesis-test if CORE-OM change (delta) from beginning to end of therapy significantly correlates with client-therapist goodbye letter concordance, a newly devised indicator. **Methods:** CORE-OM deltas were correlated with pre-CAT scores and with client-therapist Goodbye letter concordance for presence/absence of reciprocal roles (RR), target problem (TP), and traps, dilemmas, and snags (T,D,S). Deltas were compared to those identified in a systematic review by the author.

Results: Of 103 clients (36.26± 9.43 years, 79% women, receiving 15.75± 2.56 CAT sessions for anxiety, depression, and relationship problems), 53 had complete datasets. The mean pre-CAT CORE-OM score of 1.21± 0.68 dropped by 0.50± 0.54 at the end of CAT (P<0.001); (by 0.60± 0.53 without trainees, P<0.001). 51% of the whole cohort (62.5% without trainees) showed reliable improvement (≥ 0.50 points). CORE-OM deltas, which were typical of nine CAT studies from the systematic review, were correlated with pre-CAT scores (r= 0.726, P<0.001). Goodbye letter concordances (RR=81%, TP=81%, T,D,S=58%) were not significantly correlated to CORE-OM deltas. **Conclusion:** Following CAT, the cohort showed a highly significant and predictable improvement in CORE-OM measured distress, with more than half showing reliable improvement. High Goodbye letter concordance suggested shared understanding that may have

Irene Elia PhD, MSc, UKCP
Independent researcher
07866052416
me14@cam.ac.uk.
68 Highsett, Cambridge CB2 1NZ

enabled positive change but was not significantly correlated with CORE-OM deltas, perhaps because concordance indicates shared understanding not captured by CORE-OM, and CORE-OM measures distress not captured by concordance.

Introduction

Cognitive Analytic Therapy (CAT) is a time-limited therapy (usually 16 sessions, once/week) that helps clients see unhelpful behaviours, thoughts, and feelings (Ryle and Kerr, 2020). As this requires the client to observe and reflect, CAT is not used with clients actively abusing substances or in florid psychosis. However, the evidence base for effectiveness of CAT is not strong and could be improved (Calvert and Kellett, 2014, Martin et al., 2021, Hallam et al., 2021, Baker, 2003).

The effectiveness of CAT can be measured in many ways, both by independent clinical evaluation and self-reported outcome. Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM) is a validated, reproducible and widely used tool that assesses distress (Evans et al., 2002), used here to retrospectively explore CAT effectiveness over a period of 17 years in the author's private practice.

Another potential way to assess effectiveness of CAT concerns the Goodbye letters, which are usually written by clients at the end of CAT to reflect their understanding of and feelings about therapy and what remains to be worked on. The therapist also writes a Goodbye letter summarising what has been understood and some key experiences that have helped the client recognise and start to revise any problematic patterns. It has been the author's impression that concordance between client and therapist content shows that a client 'gets' what is going on with them and feels less distress because they know what they're doing and what they may need to change. On this basis, it was hypothesized that concordance between client and therapist Goodbye letters in mentioning CAT elements (target problem (TP), reciprocal role(s) (RRs), traps, dilemmas, or snags (T,D,S)), which frame self-understanding, could potentially show therapy has been effective.

While some large multi-centre studies and reviews have used CORE-OM to assess the effectiveness of an amalgamation of different psychological treatments (Gilbert et al., 2005, Stiles et al., 2015, Barkham et al., 2005), or a combination of self-report tools including CORE-OM to assess the effectiveness of CAT (Calvert and Kellett, 2014, Hallam et al., 2021), none focussed on the specific use of CORE-OM to evaluate only CAT. Also, the author is unaware of any study establishing correlational validity between CORE-OM deltas and any form of evaluation of Goodbye letters.

The aims here are to assess the effectiveness of the author's CAT practice using change in CORE-OM as an objective indicator of improvement or deterioration in global distress, and to examine the novel hypothesis that concordance of CAT elements in client and therapist Goodbye letters relates to a drop in CORE-OM score. In addition, in order to put the results in perspective a systematic review is to be udertaken to compare results of this study with those reported in peer-reviewed studies using a change in CORE-OM score (usually along with other tools) to assess CAT effectiveness.

Methods

Overview
This study was undertaken retrospectively, during the covid pandemic 2020-21, using an available sample of clients seen between 2001 and 2018 in Cambridge. If clients were on anti-depressants or other medications from their GP, they continued until they consulted their GP about stopping. No one had another form of psychotherapy while receiving CAT. Clients were given a CORE-OM to complete in session 1 (pre-CAT) and another in the penultimate or last session (post-CAT). The pre-CAT score may be referred to as the initial (or first) score and the post-CAT score as the final (or second) score. Clients wrote a Goodbye letter; the therapist wrote a Goodbye letter to each of them.

Retrospective client enrolment and cohort characteristics
Out of 103 clients who started CAT, 50 were excluded for reasons shown in Figure 1. Either they dropped out due to lack of funds/insurance (N=24) or were excluded because they hadn't completed one or both CORE-OM forms (N=17). See Table 1 for baseline characteristics of the sample of 53. Of these, 41 had a 16-session CAT, six had eight sessions, one had 10 sessions, one had 15 sessions, two had 17, and two had 24. Twenty-three of the cohort were involved in psychology/psychiatry, and of these 13 were training to be clinical psychologists/therapists. CAT
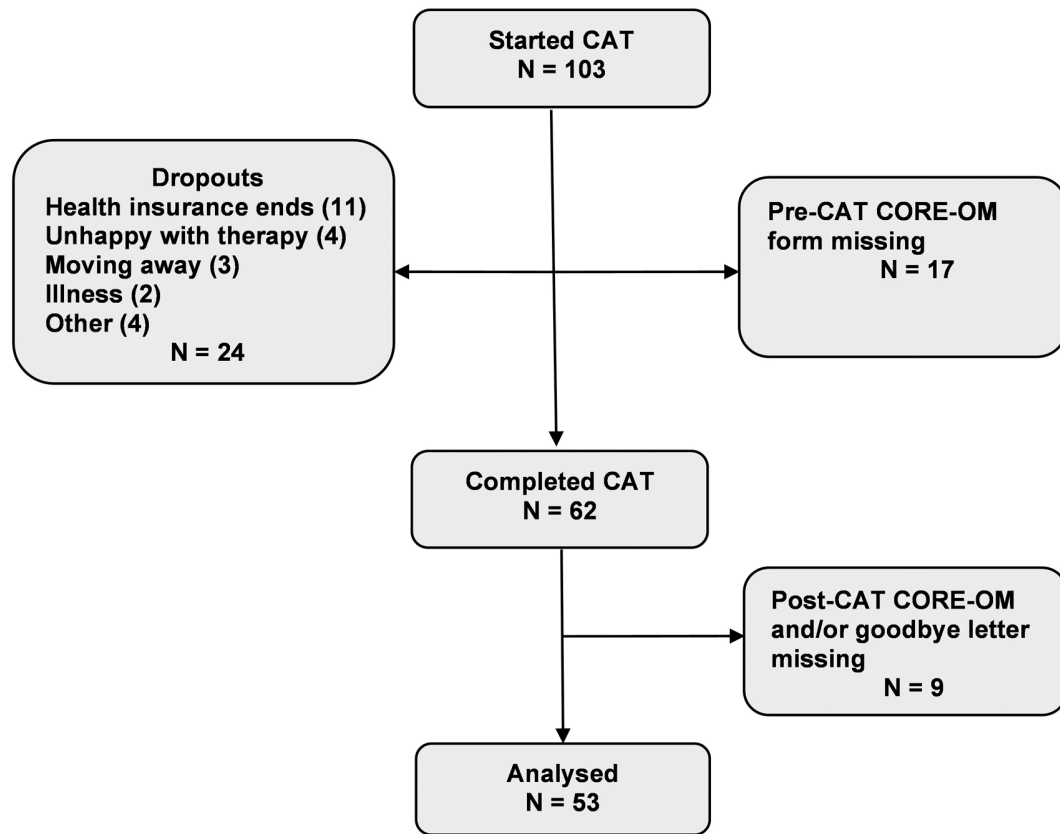
Figure 1 Flow chart of study from starting CAT to data analysis

therapy was an academic requirement for them; most would not otherwise have come for therapy.

Scoring CORE-OM and Goodbye letter concordance
The total score from CORE-OM forms was divided by the number of questions (34) to give a mean score. The change in mean score (initial minus final score; delta CORE-OM, or CORE-OM delta) was calculated, with a positive score representing improvement and a negative score deterioration. After reading the Goodbye letters and noting any mention of TP, RRs, and T, D, S. the results were analysed using two methods. In Method 1, whenever the client or therapist mentioned an element, it was labelled as Yes and Yes/Yes if they both mentioned the element. This is referred to as concordance or agreement. When a CAT element was not mentioned in either Goodbye letter (No/No), whether because they were not relevant to or only briefly/incompletely discussed in therapy, it was still considered 'concordance'. However, a relevant element's absence from only one of the letters (therapist/client: No/Yes or Yes/No) represented a lack of concordance. A more stringent approach is taken in Method 2, which holds that absence of an element in a Goodbye letter could have occurred not only due to not being included in the therapy, but also from intentional/unintentional omission from the letter or from lack of understanding by the client. So, any letter labelled No either for the therapist or the client (No/No, No/Yes, Yes/No) embodies uncertainty. With Method 2, therefore, the only certain concordance would be Yes/Yes. The remaining cases could all theoretically be concordant or non-concordant; we just don't know. Therefore, Method 2, using sensitivity/uncertainty analysis, expresses concordance as a range, from a minimum of certain concordance (Yes/Yes) to a maximum potential concordance.

Relationship of CORE-OM delta to Goodbye letter concordance
The relationship of CORE-OM delta to goodbye letter concordance was assessed by regressing CORE-OM on concordance for individual CAT elements (point-biserial correlation: 1=concordance, 0=no concordance) as well as on the sum of individual CAT element concordances.

Comparison with other studies identified through systematic literature search
A systematic review, undertaken on 28 August 2021, followed standard

Table 1 Baseline characteristics of clients and number of CAT sessions

	All subjects N=53	Without trainees N=40	Trainees N=13	P-Value Trainees vs rest
Mean age (range in years)	36.26±9.43 (21-73)	36.92±10.19 (21-73)	34.23±6.48 (27-45)	0.376[a]
Gender F/M (%/%)	42/11 (79.2/20.8)	30/10 (75.0/25.0)	12/1 (92.3/7.7)	0.181[b]
Mean number of CAT sessions	15.75±2.56	15.78±2.70	15.69±2.18	0.921[a]

Results for age and number of sessions are presented as mean ± standard deviation (SD)
[a] unpaired t-test
[b] chi square test

procedures (Egger et al., 2001, Gough et al., 2017). The key words used in MEDLINE and PsychINFO were 'cognitive' and 'analytic' and 'therapy'. Cross referencing and secondary searches were also undertaken. The inclusion criteria were as follows: adults (≥ 18 years) receiving CAT in any setting; single and multiple therapists; public and private sector; pre-post CAT designs, including randomised controlled trials with pre-post CAT in one of the trial arms; and primary outcome being CORE-OM delta (pre- minus post CAT CORE-OM measured at the end of therapy).

Statistics
Statistical analyses were carried out on the entire cohort (N=53) and on groups: 40 non-trainees, 13 trainees, 30 with non-psychology-psychiatry jobs, 23 with psychology-psychiatry jobs. In addition to the descritpive statistics shown in Table 1, CORE-OM deltas were analyzed by regression analysis and paired or unpaired t-tests. The Statistical Package for the Social Sciences (SPSS version 27) was used. P-values of <0.05 were considered significant. Results are presented as mean± SD.

Cut-off points for CORE-OM scores between this cohort ('clinical') and the general population ('non-clinical') datasets, and between reliable and non-reliable changes (either for improvement or deterioration) were established using standard methodology (Jacobson and Truax, 1991). To calculate these cut-off points, use was made of the mean± SD of the cohort's CORE-OM scores and those of the general population of Great Britain (4.8± 4.3) (Connell et al., 2007). To calculate reliable changes (not measurement error or chance), use was made of the standard deviation of the cohort's CORE-OM scores and of an internal CORE-OM consistency score (i.e., the CORE-OM reliability or reproducability score, here the Crombach's alpha) of 0.93, chosen so as to fall between 0.91 used by Connel et al. 2007 and 0.95 for clinical samples reported by others (Evans et al., 2002, Barkham et al., 2005, Trujillo et al., 2016).

Ethics
Since this study is a service evaluation, approval from the Local Ethics Committee was not sought. However, the data were handled anonymously and with confidentiality throughout the study.

Results
The study group (N=53) did not differ significantly from the people not included in the study (N=50): for age (36.26± 9.43 vs 39.90± 0.56 respectively; P=0.745), sex (F/M: 42/11; vs 37/13; P=0.529), or pre-CAT CORE-OM scores (1.21± 0.69 vs 1.24± 0.71; P=0.38). During the study, no clients attempted suicide, revealed suicidal thoughts, or were admitted to hospital for mental health issues.

Baseline characteristics of cohort and subcohorts
The baseline characteristics of the entire cohort and of subgroups of 'trainees' and 'non-trainess' are shown in Table 1. There was no significant difference in age, sex distribution, or number of sessions received between trainees and non-trainees. While non-trainees received CAT because of depression, anxiety, and/or relationship problems, the trainees needed CAT for their qualification.

CORE-OM scores, CORE-OM deltas, and clinical cut-offs
CORE-OM scores: The pre-CAT CORE-OM scores ranged from 0.06 to 2.97, with 19% having scores <0.5 (very little/no distress) and 15% having scores >2.0 (moderate to severe distress) (https://therapymeets numbers.com/made-to-measure-core/). The majority (49%) had low to moderate distress (scores of 0.6–1.5). Table 2 shows the mean pre-CAT score for the whole cohort was 1.21, and that the mean pre-CAT score for non-trainees was significantly higher (P=0.001) at 1.37.

Table 2 CORE-OM scores before and after CAT by trainee status

	Whole cohort N=53	Non-trainees N=40	trainees N=13	P-Value trainees vs non-trainees[a]
Pre-CAT CORE-OM score	1.21±0.68	1.37±0.67	0.70±0.67	0.001
Post-CAT CORE-OM score	0.71±0.47	0.77±0.49	0.53±0.37	0.114
Delta CORE-OM	0.50±0.54	0.60±0.53	0.17±0.44	0.010
P-value for delta CORE-OM[b]	<0.001	<0.001	0.19	

Results are presented as mean ± standard deviation (SD)
[a] unpaired t-test
[b] paired t-test

CORE-OM deltas

Table 2 shows a highly significant (P<0.001) reduction in the whole cohort's mean final CORE-OM score, which was on average 0.50 points (41%) less than its mean initial score. Without the trainees, the reduction was even greater: 0.60 points (44%). This was related to an'overall improvement in 91% of the entire cohort and in 93% if trainees were excluded. Trainees had smaller deltas (P=0.01).

Table 3 below shows that the mean of the initial CORE-OM scores for the psychology/psychiatry group (N=23) was significantly lower (P<0.001) than that of the rest of the cohort (N=30). Their mean delta CORE-OM score was also less than for the rest of the cohort (P<0.02) (Table 3).

	Whole cohort N=53	Non-psych[a] N=30	Psych[a] N=23	P-Value Non-psych vs Psych[b]
CORE-OM score pre-CAT	1.21±0.68	1.50±0.62	0.83±0.55	<0.001
CORE-OM score post-CAT	0.71±0.47	0.84±0.49	0.54±0.38	0.018
Delta CORE-OM	0.50±0.54	0.65±0.49	0.30±0.53	0.014
P-value for delta CORE-OM[c]	<0.001	<0.001	0.015	

Results are presented as mean ± standard deviation (SD).
[a] Non-psych=not in a psychology/psychiatry profession; Psych=in psychology/psychiatry profession
[b] unpaired t-test
[c] paired t-test

Table 3 CORE-OM before and after CAT by profession

When the trainees were excluded, so that the number of subjects in the Psych group was reduced to only 10, the P-values for comparisons with the Non-psych group became less strong (e.g., post-CAT scores (0.84± 0.49 vs 0.55± 0.42; P=0.099) or deltas (0.66± 0.50 vs 0.46± 0.62; P=0.311), but remained significant for the pre-CAT CORE-OM scores (1.01± 0.69 vs 1.50± 0.62; P=0.042).

Clinical cut-offs and reliable CORE-OM deltas

Using the procedure of Jacobson and Truax (Jacobson and Truax, 1991) on the dataset, it was shown that a reliable delta CORE-OM involved either a reduction by ≥ 0.5 (reliable improvement) or an increase by ≥ 0.5 (reliable deterioration), in agreement with Connell et al (Connell et al., 2007) and others (Evans et al., 2017b, Kellett et al., 2020) and https://therapymeetsnumbers.com/made-to-measure-core/).

The 'clinical' distress cut-off points between the cohort/subgroup (53 clients/40 without trainees) and the general population (535 people) (Connell et al., 2007) were found to be 0.763 and 0.837, respectively. Both cut-offs were lower than the 0.99 found by Connell et al (Connell et al., 2007)'because mean pre-CAT CORE-OM scores here were lower than Connell et al's. Following CAT, fewer clients had CORE-OM scores above the clinical distress cut-off (dots above horizontal line on Figure 2) than they had before CAT (dots to right of the vertical line Figure 2).
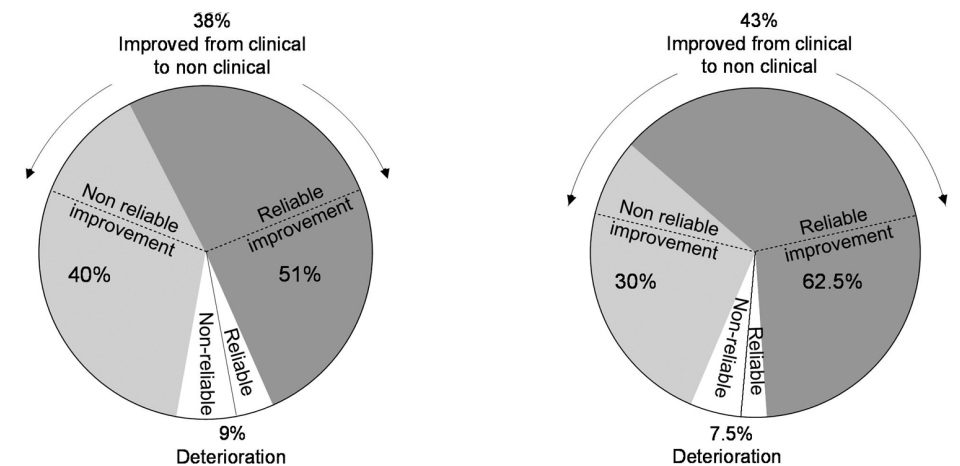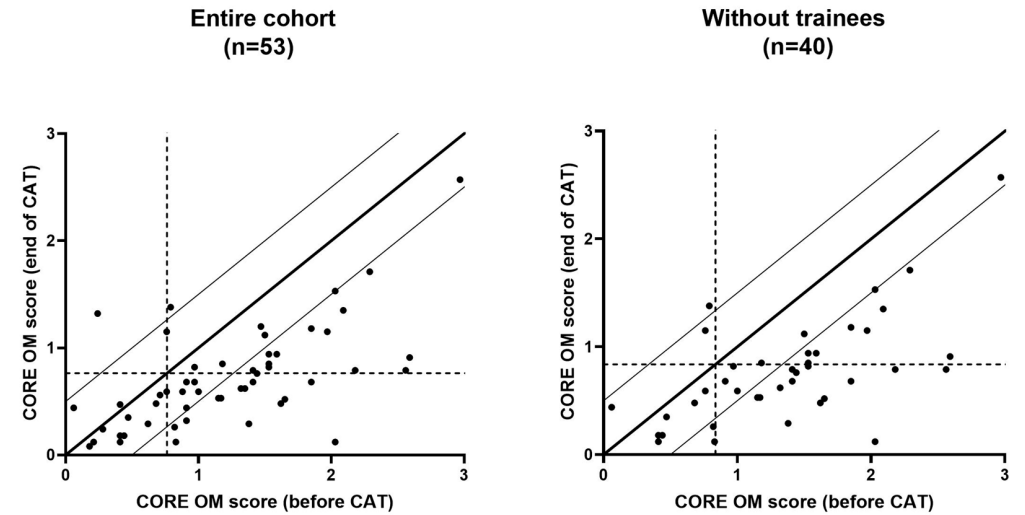

Figure 2 Effect of CAT on CORE-OM deltas (*left*, whole cohort; *right*, cohort without trainees) Thick central tramline on graphs represents no change; lower and upper tramlines indicate the cut-off points for reliabe improvement and reliable deterioration, respectively. Pie charts show overall improvement: reliable (dark gray) plus non-reliable (light gray). Dotted lines indicate the cut-off point between clinical and non-clinical status.

Most clients had a pre-CAT CORE-OM score above the clinical cut-off ('distressed'), but 28% had an initial CORE-OM score below it ('not distressed'), symbolised by the solid dots to left of the vertical lines on the graphs in Figure 2. Among those that had pre-CAT CORE-OM scores above the clinical cut-off, there was 46% improvement following CAT (47% in the cohort without trainees), and in the quarter with the greatest drop in CORE-OM score, the range of improvement was 64-93%.

More than half of the cohort showed reliably lower CORE-OM scores following CAT (reduction in distress): 51% of the entire cohort and 62.5% of the non-trainees (dots below the lowest tramline on the graphs of Figure 2). Clients (dots) that are below this tramline as well as below the horizontal dotted line on graphs in Figure 2 reliably improved from 'clinical' to 'non-clinical' distress status, corresponding to 26% of the cohort and 35% of the cohort without trainees. Reliable deterioration (dots above the highest tramline) occurred in only two clients (3.8% of whole cohort) and just one client (2.5% of cohort minus trainees). Overall, 30% of the cohort without trainees showed non-reliable change, either toward improvement or toward deterioration, as represented by the dots between the lower and upper tramlines in Figure 2; 45% with trainees.

Relationship of CORE-OM delta to pre-CAT CORE-OM score
Close inspection of the graphs in Figure 2 shows that after CAT, clients with higher initial scores showed greater improvement (greater downward displacement of dot from a 'no change' position on the central tramline). Indeed, there was a highly significant correlation (P<0.001) between the reduction in CORE-OM score (pre-CAT minus post-CAT score) and the pre-CAT score for both the entire cohort (r=0.726)) and the cohort without the trainees (r=0.692) (Figure 3).'Those with high initial scores tended to have greater absolute and greater proportional reductions in CORE-OM scores following CAT. For the entire cohort, the regression equation (post-CAT score=-0.197 + 0.575 pre-CAT score) predicts that in a client with a pre-CAT score of 0.5 the reduction in CORE-OM following CAT is 0.09 (18% reduction in distress); with a pre-CAT score of 1 it is 0.38 (38% reduction in distress); with pre-CAT of 2 it is 0.95 (48%); and with 3 it is 1.53 (51%). When the pre-CAT score is 1.21 (the mean pre-CAT value of this population), the predicted reduction of 0.50 (41%) corresponds exactly to the observed mean reduction. A similar pattern is predicted by the regression equation for the cohort without trainees (post-CAT score=-0.149 + 0.549 pre-CAT score).
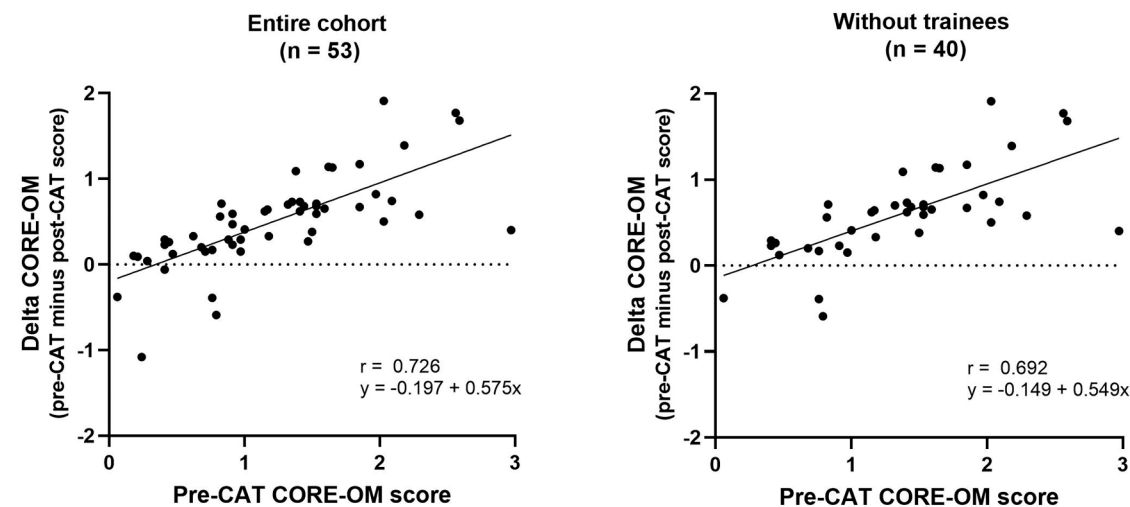

Figure 3 Relationship between delta CORE-OM and the pre-CAT score: entire cohort (left side) vs the entire cohort without the trainees (right). Solid line indicates predicted CORE-OM scores.

Figure 4 shows that the percent improvement in CORE-OM following CAT increases curvilinearly with increasing pre-CAT CORE-OM score, with
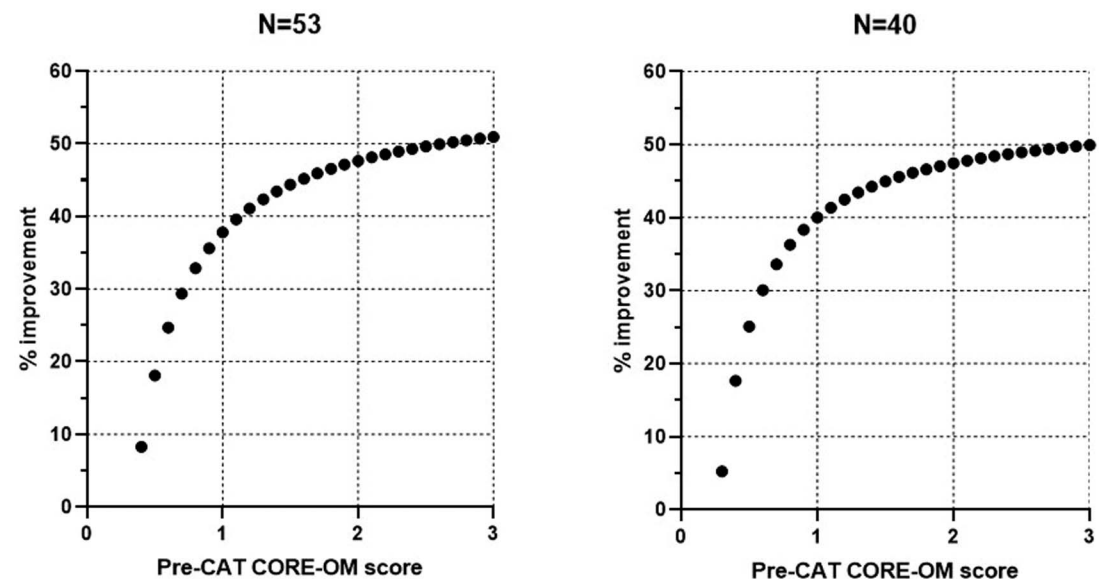

Figure 4 Relationship between Pre-CAT CORE-OM score and percent improvement in CORE-OM following CAT in the group of 53 clients (left; entire cohort; y = 77.5 - (19.7/x)) and 40 clients (right, without trainees y = 54.9 - (14.90/x)). (Percent improvement in CORE-OM = 100 times predicted CORE-OM delta – solid line in Figure 3 – divided by pre-CAT CORE-OM.)

greater percent improvement in those with higher initial pre-CAT scores. For example, those with pre-CAT CORE-OM scores from 1 to 3, (representing 65% of non-trainee clients) might expect a 40-50% improvement; those with scores from 0.5 to 1, (as in 20% of all clients) might see 25-49% improvement; and with scores from 0 to 0.5 could expect <25% improvement or even no improvement. So, improvement is seen to be predictable.

Goodbye letters: CAT elements, client-therapist concordance, relationship to delta CORE-OM
Mention of CAT elements in Goodbye letters: Table 4 shows the frequency with which CAT elements are mentioned in the Goodbye letters, from which it can calculated that the proportion of trainees (N=13) vs non-trainees (N=40) doing so was as follows for each element: RR 85% vs 80%; TP 92% vs 85%; T,D,S 54% vs 48%. (In all cases the P values were between 0.5-1.0 by Chi-squared test.)

Goodbye letter concordance: Using Method 1, the upper half of Table 4 shows that for the whole cohort, the extent of concordance between therapist and client in the goodbye letters (top left + bottom right in each set of four in Table 4) was very good for RR and TP (both scoring 43/53, 81.1%) but less good for T, D, S (31/53, 58.5%), a pattern reflected in the cohort without trainees (lower Table 4). With Method 2 (sensitivity/ uncertainty analysis) the results are reported as a range from a minimum certain concordance (Yes/Yes; lower right number in each set of four on Table 4) to a maximum of 100% (always 53/53 for the entire cohort or 40/40 for cohort minus the trainees). For the entire cohort, the concordance between Therapist and Client on CAT elements ranges from 81-100% for RR; 74-100% for TP, and 42-100% for T, D, S. For the cohort without trainees, the concordance ranges were 80-100%, 73-100% and 38-100%, respectively.

Relationship of CORE-OM delta to Goodbye letter concordance: CORE-OM delta was found to be weakly but non-significantly related to the amount of concordance between therapist and client for individual CAT elements (r=0.05 for RR, r= 0.08 for TP and r= 0.11 for T, D, S) for the whole cohort; (r= 0.09 for RR, r= <-0.01 for TP, and r=< -0.01 for T, D, S) for the cohort without trainees, and to all three elements combined: r =0.12 for the whole cohort and r= 0.17 for the cohort without trainees. After adjustment for pre-CAT CORE-OM scores, all correlations between CORE-OM deltas and concordances remained non-significant, both for

			Client		Concordance			
					Method 1		Method 2	
			No	Yes	N	%	N	%
Entire cohort (N=53)								
RR	Therapist	No	**0**	0	43	81	43-53	81-100
		Yes	10	**43**				
TP	Therapist	No	**4**	7	43	81	39-53	74-100
		Yes	3	**39**				
T, D, S	Therapist	No	**9**	4	31	58	22-53	42-100
		Yes	18	**22**				
Without trainees (N=40)								
RR	Therapist	No	**0**	0	32	80	32-40	80-100
		Yes	8	**32**				
TP	Therapist	No	**3**	5	32	80	29-40	73-100
		Yes	3	**29**				
T, D, S	Therapist	No	**7**	4	22	55	15-40	38-100
		Yes	14	**15**				

Table 4 Concordance of therapist and client letters for CAT elements (entire cohort and cohort without trainees) by Methods 1 and 2.

the results of Method 1 and the lowest % concordances for Method 2. Analysis for the huge number of other permutations - Yes/No, No/No, No/Yes - between the lowest and highest percentage points was not done.

Comparison with other studies identified through systematic literature search
Out of a total of 3,529 publications retrieved from Medline and PsychINFO, cross-referencing, and secondary literature searches only nine were eligible for inclusion (Baronian and Leggett, 2020, Birtchnell et al., 2004, Clarke et al., 2013, Darongkamas et al., 2017, Evans et al., 2017b, Kellett et al., 2013, Kellett et al., 2020, Martin et al., 2021, Williams and Craven-Staines, 2017). Although all examined the effect of CAT on CORE-OM, they were heterogenous. They differed in multiple ways: study design (pre-post CAT cohort studies or randomised controlled studies); sample size (7-53); enrolment (prospective, retrospective); underlying client conditions (chronic pain, personality disorders, bipolar disorder, anxiety/ depression or unspecified); age (mean age of studies 36-73); sex (50-82% women); type of practice (private, public); CAT format (one study involved group CAT). They also differed in the type of analysis: 'complete

case analysis' (all data exists, as in this study) vs 'intention to treat analysis' (some missing data imputed) and in the method of calculating cut-off points and reliable CORE-OM deltas. The present study tended to have pre-CAT CORE-OM scores (mean 1.21 in the entire cohort; 1.37 in the cohort without trainees) towards the lower end of the range reported by the other studies (1.30-2.18), but it produced typical CORE-OM deltas 0.50 for all vs 0.61 without trainees (cf. 0.36-0.83); typical levels of reliable improvement 50.9% vs 62.5% (cf. 41.2-71.4%); and typical levels of reliable clinical improvement 26.4% vs 35% (cf. 17.6-42.9%).

Discussion

This study has shown that more than 90% of the cohort had reduced CORE-OM measured distress following CAT, with more than half showing reliable improvement. The change was significantly correlated with pre-CAT scores, but not with Goodbye letter concordance as assessed here. The data were analysed with trainees (N=53) and without the trainees (N=40) because the trainees would not have come to therapy unless they needed to do so for their qualification. They also differed from the rest of the cohort in having significantly lower initial scores and non-significant deltas.

Effect of CAT on CORE-OM measured distress
The reduction in global distress of the 53 subjects in the cohort, measured by CORE-OM delta, was not only statistically significant it was also clinically significant, with sizable mean reduction of over 40%. Furthermore, among those with initial scores indicating clinical distress, just over a quarter showed a 63-94% reduction in CORE-OM scores, suggesting that in these individuals most of the distress was eliminated by the end of CAT. The reliable 'deterioration' (negative delta) in 2 clients could have shown they had had a bad week (CORE-OM refers only to the previous week) or possibly that therapy had increased their ability to trust and so they completed CORE-OM more openly.

The results of this study are encouraging, since 91% of the cohort showed improvement in CORE-OM score, and 51% showed reliable improvement (i.e., distinguishable from measurement error or chance), with 26% moving from clinical to non-clinical condition. Without trainees, reliable improvement was even better: 63%, with change from clinical to non-clinical in 35% of clients. Only 3.8% showed a reliable deterioration

in the entire cohort; 2.5% without the trainees. These results are intermediate amongst studies that have investigated the effect of CAT on CORE-OM (see below).

Pre-CAT CORE-OM distress was found to be less in the sub-cohort of those involved in psychology or psychiatry professions (Psych). They had lower initial CORE-OM scores (Table 3), and the psychology trainees, who were over half of this sub-cohort, had even lower initial scores. This suggests that these clients were less distressed than the rest, although it is possible that they might have been reluctant to show a high CORE-OM score, thinking they might be judged unsuitable to continue their training. Such guarding for fear of appearing 'not well enough' in the profession was mentioned by Williams' (Williams, 2013). Ideally, to examine this possibility, an independent in-depth clinical evaluation of distress would be required.

Relationship of CORE-OM delta to Pre-CAT scores
An interesting finding is the linear relationship between pre-CAT scores and the deltas that indicate the decline in distress. So, a client's pre-CAT CORE-OM score predicts the extent to which their distress improves following CAT (Figure 3 shows this linear relationship). This correlation holds from low to high pre-CAT scores, meaning as the pre-CAT score increases, there is a progressively greater absolute and proportional reduction in post-CAT CORE-OM, indicating improvement in distress. For example, a pre-CAT score of 0.5 predicts a drop of 0.09 or an 18% improvement in distress, while a pre-CAT score of 2.0 predicts a drop of 0.95 or a 48% improvement (see Figure 4). Although people providing and receiving CAT do not usually consider the extent to which improve-ment is likely to occur following CAT, this study suggests it can be predicted from the pre-CAT CORE-OM score, similar to the way risk of bone fracture in osteoporosis or risk of a cardiovascular event can be predicted. Consequently, those with the lowest pre-CAT scores, who are expected to have little or no improvement, might consider not having CAT, saving money, and improving their spirits in other ways. Still, these are typical responses; some people do better and others worse than expected; therefore, this can only be a guide.

A possible explanation for this relationship is that more severely distressed individuals are more responsive to CAT than those less severely distressed. Low initial pre-CAT CORE-OM scores in some individuals (like trainees) could affect this relationship, since they cannot exhibit large

reductions in CORE-OM scores, given that the lowest possible score is zero (the 'floor' effect). Thus, 18% of the total cohort who had'pre-CAT scores of <0.5 could not have reduced their scores by 0.5 points and so could not show a reliable improvement. This included most CAT trainees who were, however, too few to abolish the significant relationship between initial score and deltas seen for the entire cohort (N=53).

Another explanation for the relationship between the delta CORE-OM following CAT and the initial CORE-OM score concerns the statistical phenomenon of regression to the mean, which is the tendency for extremely high or low scores to come closer to the mean on retesting. The contribution of these two possible explanations cannot be accurately separated in this cohort study, although it is possible to do so in studies that include a control group (e.g., randomised control trials).

A relationship between initial CORE-OM score and the delta measured at the end of CAT does not appear to have been previously reported. However, a study at Maudsley Mental Health Trust (Evans et al., 2017a) did report a weak correlation (r=0.35) between first session CORE-OM scores and deltas following an amalgamation of different psychological therapies. Their correlation may have been much weaker than mine (r=0.73), not only because different therapies were provided by a variety of therapists (in 3-189 sessions) but also because they dealt with a more heterogeneous client group, including those suffering from severe conditions. An earlier study from the same Trust (Beck et al., 2015) also used a variety of different psychological therapies and reported an even lower correlation (r=0.31). However, this was for the relationship between CORE-OM scores at assessment (not the first session) and deltas after therapy. The waiting period from assessment to start of CAT was variable, typically several months, and so the study is not comparable to the present study.

Goodbye letters: CAT elements, client-therapist concordance, relationship to delta CORE-OM
Mention of CAT elements in Goodbye letters: It can be suggested that trainees were more likely than non-trainees to mention CAT elements in their Goodbye letters because they have recently learned the terms and/or wish to show their knowledge to their therapist. However, the data suggest that the differences between trainees and non-trainees in the use of CAT words were minor, very far from being significant, and would require very large samples to be formally examined. In any case, for any

client (trainee or not), using CAT language to report distress does not necessarily mean that their distress is better managed, but it might indicate a first step in distress reduction.

Goodbye letter concordance: Through their narrative commentary in Goodbye letters, clients may show that CAT has raised their consciousness (Weiskrantz, 1997) of some of their unhelpful attitudes, words, and actions. Therefore, inclusion of CAT elements – target problem (TP), reciprocal roles (RRs), and Traps (T), Dilemmas (D), and Snags (S)– in a client's Goodbye letter and corroborated by therapist's Goodbye narrative seemed indicative of some understanding of where change is needed.

The study found that concordance between therapist and client was high for RR (81% for full cohort and 80% for non-trainees) and for TP (81%/80%), but substantially less for T, D, S (58%/55%) (Table 4). This lower concordance of T, D, S may be because for some clients grasping reciprocal roles and target problem(s) is sufficient to carry on recognition and revision of their patterns, and so T, D, S are not introduced. Even when T, D, S are explained, they may be more complex and difficult for clients to put into words and so be omitted from Goodbye letters, raising some doubt about their concordance as an indicator of awareness of what needs to change.

In addition, there is also some statistical uncertainty about concordance, indicated by the range of results obtained by Method 2 (see Results Table 4).

While concordance would seem to suggest that clients did 'get' what was being discussed in therapy, lack of concordance may not indicate a lack of comprehension. Some clients who understood their TP, RRs, and even their T, D, S may have focussed on something else in their Goodbye letter: for example, detailing changes in their circumstances and feelings or thanking/criticising the therapist. Clients weren't given instructions about what to include in their letters. Instruction to use CAT language might have made concordance analysis easier but obviously would have biased results.

Even when there is client-therapist concordance for CAT elements, the letters may not be registering diverse aspects of the client's distress, captured by CORE-OM. And CORE-OM, although measuring distress reliably, refers only to the previous week and does not touch upon CAT-understanding developed over the previous months in therapy. CORE-OM cannot rate a client's ability to use CAT-understanding to reduce their distress, an ability important in the long-term (Ryle and Kerr, 2020).

Relationship between CORE-OM deltas and Goodbye letter concordance:
The study found no significant correlation between CORE-OM deltas and Goodbye letter concordance. This may be because CORE-OM deltas and Goodbye letter concordance are expressing and measuring different things.

> **BOX**
> **Author's reflection:** In hindsight, and for reasons indicated in the Discussion, a non-significant correlation between CORE-OM and Goodbye letter CAT-element concordance might have been expected. However, having read through more than 100 client Goodbye letters, it was learned that clients generally understood and were grateful for what was discussed, whether or not CAT elements were mentioned. The author was moved by their honesty, captivated by their ability to reflect, and pleased to see that even when disagreement or enduring resistance were touched on, there was always record of how CAT had made a deep impression. Like a good novel, a good therapy is absorbed, mulled over, revisited, and debated within the self, even years later. It was felt that concordance and other content in the Goodbye letters conveyed this spirit even if not correlated significantly to CORE-OM deltas.

As the discussion just above indicates, Goodbye letter concordance may not always be a robust indicator of reduced distress (CAT effectiveness) and is associated with some statistical uncertainty. Lack of correlation may also simply reflect that the client's letters give a more personal indication of distress change (CAT effectiveness) than CORE-OM. There could also be problems with CORE-OM as an overall measure of CAT's effectiveness as shown in studies that have used not only CORE-OM deltas to assess CAT effectiveness but also a variety of other tools, such as anxiety and depression questionnaires, personality structure questionnaires, and the Work and Social Adjustment Scale. Pre-post CAT changes registered as the effect size indices of tools other than CORE-OM can vary widely, sometimes several-fold, and while pre-post CORE-OM effect size indices were not extreme, they could differ twofold or more than that of other tools'(Evans et al., 2017b, Kellett et al., 2013). Also, a recent study observed that 71% % of clients reported at least one of the items of greatest importance to them was not covered by CORE-OM, highlighting the importance of an individualized outcome measure, such as a Goodbye letter (Sales et al., 2018).

Comparison with other studies identified through systematic literature search
The author compared the effectiveness of this CAT practice with those of CAT practice in nine other studies that used CORE-OM as an outcome measure. Five studies analysed data from only 7-17 clients, and only one had as many clients as this study. Unlike the author's private practice with just one therapist, seven of the eight studies with a one-to-one CAT format had multiple therapists (3-22); only one was in the private sector but used three therapists'(Baronian and Leggett, 2020); and only one had just one therapist (Kellett et al., 2020). The case-mix of clients in the author's practice (anxiety/depression/personal relationship difficulties) did not match that of several other studies; for example, two involved only clients with chronic pain'(Birtchnell et al., 2004, Baronian and Leggett, 2020) and one had only clients with bipolar disorder (Evans et al., 2017b). With these confounding variables, it is difficult to directly compare this study with others. Nevertheless, it was found that non-trainee clients had a mean pre-CAT CORE-OM score within the range of the other studies, albeit toward the lower end of the distribution, and a similar proportion of clients with reliable improvement and reliable clinically significant improvement. In this study, the proportion of clients with complete datasets (51%) also fell in the range of studies using CORE-OM to evaluate CAT effectiveness (≤ 41-91%) and was much higher than other retrospective studies using CORE-OM to judge effectiveness of a mixture of therapies (19% (Evans et al., 2017a); 10% (Beck et al., 2015).

Limitations

This study has several limitations.

1) While follow-up can be informative, it was not an aim to look at CORE-OMs filled in during the follow-up session, which is usually three months after CAT ends. This means that for this study, it could not be shown how sustained the improvement may have been.

2) Care should be taken not to generalise the findings of this study, which included an unusually high proportion of clients involved with the psychology/psychiatry profession.

3) The pre-post CAT study design was retrospective and uncontrolled, which makes it difficult to assess or predict if some clients would have improved without CAT. The only two RCTs that

examined the effect of a control group (receiving 'treatment as usual') on CORE-OM deltas reported very different results: one of them reported that CORE-OM deltas were almost as large as in the CAT group (Evans et al., 2017b) and the other that they were very much smaller than in the CAT group (Clarke et al., 2013).

4) The distribution and collection of CORE-OM forms in the author's CAT practice was patchy because there were no prior plans to undertake a study, with the result that only 51% were complete and amenable to paired analysis (pre-post CAT). Therefore, the present study is at risk of selection bias, although no evidence for this was found from the excluded clients' baseline characteristics – age, sex, pre-CAT CORE-OM scores – which did not differ significantly from those of included clients. Also, although data collection and entry were objective and checked, they were made by the author, which adds another small risk of bias.

5) An independent in-depth clinical evaluation of changes in each client would have complemented the CORE-OMs and Goodbye letters, but such evaluation was never available to the author.

6) Since Goodbye letters often do not include CAT elements and many CAT therapists don't even use Goodbye letters, the results of this study cannot be generalised.

Conclusion

This private practice study of clients with depression, anxiety, and/or relational problems has shown that for the cohort with and without trainees the mean reduction in CORE-OM measured distress following CAT exceeded 40% and most of that improvement was reliable. The study reports for the first time that improvements in CORE-OM measured distress are predictable from pre-CAT CORE-OM scores, which is of potential value to CAT practice. No significant relationship was found between improvements in CORE-OM measured distress and Goodbye letter concordance between client and therapist, perhaps because of the complexity of accurately comparing different aspects of CAT effectiveness (e.g., distress reduction vs. recognising and revising unhelpful patterns). □

References

Baker, J. 2003. Evaluation of CAT in GP Practice. *Reformulation,* Spring 2003, 16-17.

Barkham, M., Gilbert, N., Connell, J., Marshall, C. & Twigg, E. 2005. Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *Br J Psychiatry,* 186, 239-46.

Baronian, R. & Leggett, S. J. 2020. Brief cognitive analytic therapy for adults with chronic pain: a preliminary evaluation of treatment outcome. *British Journal of Pain,* 14, 57-67.

Beck, A., Burdett, M. & Lewis, H. 2015. The association between waiting for psychological therapy and therapy outcomes as measured by the CORE-OM. *Br J Clin Psychol,* 54, 233-48.

Birtchnell, J., Denman, C. & Okhai, F. 2004. Cognitive analytic therapy: Comparing two measures of improvement. *Psychology and Psychotherapy: Theory, Research and Practice,* 77, 479-492.

Calvert, R. & Kellett, S. 2014. Cognitive analytic therapy: a review of the outcome evidence base for treatment. *Psychol Psychother,* 87, 253-77.

Clarke, S., Thomas, P. & James, K. 2013. Cognitive analytic therapy for personality disorder: Randomised controlled trial. *British Journal of Psychiatry,* 202, 129-134.

Connell, J., Barkham, M., Stiles, W. B., Twigg, E., Singleton, N., Evans, O. & Miles, J. N. 2007. Distribution of CORE-OM scores in a general population, clinical cut-off points and comparison with the CIS-R. *Br J Psychiatry,* 190, 69-74.

Darongkamas, J., Newell, A., Hewitt-Moran, T. & Jordan, S. 2017. Clinical outcomes of congitive analytic therapy delivered by trainees. *Reformulation, Winter,* 6, 1-12.

Egger, M., Smith, G. D. & Altman, D. G. 2001. *Systematic Reviews in Health Care,* London, BMJ Publishing Group.

Evans, C., Connell, J., Barkham, M., Margison, F., Mcgrath, G., Mellor-Clark, J. & Audin, K. 2002. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *Br J Psychiatry,* 180, 51-60.

Evans, L. J., Beck, A. & Burdett, M. 2017a. The effect of length, duration, and intensity of psychological therapy on CORE global distress scores. *Psychol Psychother,* 90, 389-400.

Evans, M., Kellett, S., Heyland, S., Hall, J. & Majid, S. 2017b. Cognitive Analytic Therapy for Bipolar Disorder: A Pilot Randomized Controlled Trial. *Clinical psychology & psychotherapy,* 24, 22-35.

Gilbert, N., Barkham, M., Richards, A. & Cameron, I. 2005. The effectiveness of a primary care mental health service delivering brief psychological interventions: a benchmarking study using the CORE system. *Primary Care Mental Health,* 3, 241-251.

Gough, D., Oliver, S. & Thomas, J. 2017. *An Introduction to Systematic Reviews,* London, Sage Publishing Ltd.

Hallam, C., Simmonds-Buckley, M., Kellett, S., Greenhill, B. & Jones, A. 2021. The acceptability, effectiveness, and durability of cognitive analytic therapy: Systematic review and meta-analysis. *Psychology and psychotherapy,* 94, 8-35.

Jacobson, N. S. & Truax, P. 1991. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol,* 59, 12-9.

Kellett, S., Bennett, D., Ryle, T. & Thake, A. 2013. Cognitive analytic therapy for borderline personality disorder: therapist competence and therapeutic effectiveness in routine practice. *Clinical psychology & psychotherapy,* 20, 216-225.

Kellett, S., Easton, K., Cooper, M., Millings, A., Simmonds-Buckley, M. & Parry, G. 2020. Evaluation of a mobile app to enhance relational awareness and change during cognitive analytic therapy: Mixed methods case series. *JMIR Mental Health,* 7.

Martin, E., Byrne, G., Connon, G. & Power, L. 2021. An exploration of group cognitive analytic therapy for anxiety and depression. *Psychology and psychotherapy,* 94, 79-95.

Ryle, A. & Kerr, I. B. 2020. *Introducing Cognitive Analytic Therapy: Principles and Practice of a Relational Approach to Mental Health,* Hoboken, NJ, John Wiley & Sons, Inc.

Sales, C. M., Neves, I. T., Alves, P. G. & Ashworth, M. 2018. Capturing and missing the patient's story through outcome measures: A thematic comparison of patient-generated items in PSYCHLOPS with CORE-OM and PHQ-9. *Health Expect,* 21, 615-619.

Stiles, W. B., Barkham, M. & Wheeler, S. 2015. Duration of psychological therapy: relation to recovery and improvement rates in UK routine practice. [corrected]. *Br J Psychiatry,* 207, 115-22.

Trujillo, A., Feixas, G., Bados, A., Garcia-Grau, E., Salla, M., Medina, J. C., Montesano, A., Soriano, J., Medeiros-Ferreira, L., Canete, J., Corbella, S., Grau, A., Lana, F. & Evans, C. 2016. Psychometric properties of the Spanish version of the Clinical Outcomes in Routine Evaluation - Outcome Measure. *Neuropsychiatr Dis Treat,* 12, 1457-66.

Weiskrantz, L. 1997. *Consciousness Lost and Found,* Oxford, Oxford University Press.

Williams, B. 2013. Past hurts and therapeutic talent. *Reformulation*, 39-42.

Williams, T. & Craven-Staines, S. 2017. Exploring the efficacy of cognitive analytic therapy. *Int. Journal of CAT & RMH,* 1, 164-178.